STUDIES OF CLONED REPETITIVE DNA SEQUENCES IN THE SEA URCHIN GENOME

Thesis by

Richard Henry Scheller

In Partial Fulfillment of the

Requirements for the Degree

of Doctor of Philosophy

California Institute of Technology

Pasadena, California

1980

(Submitted April 21, 1980)

Acknowledgements

This thesis would not have been possible without the numerous collaborations within the Davidson and Britten groups. I would like to thank in particular my thesis advisor Eric Davidson for his help and guidance over the last several years. I wish to acknowledge the support of a National Institute of Health training grant over the years of my graduate work.

Abstract

The sequence organization, evolution, and transcription of cloned repetitive DNA elements in sea urchin were studied.

Total S. purpuratus DNA was reassociated such that only the reiterated sequences were duplex. The flanking single-stranded DNA was digested with S1 nuclease and synthetic Eco RI sites were ligated to the ends. These molecules were cloned in RSF2124 and recombinants were selected by their $Amp^{+}Col^{-}$ phenotype. These clones ranged in reiteration frequency from 3 to 12,000 copies per haploid genome. The mean intrafamilial divergence was as low as 4°C for some families which could be attributed entirely to polymorphism. Other families contained members that showed greater divergence, some as high as 25°C.

Nine randomly selected sequences were found to be transcribed in oocyte RNA, gastrula hnRNA, and intestine hnRNA. The concentrations of these transcripts are tissue specific and unlike single copy sequences both strands were represented in the RNAs. The transcripts were longer than the repetitive elements themselves indicating linkage to single copy sequences.

The number of copies per haploid genome was determined in S. franciscanus and L. pictus in addition to the parent species. The ratio of the reiteration frequency in S. purpuratus to S. franciscanus ranged from about 20 to 1. The copies remaining in S. franciscanus and L. pictus genomes were, however, conserved relative to average single copy DNA sequence.

The plasmid clones were used to select individual family members from total genomic λ libraries and their characteristics were investigated. Families 2034 and 2108 were found to be members of the long repetitive sequence class. The 2034 elements exist in clustered arrays while the 2108 family is dispersed. The 2109 elements were shown to be largely of the short repetitive class, however, about 10% of the members occur in long repetitive regions.

The interspersed elements were found to be flanked by low reiteration frequency or single copy sequences. Thus, a large number of genomic regions are physically linked via the 2109 repetitive sequences. The implications of these findings with respect to evolution and gene regulation are many-fold and are discussed in detail in the following chapters.

Table of Contents

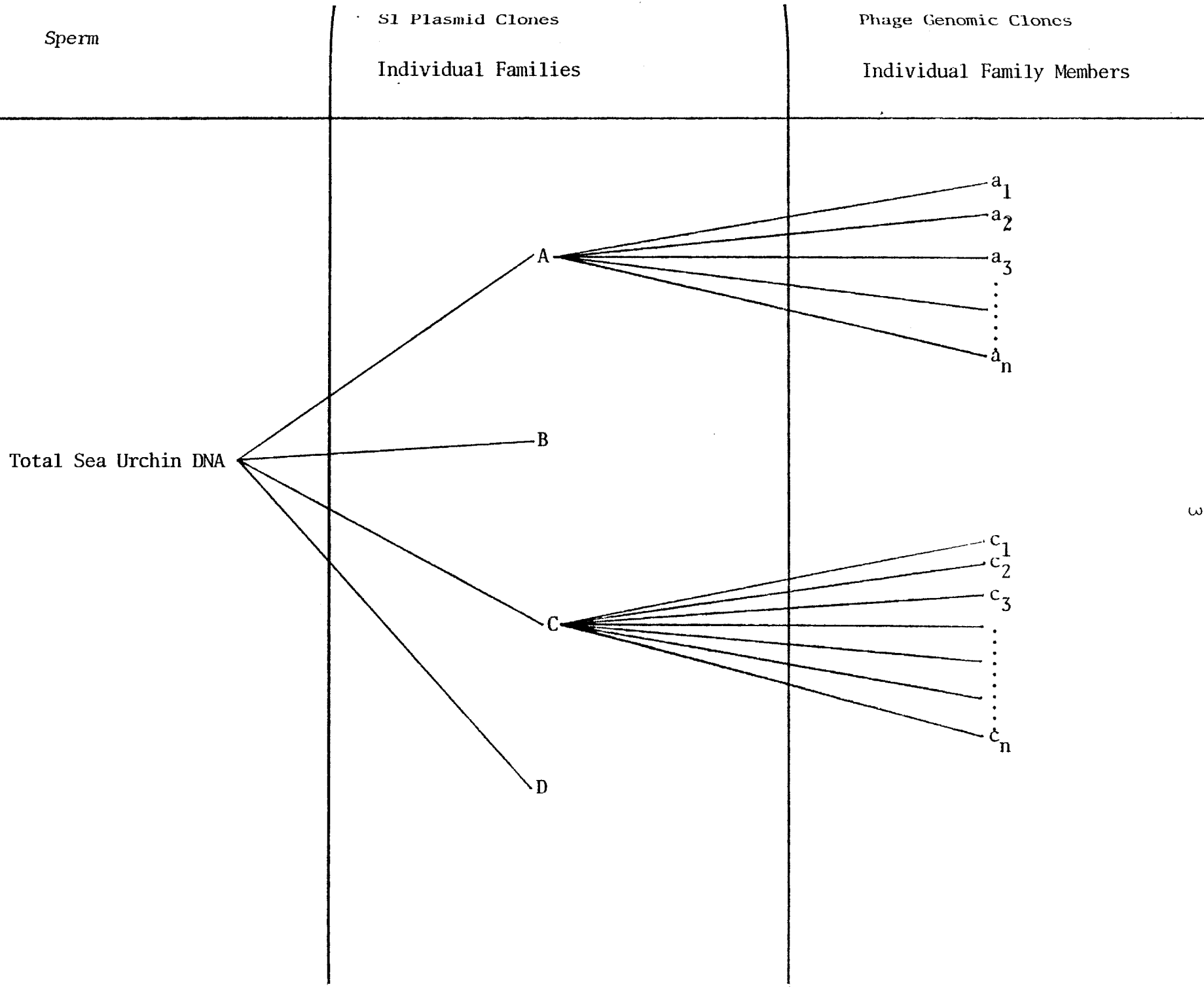
CHAPTER 1

Introduction

The reassociation kinetics of animal DNAs show large fractions which react faster than would be expected on the basis of the cellular DNA content (Britten and Kohne, 1968). This ubiquitous feature of animal genomes not found in bacteria or viruses indicates that a certain fraction of the DNA is present in multiple copies. This dissertation is a characterization of the sequence organization, evolution and transcription of repetitive sequences in the sea urchin S. purpuratus. Two rounds of recombinant DNA molecular cloning in plasmids and bacterophage allowed the isolation of repetitive sequence family members as outlined in Fig. 1.

S. purpuratus provides an excellent model system for investigating repetitive sequence function. This species is of intermediate evolutionary complexity and has a genome size of .8 pg. 25% of the DNA is repetitive and the unique DNA represents 8.3×10^8 nucleotide pairs (Graham et al., 1974). This information is complemented by the vast amount of knowledge regarding sea urchin embryogenesis (for review see Davidson, 1976). Early workers characterized the morphological events of early development and demonstrated that blastomeres from the four cell stage could give rise to adults.

Development actually begins with maturation of the oocyte which contains RNA transcribed from about 6% of the genome or 3.7×10^7 nucleotide pairs of RNA (Galau et al., 1976; Hough-Evans et al., 1977). This RNA is known as a maternal messenger RNA as it is stored in the egg until needed soon after fertilization (Hough et al., 1975). The fourth cell division produces 4 large macromeres and 4 small micromeres which are positioned at the vegetal pole. The micromeres contain a distinct subset of the RNA molecules representing 75% of the complexity of the total embryo (Ernst et al., 1980). Protein synthesis increases 10-20 fold after fertilization; however, the pattern of prevalent protein molecules remains


Sperm
S1 Plasmid Clones
Individual Families
Phage Genomic Clones
Individual Family Members
Total Sea Urchin DNA
A
B
C
D
a1
a2
a3
an
c1
c2
c3
cn

qualitatively the same until gastrulation where the 600 cell embryo shows minor changes in the complex pattern (Epel, 1967; Brandhorst, 1976). The feeding pleutus consists of about 1200 cells and is the culmination of 90 hr of development during which there has been little or no increase in embryo mass. Many differentiated tissues exist including skeletal, digestive and motility systems. Subsets of the maternal messenger RNA molecules are the only detectable sequences on the polysomes until the gastrula stage where an additional complexity of 2.6×10^6 nucleotide pairs is observed. Adult tissues are characterized by relatively small messager RNA complexities of about 6×10^6 nucleotide pairs (Galau, 1974, 1976).

RNA from the nuclei of cells is of larger complexity than the cytoplasm and represents transcription from one-third of the genome or 1.7×10^8 nucleotide pairs (Hough et al., 1975). Furthermore, genes which are not represented on polysomes actively making proteins in a particular tissue are transcribed into nuclear RNA of that tissue (Wold et al., 1978). The size of the nuclear RNA transcript is longer than the polysomal messenger RNA and the turnover rate is about 25 min, about 20 times faster than message. These data indicate processing of the primary transcript molecules. While the nuclear RNAs of gastrula embryos and adult intestine share most of their complexity, 3.5×10^7 nucleotides were found specific to the intestine heterogeneous nuclear RNA (Ernst, 1979). These sequences may be of a regulatory nature, since as mentioned previously, the bulk structural genes are transcribed in all nuclear RNAs. These results with whole cell probes have been confirmed and expanded upon using cloned tracers specific to messenger RNA sequences (Lev et al., 1980; Lee et al., 1980).

The above presentation is merely a description of a few of the processes involved in sea urchin development. A central theme is lacking. What is the link between the information in the genome and the morphogenesis of an organism?

It is the purpose of the remainder of this discussion to review the data concerning the possible role of repetitive sequence in this process.

There are about 5000 repetitive sequence families in the sea urchin genome with an average reiteration frequency of 100 (Eden et al., 1977). Long DNA can be denatured and allowed to renature to a point where only reiterated sequences are in duplexes. If this reaction mixture is treated with S1 nuclease, the single stranded DNA will be specifically degraded. The size distribution of the spared molecules can be resolved by passing the mixture over a Sepharose CL-2B Column providing effective separation due to differences in molecular weights. An interesting elution pattern is observed. About half of the molecules are excluded indicating a mean length of about 2000-3000 nucleotides, while the rest of the repetitive elements are retarded by the Sepharose chromatography indicating a mean length of 300 nucleotide pairs (Britten et al., 1976). These repetitive sequences are found in all higher eucaryotes; however, the ratio of long to short elements has been found to vary (Crain et al., 1976a,b; Manning et al., 1975).

A typical region of the sea urchin genome contains repetitive sequences interspersed with single copy DNA at an average distance of 2000 nucleotides (Graham et al., 1974). Repetitive sequences do however exist in other modes of organization as well, some of which will be discussed later. The interspersed repetitive sequences are represented in nuclear RNA of higher organisms reflecting the pattern of the genome (Federoff et al., 1977; Smith et al., 1974). In addition to this observation, Davidson et al. (1975b) showed that mRNA molecules in sea urchin embryos are transcribed from DNA sequences adjacent to repetitive regions. These observations are strong support for the Britten and Davidson model of coordinate gene regulation (Britten and Davidson, 1969).

The concept behind this model is that genes are organized into groups called batteries. The products of these genes are presumed to be proteins which perform related functions. A battery could, for example, consist of a set of enzymes required in a particular metabolic pathway. It was further proposed that what physically links the component genes of a battery are the members of a repetitive sequence family. These authors also pointed out that much of eucaryotic evolution may have occured by changing battery composition, rather than developing completely new enzymatic activities or structural components (Britten and Davidson, 1971). To further test these hypotheses, individual repetitive sequence families and individual family members were isolated using recombinant DNA techniques.

S1 treated repetitive elements are not flanked by naturally occurring restriction enzyme sites. In order to clone these molecules, we synthesized DNA molecules containing restriction enzyme recognition sites for Eco RI Bam I, and Hind III (Chapter 2). These self-complementary molecules, in addition to the 6 base recognition sites, were flanked by CC on the 5' end and GG on the 3' end. This stabilized the 10 base pair duplexes resulting in a T_m of 42°C in .1 M NaCl. Sea urchin DNA greater than 2000 nucleotides in length was denatured and renatured, so that only the repetitive sequences had time to nucleate and form duplexes. Digestion of this reaction mix with S1 nuclease cleaved the single stranded DNA which consisted of unreacted repetitive sequences and single copy DNA. The synthetic Eco RI restriction enzyme recognition site was ligated to the S1 resistant duplexes using T4 DNA ligase. After cleavage with Eco RI, the repeated sequences were ligated to the plasmid RSF2124 which contains a gene for resistance to the antibiotic, ampicillin. These chimeric molecules were used to transform E. coli, providing clones of individual repetitive sequence elements could be propagated in large amounts for further studies.

The overall characteristics of these families were found to be representative of the genomic repetitive elements. The reiteration frequencies varied from 3 to 12,000 copies per haploid genome while the lengths of these cloned fragments ranged from as short as 125 nucleotides to as long as 1100 nucleotides. Shorter clones may have been missed, due to the difficulty in detecting very short fragments. It had been known since the early measurements of Eden et al. (1977) that the members of a repetitive sequence family were not identically homologous. To further investigate this issue, hybrids were formed between sea urchin genomic DNA and the cloned repetitive element. The thermal stability of these hybrids was determined and used as a measure of sequence divergence within the family. Thirteen of these families lacked any detectable highly divergent family members while five families were found to contain many, and in some cases, almost all divergent members. The functional significance of the intra-family divergence is not understood. It is possible that many sequences of a nondivergent family are constrained within the family structure by selective pressure within the S. purpuratus species (Chapter 3).

The evolution of these families was studied by determining the thermal stability and reiteration frequency of the cloned fragments in two other species, S. franciscanus and L. pictus. The ratio of the repetition frequency in S. purpuratus to S. franciscanus ranged from 20 for clone 2133A to about 1 for clone 2108. S. franciscanus hybrids of two clones, 2109B and 2090, have thermal stabilities within 1°C of the purpuratus hybrids compared to a ΔT_m of 10°C for the average single copy DNA. The repetitive sequences have therefore diverged in general less than the average single copy DNA sequence and must have a selective constraint maintaining the sequence between species.

Yet while constraint on the sequence exists, the size of the family was found to vary between species. This observation can be understood if different

families have the same functions in related genomes. Repetitive families could be substituting for each other effecting an evolutionary change at many different genomic locations simultaneously (Chapter 7).

The repetitive sequence clones were radioactively labeled at their 5' ends, strand separated on polyacrylamide gels and stripped of contaminating complementary strands. These labeled fragments were found to hybridize with RNA proving that the repetitive elements are transcribed. Unlike single copy sequences, however, both strands were found represented in the RNAs. This is most likely the result of asymmetric transcription from several different locations where the repeat has either orientation with respect to the transcription unit. The repeat transcripts were found to be much longer than the repetitive sequence (i.e., several thousand nucleotides) indicating that they probably contain single copy sequences as well. The pattern of representation for the cloned repetitive elements was found to be tissue specific. For instance, the representation of repeat A might be 10 times higher than repeat B in oocyte RNA, and the opposite true for intestine RNA. This is consistent with the idea that different regions of the genome are expressed coordinately with the physical link between regions the repetitive element.

Viewing these results in terms of the regulatory hypothesis presented earlier and the representation of embryonic messenger RNAs in adult nuclear RNA led to further ideas on the possible function of repetitive sequences (Davidson and Britten, 1979; and Chapter 6). It was proposed that the concentration of the repetitive sequences in nuclear RNA serves to regulate which genes become represented on the polysomes. The RNA transcript itself may be a diffusible regulatory element and act either through DNA-RNA or RNA-RNA interactions (Chapters 5 & 6).

Costantini et al. (1980) have isolated repeat containing RNA and shown that at least about half and possibly all of the maternal messenger RNA stored in the mature sea urchin oocyte contains repetitive elements covalently linked to single copy sequences. Their analysis concluded that the few hundred kinds of prevalent repeat transcripts are each associated with between 10 and 50 different rare messages physically defining batteries. Moore et al. (1980) showed that while the reiteration frequency of the cloned fragments has changed the prevelance in oocyte, RNA has remained the same. These results lead one to believe that the transcribed repeats are conserved and that possibly their concentration in the oocyte is important.

The next level of understanding would involve the isolation of manyindividual membe sea urchin DNA from several species were inserted into the λ bacterophage Charon 4 (Maniatis et al., 1978). The average insert size was approximately 15 kb and a total of about 1 million S. purpuratus recombinants have been obtained over the last 2 years. The libraries, as they are called, contain enough recombinants such that there is a greater than 90% chance that any single copy sequence will be represented. The plasmid repetitive sequence clones were radioactively labeled and used to screen the libraries for individual family members. The genome screening showed immediately that the repetitive elements were dispersed throughout the genome in that the number of positive spots obtained was consistent with there being one or few copies of the repeat per recombinant.

The 2109 family was usually found to be flanked by single copy DNA, although some members were in long repetitive regions. These elements usually occurred singly on a λ clone; however, 4 of the 30 clones investigated contained two copies of the repeat. Even these few occurrences of greater than one copy per clone are above the random expectation indicating an interesting organization of the family members with respect to each other. The 2108 and 2034 repetitive

elements were found to be flanked by repetitive DNA of the same reiteration frequency. However while the 2034 family often contained many copies on a given clone, 2108 was to a large extent dispersed (Chapter 8). While both of these elements are long, they have a very different sequence organization and, as will be seen in the following paragraphs, the intra-family sequence relationships differ as well.

To generate a more detailed picture of the sequence relationships between family members, the thermal stabilities of hybrids between the plasmid clone and individual λ clones were determined. The 2034 hybrids all melted $\leq$ 6°C below the self T_m. Clone λ2034-4 which contains elements from the 1.7 kb prominent genomic land (Moore et al., 1980b), and melts only 2°C below the self T_m indicating polymorphism less than the 4°C observed for the average single copy sequence (Britten et al., 1978). The 2108 λ clones, however, show a different organization. There are two types of hybrids formed, those with stability $\leq$ 6°C below the self T_m, and those with ΔT_ms $\geq$ 20°C. No intermediate stability fragments were found.

The clones that showed (high) ΔT_ms with the 2108 tracer were found to contain well-matched members in the genome. These observations defined the structure of this family as one of several subfamilies containing closely related members that are all distantly related to each other. When restriction enzyme fragments from bands flanking the region of 2108 homology were reacted with digests of the set of 2108 clones bound to nitrocellulose, an interesting phenomenon arose. Closely related sequences were found to be colinear for a long distance $\geq$ 5 kb. In contrast, the order of the components between subfamilies was different. For example, if the order of elements within a subfamily is ABC, these same sequences might be found in the order BAC in a different subfamily. The size of the 2108 superfamily was estimated to be about 500 copies, while we believe the average size of a subfamily to be 20 copies/haploid genome.

The 2109 family was found to have a similar organization, although it was even more involved. Hybrids between these elements were found to have clustered T_ms. This was interpreted as defining increasingly longer homologous regions. For example, clones of type I might have a single colinear element A, type II AB, and type III ABC. Heteroduplexes of clones of type I with clones of type I, II, or III would have the same T_m call it T_m (I). It follows that T_m (I) < T_m (II) < T_m (III) and the reason for this arrangement is the increasing length of the homologous region (Scheller et al., 1980a). The long repetitive sequences in sea urchin are less divergent (Moore et al., 1980b) indicating that they are perhaps of recient origin. As a family evolved, it would then accumulate base changes while it also scrambled its organization with other repetitive elements.

Wensink et al. (1980) have shown this type of sequence arrangement in Drosophila using cloned long repetitive regions. Fragments from these clones reacted in different orders with fragments from other genomic regions. Other studies on sea urchin long repeat sequences are also consistent with these ideas.

These data show that repetitive sequence families have a complex structure that is certainly not completely understood in terms of the evolution and function of these elements. It should be pointed out that in considering these questions, the blocks that are seen to be the elements used to build repeat families may also be the units of functional significance.

The nucleotide sequences of the plasmid clones have demonstrated that the interspersed elements form discrete families (Posakony et al., 1980). The presence of stop codons in most of the elements in all reading frames is further evidence that these sequences do not code for proteins. The interesting observation was made that there exist a large number of 6-15 nucleotide inverse repeats centered on what is also a direct repeat. These structures arise from the duplication (either direct or inverse) of what is already an inverse repeat. The intra-family

block relationships were demonstrated to be correct with primary sequence data. The blocks themselves were as small as 18 NTP and ranged in divergence from 2.5-15.5%. The meaning of these arrangements and how they were generated is unknown. It seems, however, that the intricate network of interactions physically linking different regions of the genome has a new dimension added to it. A region containing elements ABC is linked to all regions containing either element. Another occurrence of B may be in the context of BD linking A to D through B, and so on. These relationships may be important in defining gene batteries and could encode the information needed direct to the pathway of ontogeny.

The organization of repetitive sequences in the vicinity of gene regions has been investigated. These studies place repetitive elements in several positions with respect to the protein coding region and probable transcription unit. In Dictyostelium, a repetitive family has been shown to occur at the 5' ends of messenger RNA sequences (Kimmel and Firtel, 1979; Kindle and Firtel, 1979). The initiation of transcription probably occurs midway through the repetitive element. These sequences are asymmetrically transcribed, unlike the sea urchin repeats discussed earlier. The rabbit globin gene cluster consists of 44 kb and contains 4 linked genes. At least 5 tr>rsy families are present in no fewer than 20 elements ranging in size from 140-1400 nucleotides. These sequences are present in the regions between the genes and are represented many times in the genome outside the cluster (Shen and Maniatis, 1980). The rat albumin gene has 13 introns, 2 of which contain repetitive sequences and are therefore almost certainly transcribed repeats that are spliced out of the primary transcript (Sargent et al., 1979). In addition to this information, a cloned sea urchin polysomal poly(A)+ RNA has been found to contain a repetitive sequence on the 3' end of the message. In contrast to the Dictyostelium family transcription may end rather than initiate midway through these repeats (Scheller et al., 1980b).

At least 2 families of repetitive sequences have been found to give rise to short transcripts. These families have been implicated in RNA splicing and DNA replication (Rubin et al., 1980; Jelinek et al., 1980). Some of these RNAs have been shown to be hydrogen bonded to hnRNA and cytoplasmic message (Jelinek et al., 1978a,b). The transcription of these families, however, differs from the repetitive families discussed earlier in that the transcribed repetitive elements are thought not be be parts of larger transcription units. If repetitive elements as a class have a generalized function, it is possible that the pattern of interspersion is superimposed upon the transcription unit intron exon pattern of the genome. In other words, repetitive elements in all these different locations would function in similar ways. On the other hand, the existence of 5' or 3' families and short RNAs would argue that the position of the element was important for the function and that the function or at least the mechanisms of action of different families may be different. Transcribed repeats that are parts of introns may function at the level of processing of the primary transcript (Davidson, et al., 1977). While the 5' repeats may function at the level of initiation of transcription, it seems as though these and the 3' families leave at least remnants of the repeat on the fully processed polysomal message. The presence of the repetitive element in the cytoplasm is probably important for the ultimate function, perhaps via a subcellular localization process or translational regulation.

While the mechanism of amplification and dispersal of all families may be similar different families probably have very different functions. These functions remain to date unknown. The transcription, evolutionary conservation and intricate pattern of organization seem to require a central role in the biological processes of higher organisms for repetitive elements.

References

Brandhorst, B. P. (1976). Two-dimensional gel patterns of protein synthesis before and after fertilization of sea urchin eggs. Develop. Biol. **52**, 310.

Britten, R. J., and Davidson, E. H. (1969). Gene regulation for higher cells: a theory. Science **165**, 349-357.

Britten, R. J., and Davidson, E. H. (1971). Repetitive and nonrepetitive DNA sequences and a speculation on the origins of evolutionary novelty. Quart. Rev. Biol. **46**, 111-138.

Britten, R. J., Graham, D. E., Eden, F. C., Painchaud, D. M., and Davidson, E. H. (1976). Evolutionary divergence and length of repetitive sequences in sea urchin DNA. J. Mol. Evol. **9**, 1-23.

Britten, R. J., and Kohne, D. (1968). Repeated sequences in DNA. Science **161**, 529-540.

Costantini, F. D., Britten, R. J., and Davidson, E. H. (1980). Message sequences and short repetitive sequences are interspersed in sea urchin poly(A)+ RNAs. Nature, submitted.

Crain, W. R., Eden, F. C., Pearson, W. R., Davidson, E. H., and Britten, R. J. (1976a). Absence of short period interspersion of repetitive and nonrepetitive sequences in the DNA of Drosophila melanogaster. Chromosoma (Berl.) **56**, 309-326.

Crain, W. R., Davidson, E. H., and Britten, R. J. (1976b). Contrasting patterns of DNA sequence arrangement in Apis mellifera (Honey bee) and Musca domestica (Housefly). Chromosoma (Berl.) **59**, 1-12.

Davidson, E. H. (1976). Gene Activity in Early Development. (New York: Academic Press).

Davidson, E. H., and Britten, R. J. (1973). Organization, transcription and regulation in the animal genome. Quart. Rev. Biol. **48**, 565-613.

Davidson, E. H., and Britten, R. J. (1979). Regulation of gene expression: possible role of repetitive sequences. Science **204**, 1052-1059.

Davidson, E. H., Galau, G. A., Angerer, R. C., and Britten, R. J. (1975a). Comparative aspects of DNA sequence organization in metazoa. Chromosoma (Berl.) **51**, 253-259.

Davidson, E. H., Hough, B. R., Klein, W. H., and Britten, R. J. (1975b). Structural genes adjacent to interspersed repetitive DNA sequences. Cell **4**, 217-238.

Davidson, E. H., Klein, W. H., and Britten, R. J. (1977). Sequence organization in animal DNA and a speculation on hnRNA as a coordinate regulatory transcript. Dev. Biol. **55**, 69-84.

Eden, F. C., Graham, D. E., Davidson, E. H., and Britten, R. J. (1977). Exploration of long and short repetitive sequence relationships in the sea urchin genome. Nucl. Acids Res. **4**, 1553-1567.

Epel, D. (1967). Protein synthesis in sea urchin eggs. A "late" response to fertilization. Proc. Natl. Acad. Sci. USA **57**, 899.

Ernst, S. G., Britten, R. J., and Davidson, E. H. (1979). Distinct single-copy sequence sets in sea urchin nuclear RNAs. Proc. Natl. Acad. Sci. USA **76**, 2209-2212.

Ernst, S. G., Hough-Evans, B. R., Britten, R. J., and Davidson, E. H. (1980). Limited complexity of the RNA in micromeres of 16-cell sea urchin embryos. Submitted, Developmental Biology.

Federoff, N., Wallauer, P. K., and Wall, R. (1977). Intermolecular duplexes in heterogeneous nuclear RNA from HeLa cells. Cell **10**, 597-610.

Galau, G. A., Britten, R. J., and Davidson, E. H. (1974). A measurement of the sequence complexity of polysomal messenger RNA in sea urchin embryos. Cell **2**, 9.

Galau, G. A., Klein, W. H., Davis, M. M., Wold, B. J., Britten, R. J., and Davidson, E. H. (1976). Structural gene sets active in embryos and adult tissues of the sea urchin. Cell **7**, 487-505.

Graham, D. E., Neufeld, B. R., Davidson, E. H., and Britten, R. J. (1974). Interspersion of repetitive and non-repetitive DNA sequences in the sea urchin genome. Cell **1**, 127-138.

Hough, B. R., Smith, J. J., Britten, R. J., and Davidson, E. H. (1975). Sequence complexity of heterogeneous nuclear RNA in sea urchin embryos. Cell **5**, 219-299.

Hough-Evans, B. R., Wold, B. J., Ernst, S. G., Britten, R. J., and Davidson, E. H. (1977). Appearance and persistence of maternal RNA sequences in sea urchin development. Dev. Biol. **60**, 258-277.

Jelinek, W., Evans, R., Wilson, M., Salditt-Georgieff, M., and Darnell, J. E. (1978). Oligonucleotides in heterogeneons nuclear RNA: similarity of inverted repeats and RNA from repetitious DNA sites. Biochemistry **17**, 2776-2783.

Jelinek, W., and Leinwand, L. (1978). Low molecular weight RNAs hydrogen-bonded to nuclear and cytoplasmic poly(A)-terminated RNA from cultured Chinese hamster ovary cells. Cell **15**, 205-214.

Jelinek, W. R., Toomey, T. P., Leinwand, L., Duncan, C. H., Biro, P. A., Choudary, P. U., Weissman, S. M., Rubin, C. M., Houck, C. M., Deininger, P. L., and Schmid, C. W. (1980). Ubiquitous, interspersed repeated sequences in mammalian genomes. Proc. Natl. Acad. Sci. USA, in press.

Kimmel, A. R., and Firtel, R. A. (1979). A family of short, interspersed repeat sequences at the 5' end of a set of Dictyostelium single-copy mRNAs. Cell **16**, 787-796.

Kindle, K. L., and Firtel, R. A. (1979). Evidence that populations of Dictyostelium single-copy mRNA transcripts carry common repeat sequences. Nucl. Acids Res. **6**, 2403-2422.

Lee, A. S., Thomas, T. L., Lev, Z., Britten, R. J., and Davidson, E. H. (1980). Four transcript sizes from a cloned sea urchin gene coding for rare early embryo messages. Proc. Natl. Acad. Sci. USA, in press.

Lev, Z., Thomas, T. L., Lee, A. S., Angerer, R. C., Britten, R. J., and Davidson, E. H. (1980). Developmental expression of two cloned sequences coding for rare sea urchin embryo messages. Dev. Biol., in press.

Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K., and Efstratiadis, A. (1978). The isolation of structural genes from libraries of eucaryotic DNA. Cell **15**, 687-701.

Manning, J. E., Schmidt, C. W., and Davidson, N. (1975). Interspersion of repetitive and nonrepetitive DNA sequences in the Drosophila melanogaster genome. Cell **4**, 141-155.

Moore, G. P., Costantini, F. D., Posakony, J. U., Davidson, E. H., and Britten, R. J. Evolutionary conservation of repetitive sequence expression in sea urchin egg RNAs. Science, in press.

Moore, G. P., Pearson, W., Davidson, E. H., and Britten, R. J. Studies of the evolution of sequence organization in sea urchin. In preparation.

Posakony, J. W., Scheller, R. H., Anderson, D. M., Britten, R. J., and Davidson, E. H. Nucleotide sequences of repetitive DNA. In preparation.

Rubin, C. M., Houck, C. M., Deininger, P. L., Friedmann, T., Schmid, C. W. Partial nucleotide sequence of the 300 NT interspersed repeated human DNA sequences. In preparation.

Sargent, T. D., Wu, J. R., Sala-Trepat, J. M., Wallace, R. B., Reyes, A. A., Bonner, J. (1979). The rat serum albumin gene: analysis of cloned sequences. Proc. Natl. Acad. Sci. USA **76**, 7, 3256-3260.

Scheller, R. H., Anderson, D. M., Posakony, J. W., McAllister, L. B., Britten, R. J., and Davidson, E. H. Complex sequence homologies of repetitive elements. In preparation.

Scheller, R. H., Anderson, D. M., Britten, R. J., and Davidson, E. H. Unpublished observations.

Shen, Che-Kun, and Maniatis, Tom. (1980). The organization of repetitive sequences in a cluster of rabbit β-like globin genes. Cell. In press.

Smith, M. J., Chamberlin, M. E., Hough, B. R., and Davidson, E. H. (1974). Repetitive and non-repetitive sequences in sea urchin hnRNA. J. Mol. Biol. **85**, 103-126.

Wensink, P. C., Tabata, S., Pachl, C. (1980). The clustered and scrambled arrangement of moderately repetitive elements in Drosophila DNA. Cell. In press.

Wold, B. J., Klein, W. H., Hough-Evans, B. R., Britten, R. J., and Davidson, E. H. (1978). Sea urchin embryo mRNA sequences expressed in the nuclear RNAs of adult tissues. Cell **14**, 941-950.

CHAPTER 2

Reprinted from
8 April 1977, Volume 196, pp. 177-180

SCIENCE

Chemical Synthesis of Restriction Enzyme Recognition Sites Useful for Cloning

Abstract. *By a triester chemical synthesis method, three decameric DNA's have been made; these act as substrates for several restriction endonucleases, including Eco RI, Bam I, and Hind III. These homogeneous decamers form duplexes that can be efficiently blunt-end ligated to themselves or to other DNA molecules by the action of T4 DNA ligase and thus are useful tools for molecular cloning experiments.*

Restriction endonucleases recognize specific sequences in double-stranded DNA. Many of these endonucleases cleave their recognition sites to produce blunt ends, while others such as Eco RI, Bam I, and Hind III, leave cohesive or "sticky" ends (*1–3*). DNA fragments with these cohesive ends can be inserted easily into, and subsequently excised from, a suitable cloning vehicle such as the plasmid pMB9 (*4*), which has one site each for Eco RI, Bam I, and Hind III.

Unfortunately, cohesive ends are lacking in many DNA fragments of interest, such as randomly cleaved DNA, complementary DNA (cDNA), and many restriction endonuclease fragments. We have recently introduced a new general procedure suitable for cloning almost any DNA molecule (*5*). This method (Fig. 1) involves the addition by blunt-end ligation of short DNA segments to both ends of the DNA to be cloned. The added segments contain restriction endonuclease cleavage sites, and treatment with the corresponding endonuclease leaves cohesive ends on the subject DNA. This DNA then can be cloned by incorporating it into a plasmid or other vector that has been opened with the same restriction enzyme. The feasibility of this new method was demonstrated by adding an octadeoxyribonucleotide containing the Eco RI recognition sequence to a 21-base-pair duplex bearing the *lac* operator sequence. A clonable, excisable DNA fragment was obtained, and was shown to function as *lac* operator should, both in vitro and in vivo (*5*).

We report here the chemical synthesis by the improved triester method of Itakura *et al.* (*6*) of three new decameric "linker" molecules (see Fig. 1), which together contain seven different recognition sites cleavable by 23 different restriction endonucleases including Eco RI, Bam I, and Hind III (*2*). These linker molecules should add great flexibility to cloning methodology, and also provide substrates for physicochemical studies on restriction enzyme recognition, methylation, and cleavage mechanisms.

The synthetic scheme in Fig. 2 outlines the strategy used in the convergent synthesis of the Bam I decamer by the triester method, using 2,4,5-triisopro-

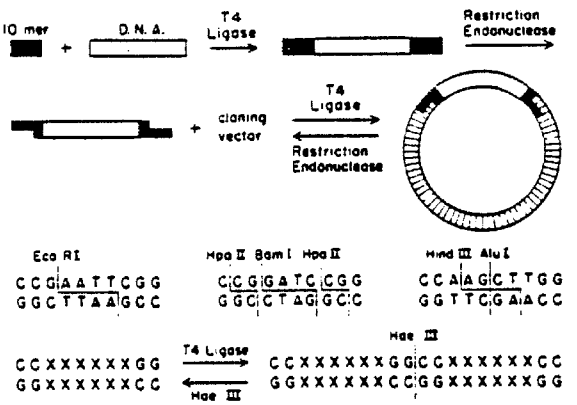


Fig. 1. (Top) A decameric linker (10 mer) bearing a restriction enzyme site is joined by T4 DNA ligase to both ends of the DNA to be cloned, and cohesive ends are then produced by treatment with restriction endonuclease. This "sticky-ended" DNA then can be incorporated into a vector that has been cut open with the same restriction endonuclease. (Bottom) The three chemically synthesized decameric linker molecules investigated in this study. The ends of the decamer sequences are designed so that, after blunt-end ligation with T4 DNA ligase, a site recognized by the Hae III restriction enzyme is produced.

pylbenzenesulfonyl tetrazolide (TPST) as a condensing reagent (*7*). The other two decamers were made similarly. Procedures and reaction conditions are given in the legend of Fig. 2. Two milligrams of fully deblocked Eco RI decamer were obtained, whereas for the Bam I and Hind III decamers the yields were 26 mg and 80 mg, respectively. The lower yield of synthesis of the Eco RI decamer may be attributable to the particular deblocking conditions used for the removal of the β-cyanoethyl group on the 3' ends. These conditions subsequently were changed for the Hind III and Bam I decamer syntheses (see legend of Fig. 2). The pyridine-ethanol mixture provided a faster, more efficient removal of β-cyanoethyl protecting groups than tetrahydrofuran did. The use of pressure silica gel chromatography for separation of protected oligonucleotides greatly aids the triester synthesis. Chromatography under pressure leads to improved and faster separation (4 hours). It also produces 5 to 10 percent higher yields at each step, resulting in significant increases in yields of the final product. Moreover, use of NH_4OH rather than NaOH in organic solvents for removing protecting groups also increases the yield and purity. The three decamers travel as a single band on 20 percent polyacrylamide–7*M* urea gel electrophoresis, with the same R_p value as bromphenol blue (Fig. 3).

Because the DNA strands synthesized are self-complementary, they form the duplexes shown in Fig. 1. These duplexes are relatively stable, having a T_m of about 48°C in 0.3*M* NaCl, 0.1 m*M* EDTA, 10 m*M* tris-HCl, *p*H 7.5, at a DNA concentration of 50 μg/ml. These duplexes are homogeneous and act as restriction enzyme substrates (Fig. 3). The Eco RI decamer is totally cleaved by its endonuclease to produce the labeled trimer as shown in Fig. 3, lanes 1 and 2. The Bam I decamer is also completely digested by its endonuclease to a labeled trimer and a second band running close to the [γ-^{32}P]ATP marker (lanes 3 to 6). Since this second band does not appear in the decamer without enzyme, it most likely can be ascribed to a nuclease contamination of the enzyme preparation. The Hind III decamer is cleaved to two identical five-base-pair pieces by Alu I, which recognizes adenine-guanine-cytosine-thymine (AGCT), but is not affected by Hsu I endonuclease, which recognizes AAGCTT (lanes 12 to 14). (Hsu I and Hind III recognize the same base sequence.) However, if the decamer is first polymerized with T4 DNA ligase under blunt-end joining conditions (legend of Fig. 3), then the polymer is cleaved by Hsu I endonuclease to pieces of approximately 30 base pairs in length, with no further degradation of these pieces (lanes 15 and 16). Thus it appears that Hsu I, unlike Eco RI, Bam I, or Alu I, requires a DNA strand of 30 base pairs or more as a substrate.

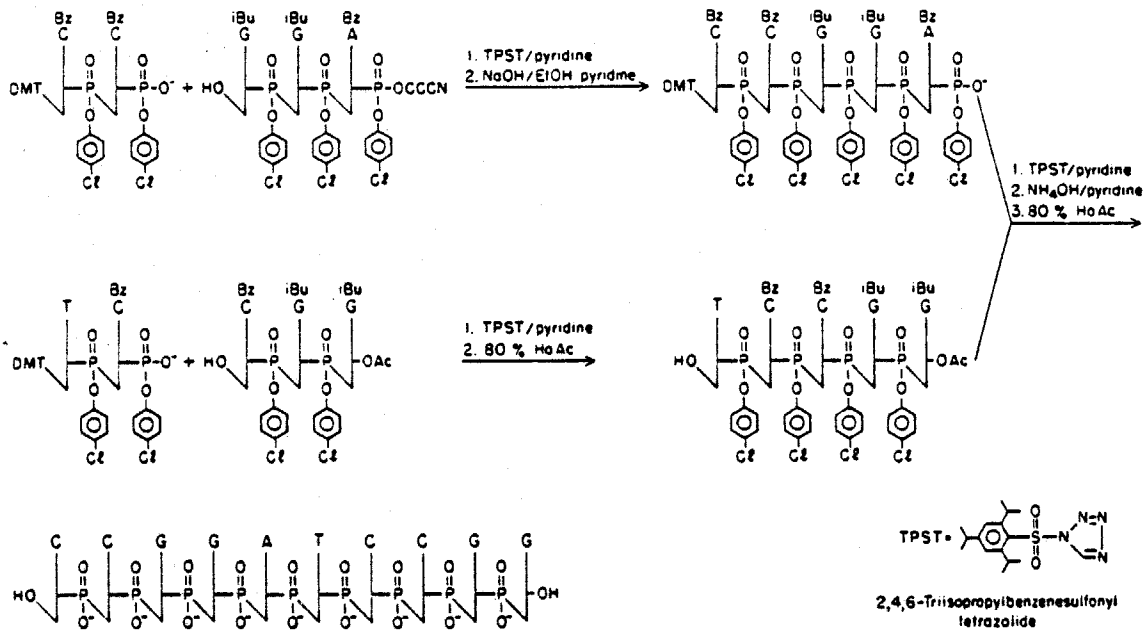
A comparable story is found with the Bam I decamer. It has two Hpa II sites per decamer (Fig. 1). In spite of this, Hpa II will not digest the decamer, apparently because the sites are too near the ends. However, if the decamers are polymerized with the T4 DNA ligase, then the polymer is quantitatively cleaved to smaller fragments by Hpa II endonuclease (lanes 8 to 10). The specific recognition by the Hpa II enzyme of the outer four base pairs of the Bam I decamer after polymerization is a confirmation of the sequence of the ends.

All three of these decameric linker molecules have CC at the 5' end and GG at the 3' end. Hence, blunt-end ligation produces a new Hae III restriction endonuclease site (GGCC). Polymers of the three linkers are cleaved back to decamers by Hae III enzyme (lane 19 and data not shown). Recognition of the central portion of each decamer by its proper endonuclease, and recognition of the outer two base pairs by Hae III, confirm the sequences as synthesized. The complete cleavage of the molecules as judged by the appearance of the gels allows one to assess the overall purity in terms of length and sequence homology as 90 percent or better.

For cloning DNA fragments, three methods for joining to the vector are in common use: (i) restriction enzyme "sticky ends" (*8*); (ii) tailing, either with polydeoxyadenylate [poly(dA)] or polydeoxythymidylate [poly(dT)] (9) or, more recently, with polydeoxyguanylate and polydeoxycytidylate [(*11*); and A. Otsuka, personal communication]; and (iii) blunt-end ligation by T4 ligase (*12*). This latter method is based on the ability of T4 ligase to rather efficiently join even or blunt-ended DNA duplexes under the appropriate conditions (*13*). All of these methods are useful, but none is completely general.

The three "linkers" described here and the procedure outlined in Fig. 1 should allow almost any DNA fragment to be cloned in an existing vector and also aid in the construction of new cloning vectors. Eco RI, Bam I, or Hind III cohesive ends can be added at will to any DNA bearing 5'-phosphate/3'-hydroxyl flush ends. If not already blunt-ended, most DNA's can be made so, either by repair with DNA polymerase [see (*12*)] or treatment with S1 nuclease (*14*).

Suitable plasmid vectors are known for insertions at all three restriction sites: Eco RI, Bam I, and Hind III. The plasmid pMB9, mentioned earlier, has been described (*4*). The plasmid RSF2124 has an Eco RI site within its colicin gene (*10*), and therefore one can screen for insertion by looking for disruption of colicin production after recombination. The plasmids pBR316 and pBR313 have both ampicillin and tetracycline resistance genes, with a Bam I and Hind III site in the latter



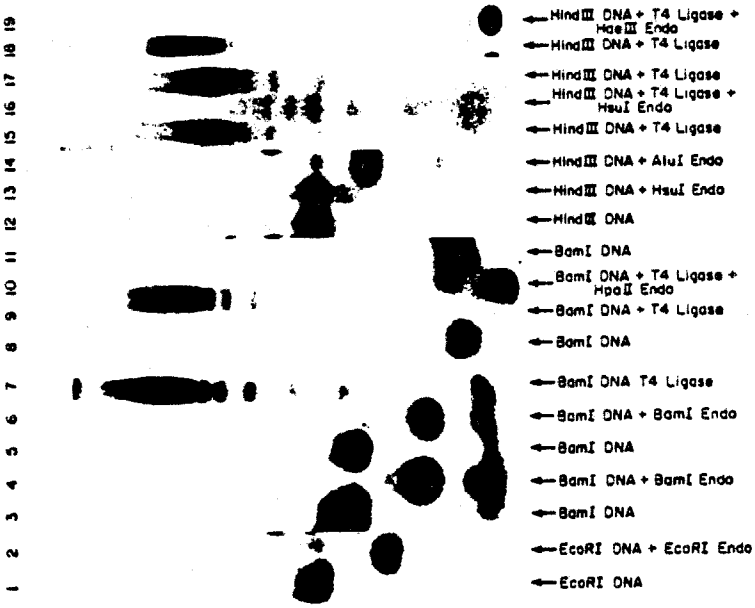


Fig. 2 (above). The convergent triester synthesis scheme used in preparing the Bam I decamer is shown. The other two decamers were made similarly. Most procedures, chemicals, and solvents have been described (6, 7); only new improvements are given here. Purification of the protected polynucleotide fragments was aided by the use of medium-pressure silica gel chromatography. Chromatronix glass columns with Fluid Metric RPG-50 solvent pumps were used to purify the intermediate and decameric products. The 2-inch columns were run at 25 pounds per square inch, 1-inch columns at 100 pounds per square inch, and ½-inch columns were run at 175 pounds per square inch in chloroform, with 5 to 10 percent methanol, depending on the polarity of the molecule being eluted. Removal of the β-cyanoethyl phosphate protecting groups at intermediate stages of the synthesis was carried out in a solution of 0.1*M* NaOH and tetrahydrofuran for preparation of Eco RI decamers, and in a solution of 0.05 *M* NaOH and a mixture of pyridine and ethanol (1 : 1) for the Bam I and Hind III decamers. All reactions were monitored on thin-layer silica gel plates (6). Final deblocking of the decameric products was done in a 1 : 2 mixture of pyridine and 30 percent NH_4OH at room temperature for 2 days, or at 70°C for 3 hours, followed by treatment with 80 percent acetic acid for 15 minutes. The deblocked products then were run on cellulose thin-layer chromatography plates in a mixture of 100 ml H_2O and 60 ml of isobutyric acid for 16 hours. The plates were dried, and the slowest moving band was recovered from the cellulose in 3 percent NH_4OH. Evaporation of solvent gave a white residue, which was dissolved in 70 m*M* phosphate buffer at *p*H 7, heated to 80°C, and then allowed to cool slowly at 4°C. The DNA duplex was then chromatographed on a Sephadex G-75 column (1 by 100 cm) at 4°C, eluted with 0.1*M* ammonium bicarbonate buffer, *p*H 7.5. Abbreviations: DMT, 4,4-dimethoxytrytyl protecting group for the 5′ end; CCCN, β-cyanoethyl protecting group for the 3′ phosphate; Ac, acetate protecting group for the hydroxyl end. For details about TPST, see (7). Fig. 3 (right). Accumulated electrophoretograms showing restriction enzyme reactions and blunt-end ligation. An explanation of each lane is in the text. The DNA's were phosphorylated and ^{32}P-labeled with T4 polynucleotide kinase (5), then treated as indicated. Conditions for T4 DNA ligase (blunt-end ligation) were 66 m*M* tris, *p*H 7.6, 6.6 m*M* $MgCl_2$, 10 m*M* dithiothreitol, 0.4 m*M* adenosine triphosphate (ATP), 1 μ*M* [5′-^{32}P]DNA, approximately 3.0 units of T4 ligase, 27°C, 16 hours, 15-μl volume. Conditions for Eco RI were 100 m*M* tris, *p*H 7.5, 5 m*M* $MgCl_2$, 100 m*M* NaCl, 0.02 percent Triton X100, 1 μg of [5′-^{32}P]DNA, 20°C, 5 hours, 10-μl volume. Conditions for Bam I were 100 m*M* tris, *p*H 7.5, 5 m*M* $MgCl_2$, 1 μg of [5′-^{32}P]DNA, 27°C, 5 hours, 10-μl volume. Conditions for Hsu I, Hpa II, and Hae III were T4 DNA ligase buffer, 1 μ*M* T4 ligated [5′-^{32}P]DNA, 37°C, 4 hours. Alu I: 6.0 m*M* tris, *p*H 8.0, 6.0 m*M* $MgCl_2$, 6.0 m*M* β-mercaptoethanol, 1 μg of [5′-^{32}P]DNA, 27°C, 5 hours, 10-μl volume. The enzymes Eco RI, Bam I, Hsu I, Alu I, Hae III, T4 DNA ligase, and T4 polynucleotide kinase were the gifts of Drs. H. Heyneker and P. J. Greene of the University of California, San Francisco. Hpa II was purchased from New England Bio-Labs. Gel electrophoresis was conducted in 20 percent acrylamide with 1.5 percent bis, 7*M* urea, at 15 volt/cm for various times in 0.05*M* tris-borate, *p*H 8.3.

(R. Rodriguez, F. Bolivar, H. Goodman, H. Boyer, and M. Betlach, personal communication). Hence, incorporation of DNA into a Bam I site can be monitored by looking for ampicillin resistance and tetracycline sensitivity in these two plasmid strains.

The linkers reported here have been used to clone eukaryotic DNA and complementary DNA (cDNA). As described (*15*), the Eco I decamer has been ligated to repetitive DNA from sea urchin sperm, and the resultant cohesive-ended fragments were cloned in RSF2124 plasmids. These clones will provide a source of repetitive DNA in quantity, helping to determine new features of the eukaryotic genome. The Bam I decamer has been added to cDNA's and cloned to the tetracycline gene of the pB313 plasmid (J. Shine and H. Goodman, personal communication). The decamers have also been shown to be useful for constructing new cloning vehicles by converting one restriction site to another, for example, Hind III to Eco RI (H. Heyneker, personal communication), and Eco RI to Bam I (A. Riggs, unpublished data). Therefore, the chemically synthesized decameric linkers reported here are generally applicable tools for molecular cloning experiments.

The three decameric linkers themselves were constructed with CC . . . . . GG at the ends for stability (T_m = 48°C), resulting in efficient blunt-end ligation, and so that polymerized decamers could be cloned and cut apart again by Hae III restriction enzyme. Cloning would provide a means of obtaining a steady supply of more linkers. However, "polylinker" clones have not yet been obtained, perhaps because of in vivo instabilities. Therefore, triester chemical synthesis may prove easier than cloning for moderate scale production of DNA molecules as small as decamers. At present, 25-mg amounts of greater than 90 percent pure material can be made in a relatively short time.

Note added in proof: The linker synthesis has also been pursued in parallel with this work by Bahl *et al.* (*16*).

RICHARD H. SCHELLER
RICHARD E. DICKERSON
Norman W. Church Laboratory of Chemical Biology, California Institute of Technology, Pasadena 91125
HERBERT W. BOYER
Department of Biochemistry and Biophysics, University of California Medical Center, San Francisco 94143
ARTHUR D. RIGGS
KEIICHI ITAKURA
City of Hope National Medical Center, Duarte, California 91010

References and Notes

1. J. Hedgpeth, H. M. Goodman, H. W. Boyer, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 3448 (1972).
2. The Eco RI decamer (CCGAATTCGG) is recognized by Eco RI and Eco RI*. The Bam I decamer (CCGGATCCGG) contains the specific sequences for Bam I, Bst I, Bsu NI, Dpn I, Mbo I, Mos I, Hpa II, Hap II, and Mno I. The Hind III decamer (CCAAGCTTGG) contains sequences for Alu I, Bbr I, Chu I, Hinb III, Hind III, and Hsu I. After polymerization, each decamer will contain the sequence GGCC recognized by Hae III, Blu I, Bsu R, Hhg I, Pal I, and Sfa I. For a review of restriction enzymes and their recognition sequences, see R. Roberts, *Critical Rev. Biochem.*, in press.
3. R. Old, K. Murry, G. Roizes, *J. Mol. Biol.* **92**, 331 (1975).
4. R. L. Rodriguez, F. Bolivar, H. M. Goodman, H. W. Boyer, M. Betlach, in *ICH-UCLA Symposia on Molecular and Cellular Biology*, P. Nierlich and W. J. Rutter, Eds. (Academic Press, New York, 1976), vol. 5, pp. 471–477.
5. H. L. Heyneker, J. Shine, H. M. Goodman, H. Boyer, J. Rosenberg, R. E. Dickerson, S. A. Narang, K. Itakura, S. Lin, A. D. Riggs, *Nature (London)* **263**, 748 (1976).
6. K. Itakura, N. Katagiri, C. P. Bahl, R. H. Wightman, S. A. Narang, *J. Am. Chem. Soc.* **97**, 7327 (1975).
7. J. Stawinski, T. Hayumi, S. A. Narang, *Can. J. Chem.* **54**, 670 (1976).
8. A. C. Y. Chang and S. H. Cohen, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 1030 (1974).
9. D. A. Jackson, R. H. Symons, P. Berg, *ibid.* **69**, 2904 (1972).
10. M. So, R. Gill, H. S. Falkow, *Mol. Gen. Genet.* **142**, 239 (1975).
11. F. Rougeon, P. Kourilsky, B. Mach, *Nucleic Acids Res.* **2**, 2365 (1975).
12. K. Backman, M. Ptashne, W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 4174 (1976).
13. V. Sgaramella, J. H. Van de Sande, H. G. Khorana, *ibid.*, p. 1468.
14. G. S. Ghangas and R. Wu, *J. Biol. Chem.* **250**, 4601 (1975).
15. R. Scheller, T. Thomas, A. Lee, W. Niles, W. Klein, R. Britten, E. Davidson, *Science* **196**, 197 (1977).
16. C. P. Bahl, K. J. Marians, R. Wu, J. Stawinski, S. A. Narang, *Gene* **1**, 81 (1977).
17. We thank J. Rosenberg and W. Gilbert for their stimulating contributions to the development of the linker concept. We also thank A. S. Lee, H. Heyneker, and J. Shine for help in demonstrating some of the enzyme reactions discussed in this report; and L. Shively for technical help. Supported by NIH grants GM-12121 and HD-04420, NSF grants PCM75-05886 and GB-26517, and an NIH predoctoral traineeship to R.H.S. This is contribution No. 5457 from the Norman W. Church Laboratory of Chemical Biology, California Institute of Technology.

4 February 1977

CHAPTER 3

Reprinted from SCIENCE, 8 April 1977, Volume 196, pp. 197-200

Clones of Individual Repetitive Sequences from Sea Urchin DNA Constructed with Synthetic Eco RI Sites

Abstract. *Interspersed repetitive sequences were isolated from sea urchin DNA by renaturing to low C_0t followed by treatment with nuclease S1. Synthetic Eco RI sites were ligated onto the repetitive sequence elements, which were then inserted at the Eco RI site of plasmid RSF2124 and cloned. The repetitive sequences can be excised from the plasmid with Eco RI for further study.*

To facilitate studies of individual repetitive DNA sequence families, we have constructed recombinant DNA plasmids containing repetitive DNA from *Strongylocentrotus purpuratus* (*1*). Our object was to obtain cloned sequences terminated at the ends of interspersed repetitive sequence elements (*2*), rather than where restriction enzyme sites happen to fall. This was accomplished through the use of synthetically prepared Eco RI restriction sites that were ligated to the repetitive DNA fragments. The product of the ligation reactions was cloned to provide a source of plasmids from which individual repetitive DNA sequences could be reisolated in relatively large amounts.

Repetitive DNA duplex was prepared as follows (details may be found in the legend to Fig. 1). Sea urchin DNA sheared to a single-strand weight mean length of about 2000 nucleotides was renatured to C_0t 40 (C_0 is the initial concentration of nucleotides in moles per liter, and t is the time in seconds) and then treated with single-strand specific S1 nuclease. Previous studies (*2*, *3*) have shown that by this point most repetitive sequences in the sea urchin genome are fully renatured. The nuclease digestion removes the nonrepetitive regions flanking the interspersed repetitive sequences, since these remain single-stranded at C_0t 40 (*2*). The resistant DNA sequence consists almost entirely of repetitive duplexes. We have shown earlier that a majority of the S1 nuclease resistant duplexes are 300 to 400 nucleotides in length, and that this is the characteristic length of interspersed repeats in the sea urchin genome (*2*, *3*). The resistant repetitive duplexes are expected to have 3'-hydroxyl and 5'-phosphoryl termini (*4*).

A symmetrical decamer containing the Eco RI restriction site was synthesized by a triester chemical synthesis method (*5*). The structure of this decamer is shown in Fig. 1. The 5'-hydroxyl termini of the Eco RI decamers were labeled with ^{32}P, using T4 polynucleotide kinase. The decamer was covalently linked to the repetitive sea urchin DNA duplexes by blunt-end ligation with T4 DNA ligase (*6*). The expected products of the blunt-end ligation are the repetitive sea urchin DNA fragments bearing covalently linked Eco RI decamers, plus polymerized Eco RI decamers. The ligated mixture was next digested with Eco RI endonuclease. Eco RI treatment should yield the repeti-

tive sea urchin DNA with covalently linked Eco RI sticky ends labeled with ^{32}P, plus Eco RI cleavage products of polymerized and monomeric decamer. Sea urchin DNA containing covalently attached ^{32}P-labeled Eco RI cohesive ends was separated from the other cleavage products by gel filtration. From the amount of ^{32}P associated with the sea urchin DNA in the exclusion peak, the specific activity of the Eco RI decamer, and the DNA mass recovered, we calculated that 30 percent of the ends of the sea urchin DNA molecules were covalently linked to an Eco RI site. Thus about 10 percent of the repetitive DNA duplexes could be expected to carry Eco RI sites on both ends. This preparation was reacted with Eco RI–cleaved DNA isolated from the plasmid RSF2124, with T4 DNA ligase included to seal the duplexes formed from the Eco RI cohesive ends.

The procedures used to prepare the repetitive DNA and to construct the clones is illustrated in Fig. 1.

RSF2124 is a colicinogenic (E1) plasmid carrying a locus for ampicillin resistance (*amp*^{R}) and a single Eco RI site (*7*). Transfection of *Escherichia coli* strain C600 with the recombinant plasmids was accomplished essentially according to Cohen *et al.* (*8*). Transformation efficiency was from 10^{-7} to 10^{-6} Amp^R colonies per molecule of plasmid DNA. The single Eco RI site in the RSF2124 genome lies within the colicin structural gene or in a control region for colicin E1 production (*7*). Insertion of a segment of DNA into the plasmid genome with the use of the Eco RI cohesive ends thus results in dysfunction of the colicin E1 locus. This property was exploited to assay the Amp^R colonies for insertion of the sea urchin DNA fragments into the plasmid genome. A double-layer technique (*9*) was used to detect colicin-producing clones, as described in the legend to Fig. 1. The frequency of Amp^R clones which were *col*$^-$ was about 10 percent.

Insertion of the Eco RI–sea urchin DNA fragments in the RSF2124 genome should introduce a second Eco RI site into the hybrid plasmid genome. Eco RI cleavage of hybrid plasmid DNA should thus result in unit length linear plasmid DNA molecules plus repetitive DNA fragments. We examined a number of *amp*^{R}*col*$^-$ clones and found that more than 65 percent contain short DNA duplexes which can be excised from the plasmid genome with Eco RI. We consider this a minimum estimate because of occasional difficulties encountered in detecting small inserted fragments representing 3 percent or less of the DNA mass. In other experiments in which long-

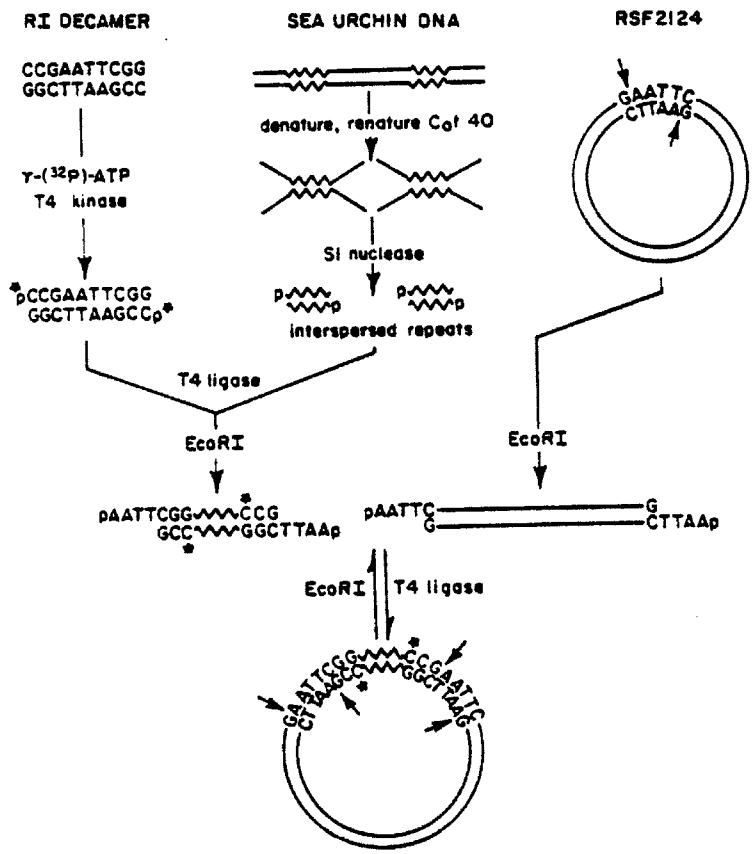


Fig. 1. Enzymatic steps in the construction of recombinant plasmids containing repetitive sea urchin DNA. Repetitive sea urchin DNA (double wavy lines), Eco RI decamer (shown in its entirety), and RSF2124 DNA (double straight lines) were ligated and cloned as follows. The vertical arrows (↑↓) indicate sites of Eco RI cleavage. The star (★) indicates the ^{32}P label at the 5′ termini. Repetitive sea urchin DNA was isolated by renaturing 2000-nucleotide long sperm DNA to C_0t 40, followed by digestion of the single-strand tails with 200 μl of S1 nuclease per milligram of DNA at 37°C for 45 minutes (*3*). Ten nanomoles of Eco RI decamer were 5′ phosphorylated with 7 units of T4 polynucleotide kinase and 50 nanomoles of [γ-^{32}P]ATP (adenosine triphosphate) for 1 hour at 37°C in 0.05*M* tris-HCl (*p*H 9.0), 0.01*M* $MgCl_2$, 0.005*M* dithiothreitol (DTT). A 60-fold excess of ^{32}P-labeled Eco RI decamer (16 μg) was ligated for 17 hours at room temperature to the repetitive DNA fragments (8 μg) with 10 μl of T4 DNA ligase in 0.04 ml, 66 m*M* tris-HCl (*p*H 7.6), 6.6 m*M* $MgCl_2$, 10 m*M* DTT, 0.01 percent NP40, and 0.4 m*M* ATP. This reaction mixture was adjusted with 100 m*M* tris-HCl (*p*H 7.6), 50 m*M* NaCl, 10 m*M* $MgCl_2$, and 0.02 percent NP40 and digested with 15 μl Eco RI endonuclease for 3 hours at 37°C. The Eco RI digest was deproteinized with an equal volume of $CHCl_3$ and isoamyl alcohol (24 : 1). The cleavage products were resolved on a Sephadex G-200 column. Supercoil RSF2124 plasmid DNA, cleaved with Eco RI essentially as described above, was mixed with a 40-fold molar excess of G-200–excluded sea urchin DNA in 66 m*M* tris-HCl (*p*H 7.6), 20 m*M* NaCl, 1 m*M* EDTA, 10 m*M* $MgCl_2$, 10 m*M* DTT, and 66 μ*M* ATP in a total volume of 30 μl. One unit of T4 DNA ligase (Miles Laboratories) was added, and the reaction mixtures were incubated at 15°C for 18 hours. Five microliters of this reaction mixture, diluted 200-fold with 0.05*M* $CaCl_2$, was used to transform *E. coli* strain C600 (RecA$^+$r$^-$m$^-$) to ampicillin resistance essentially according to Cohen *et al.* (*8*). Transformed cells were plated on agar plates containing ampicillin (15 μg/ml). Amp^R colonies were assayed for colicin production by a double-overlay method (*9*). Colonies were transferred to duplicate plates and grown overnight. One plate was retained as a master, and the cells on the second plate were killed with $CHCl_3$. The latter plates were overlaid with top agar containing 1×10^8 colicin-sensitive bacteria (*E. coli* HB101). The plates were incubated overnight and then scored for the absence of colicin production (*col*$^-$). Col$^+$ colonies are surrounded by a clear halo, indicating lack of bacterial growth, while col$^-$ colonies are overgrown by the colicin-sensitive cells. Col$^-$ colonies were stored in stab cultures and assayed for the presence of an inserted sea urchin fragment.

er sea urchin DNA fragments derived from an Eco RI digest of sea urchin DNA were inserted into the same plasmid vector, about 80 percent of the *amp*^{R}*col*$^-$ clones were found to contain sea urchin DNA.

The lengths of the inserted repetitive DNA fragments were estimated by gel electrophoresis, with the use of an Hae III digest of the parent plasmid as a size standard. The lengths of the inserted DNA sequences from 20 clones have now been measured. Of these, 17 are less than 1000 nucleotides. The mode value of their length distribution is 300 to 400 base pairs. Two inserts are in the range of 1000 to 1200 nucleotides, and one is about 2000 nucleotides. Although the sample is small, this distribution is in excellent agreement with the distribution of fragment lengths in the original nuclease S1 digest and with the range of repetitive sequence lengths in the sea urchin genome. More than 90 percent of the repetitive sequence elements in this genome are 300 to 400 nucleotides long, with less than a few percent ≥ 2000 nucleotides. Our data thus suggest that the population of repetitive sea urchin DNA duplexes inserted into the RSF2124 vector is reasonably representative of the repetitive DNA sequences in the sea urchin genome.

To verify that the DNA sequences inserted into the plasmid genome in fact consist of repetitive sea urchin DNA, the cloned inserts were reassociated with sheared sea urchin DNA. The following procedure was used. The cloned DNA fragments were excised with Eco RI and labeled at the 5′ termini with ^{32}P by the T4 kinase method (*10*) (Fig. 2). The DNA was renatured after labeling and placed on 3 percent agarose gels along with an Hae III digest of RSF2124. An example is shown in Fig. 2A, which illustrates the preparation of the repetitive DNA insert from clone CS2108. After renaturation, the labeled DNA insert from this clone migrates to a position corresponding to a 190 base pair fragment, where it can be detected either by ethidium bromide staining or by autoradiography. The labeled DNA was eluted from the gel and reassociated with excess sheared sea urchin DNA approximately 500 nucleotides in length. The fraction of DNA reassociated was determined by hydroxyapatite chromatography and is plotted as a function of the driver DNA C_0t in Fig. 2B. The self-reaction of the labeled CS2108 fragment was monitored in separate reactions (not shown). At the tracer concentration used, the rate of the self-reaction was less than 10 percent of the driver reaction and thus does not affect the results. A least-squares solution to the kinetic data reproduced in Fig. 2B yields a rate constant of $3.16 \times 10^{-2} M^{-1}$ sec^{-1} for the reassociation of the CS2108 fragment with sea urchin DNA. The rate constant for the reassociation of single-copy sea urchin DNA of similar fragment length with whole DNA is $1.13 \times 10^{-3} M^{-1}$ sec^{-1}. Thus clone CS2108 contains a DNA sequence which is repeated approximately 25 times per haploid genome. It is unlikely that the reiteration frequency of the CS2108 sequence could differ from this number by more than a factor of two or three, taking into account the limited kinetic data and uncertainties in correcting the observed rate constants for minor length differences and the effect of base pair mismatch.

The repetitive sea urchin DNA clones constructed by the blunt-end ligation

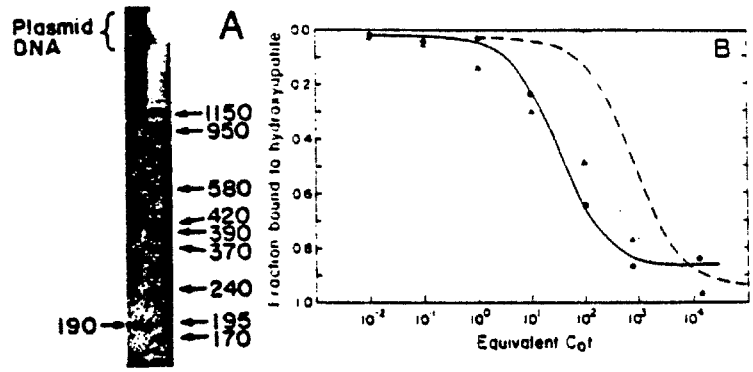


Fig. 2. Analysis of a recombinant plasmid clone constructed by blunt-end ligation. (A) Detection and end labeling of inserted sea urchin fragment. *Amp*^{R}*col*$^-$ clone CS2108 was grown in 40 ml of L broth plus ampicillin (15 μg/ml) to 4×10^8 to 6×10^8 cells per milliliter, and the plasmid content was amplified by incubation overnight in chloramphenicol (200 μg/ml). The cells were harvested by centrifugation and washed with a solution of 10 m*M* tris-HCl (*p*H 7.4) and 1 m*M* EDTA. They were again centrifuged and resuspended in 3 ml of 25 percent sucrose, 50 m*M* tris-HCl (*p*H 8.0), and 40 m*M* EDTA at 0°C. Lysozyme (5 mg; Sigma) was added, and the solution was maintained at 0°C for 5 minutes. Then 0.5 ml of 0.5*M* EDTA and 0.375*M* EGTA (*p*H 8.0) was added for 5 minutes at 0°C. Four ml of Brij lysing buffer [1 percent Brij 58, 0.4 percent sodium deoxycholate, 0.063*M* EDTA, 0.05*M* tris-HCl (*p*H 8.0)] was added and mixed vigorously. Chromosomal DNA was removed from the lysate by centrifugation at 50,000 rev/min for 45 minutes (Beckman/Spinco 50 Ti rotor). Superhelical plasmid DNA was isolated from the cleared supernatant by CsCl–ethidium bromide density gradient centrifugation (ethidium bromide at 250 μg/ml; $\rho_0 = 1.669$ g/cm^3) in a 50 Ti rotor at 39,000 rev/min. Superhelical DNA was removed from the gradient by side puncture. Ethidium bromide was removed by extraction with isopropanol. Plasmid DNA was dialyzed against 100 m*M* tris-HCl (*p*H 7.5), 100 m*M* NaCl, 5 m*M* $MgCl_2$ at 4°C. Recombinant plasmid DNA (15 μg) was digested with 2 μl of Eco RI at 37°C for 30 minutes. This digest was concentrated by flash evaporation to 0.5 ml and passed over a 0.5-ml Chelex 100 (BioRad) column. The DNA was precipitated at −70°C for 2 hours with 1 volume of isopropanol. The precipitate was collected by centrifugation at 10,000*g* for 30 minutes. The dried precipitate was resuspended in 10 m*M* tris-HCl (*p*H 8.0) and reacted with 1 μl of bacterial alkaline phosphatase (Worthington) at 37°C for 30 minutes. Ten micrograms of superhelical plasmid DNA carrier and 1 volume of 0.3*M* sodium acetate (*p*H 6.8) was added. This solution was extracted twice with a mixture of 80 percent phenol (*p*H 8.0) and 20 percent chloroform: isoamyl alcohol (24 : 1), and once with ether, then precipitated as described above. Labeling at 5′ termini was done according to Maxam and Gilbert (*10*) with [γ-^{32}P]ATP, specific activity 1200 c/mmole (Winston Salser's laboratory, University of California, Los Angeles). Kinase reactions were deproteinized as described above, and precipitated with 2.5 volumes of 100 percent ethanol at −70°C as described above. The dried precipitate was resuspended in a volume of 200 μl of 0.1*M* NaCl, 0.05*M* tris-HCl (*p*H 7.2). The concentration of the end-labeled CS2108 DNA fragment was determined by its rate of renaturation, and from this the specific activity was calculated to be about 1×10^6 to 2×10^6 count/min per microgram. A portion of the labeled DNA was heat denatured and renatured at 60°C overnight. The renatured DNA fragment was loaded on a 3 percent agarose gel and subjected to electrophoresis (gel 1) along with an Hae III digest of RSF2124 DNA (gel 2) at 60 volts for 3.5 hours. Gels were stained with ethidium bromide (1 μg/ml). Lengths of the Hae III fragments of RSF2124 are indicated in the figure. (B) Reassociation of end-labeled Eco RI fragment from a blunt-end repeat clone with excess sea urchin DNA. The 190 base pair end-labeled CS2108 fragment from (A) was eluted from the gel by dissolving in 5*M* $NaClO_4$ and binding the DNA to hydroxyapatite in 0.12*M* phosphate buffer (PB), followed by thermal elution of the ^{32}P-labeled DNA. This DNA was reassociated with excess 500 nucleotide long (5×10^6 mass ratio) sheared sea urchin DNA in 0.12*M* PB at 55°C. The fraction ^{32}P-labeled DNA (●) and driver DNA (▲) reassociated was determined by hydroxyapatite chromatography in 0.12*M* PB at 55°C. The solid line is the least-squares solution for the reaction of ^{32}P-labeled DNA driven by total sea urchin DNA. The rate constant determined for this reaction is $3.16 \times 10^{-2} M^{-1}$ sec^{-1}. The dashed line illustrates the reassociation of nonrepetitive sea urchin DNA with whole DNA (data not shown). The rate constant for this reaction is $1.13 \times 10^{-3} M^{-1}$ sec^{-1}. The dotted line shows the renaturation of total sea urchin DNA (*2*).

method described here provide a source from which we are able to obtain relatively large quantities of individual repetitive DNA sequences. Such sequences can be used to isolate all the related members of given repetitive sequence families from the genome. The method described can, of course, be used for cloning any DNA sequence without the requirement that it be terminated by specific restriction enzyme sites (6), and without the addition of substantial homopolymer sequences to the cloned fragment.

RICHARD H. SCHELLER

Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena 91125

TERRY L. THOMAS, AMY S. LEE
WILLIAM H. KLEIN, WALTER D. NILES
ROY J. BRITTEN, ERIC H. DAVIDSON

Division of Biology, California Institute of Technology, Pasadena 91125

References and Notes

1. This work was done in compliance with the NIH Guidelines for Recombinant DNA Research.
2. D. E. Graham, B. R. Neufeld, E. H. Davidson, R. J. Britten, *Cell* **1**, 127 (1974).
3. F. C. Eden, D. E. Graham, E. H. Davidson, R. J. Britten, *Nucleic Acids Res.*, in press.
4. T. Ando, *Biochim. Biophys. Acta* **114**, 158 (1966).
5. R. H. Scheller, R. E. Dickerson, H. W. Boyer, A. D. Riggs, K. Itakura, *Science* **196**, 177 (1977).
6. V. Sgaramella, J. H. Van de Sande, H. G. Khorana, *Proc. Natl. Acad. Sci. U.S.A.* **67**, 1468 (1970); H. L. Heyneker, J. Shine, H. M. Goodman, H. W. Boyer, J. Rosenberg, R. E. Dickerson, S. A. Narang, K. Itakura, S-Y. Lin, A. D. Riggs, *Nature (London)* **263**, 748 (1976); C. P. Bohi *et al.*, *Gene* **1**, 81 (1977).
7. M. So, R. Gill, S. Falkow, *Mol. Gen. Genet.* **142**, 239 (1975).
8. S. N. Cohen, C. Y. Chang, L. Hsu, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 2110 (1972).
9. P. Fredericq, *Annu. Rev. Microbiol.* **11**, 7 (1957).
10. A. M. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560 (1977).
11. T4 polynucleotide kinase, T4 DNA ligase, and Eco RI endonuclease were gifts of P. Green, H. Heyneker, and H. Boyer. Supported by NIH grants HD-03753 and GM-20927 and by NSF grant BMS 75-07359. R.H.S. is supported by an NIH predoctoral training grant; T.L.T. by an NIH postdoctoral fellowship; A.S.L. by fellowship J-289 and W.H.K. by Lievre fellowship J-340, both from the American Cancer Society, California Division.

7 February 1977

CHAPTER 4

Cell, Vol. 14, 889–900, August 1978.

Characteristics of Individual Repetitive Sequence Families in the Sea Urchin Genome Studied with Cloned Repeats

William H. Klein, Terry L. Thomas, Cary Lai,* Richard H. Scheller,† Roy J. Britten and Eric H. Davidson
Division of Biology
California Institute of Technology
Pasadena, California 91125

Summary

Cloned repetitive sequences from the S. purpuratus genome a few hundred to approximately 1000 nucleotides long were used to investigate the characteristics of individual repetitive sequence families. They were terminally labeled by the kinase procedure and reacted with sheared S. purpuratus DNA. Repetition frequencies were measured for 26 individual families and were found to vary from a few to several thousand copies per genome. Estimates of sequence divergence were made for 18 cloned repeat families by measuring thermal stability of the heteroduplexes formed between the genomic DNA and the cloned fragments, compared with that of the renatured cloned fragments. The difference was <4°C for three of the 18 families, and <10°C for 13 of the 18 families. These 13 repetitive sequence families lack any detectable highly divergent sequence relatives, and the results reported are shown not to change when the renaturation criterion is lowered below 55°C in 0.18 M Na^+. Five of the 18 cloned families displayed greater sequence divergence. The average sequence divergence of the total short repetitive sequence fraction of S. purpuratus DNA was found to match closely the average of the divergences of the cloned repeat sequences.

Introduction

The renaturation kinetics of animal DNAs show that a large variety of different repetitive sequences exist within each genome. It follows that each genome includes many individual repetitive sequence families. The repetitive sequence family is defined empirically as that set of repeat sequence elements sufficiently homologous to renature and form stable base-paired duplexes. In application, this simple definition proves to be far from elementary. It has been known since the initial studies of Britten and Kohne (1967, 1968) that the sequences of homologous repeats are often not identical. Thus when repetitive sequences are renatured, the thermal stability of the reaction products is typically lower than expected for native DNA duplexes of equal length. The question arises whether it is correct to count repeat families as discrete sets of repetitive sequence elements each containing a certain number of members, and each clearly delimited from single-copy sequence. Suppose, for example, that additional, more distantly related family members were continuously included as the renaturation criterion is lowered to the level where base pair specificity is lost. In this case, the true repetition frequency and the amount of sequence divergence within repeat families, as well as the number of such families in a given genome, could be defined only in the relative terms of an arbitrary set of renaturation conditions. Fortunately, studies on various repetitive sequence fractions indicate that very highly divergent repeat families are not a dominant feature of most animal genomes, although such families certainly exist (Weinblum et al., 1973; Mizuno and Macgregor, 1974; for a review of earlier investigations see, for example, Davidson and Britten, 1973).

The genome of the sea urchin Strongiyocentrotus purpuratus provides a typical example of repeat sequence divergence. Graham et al. (1974) showed that the low Cot renaturation kinetics of S. purpuratus DNA measured at 50°C, 0.21 M Na^+ are barely distinguishable from the kinetics obtained at 60°C in 0.18 M Na^+. Furthermore, the mean thermal stability of renatured sea urchin repetitive DNA has been reported in previous work to be at least 73–75°C (0.18 M Na^+), well above the standard reaction criterion. Some repeat fractions display an even higher thermal stability which approaches that of renatured single-copy DNA (Graham et al., 1974; Britten et al., 1976; Eden et al., 1977). Sea urchin DNA appears to contain several thousand different families of repetitive sequences, as do most genomes which have been examined. The fractionation methods so far applied do not separate individual repetitive sequence families. It has been possible to measure quantitatively only the average characteristics of large classes of repetitive sequences, and our knowledge has consequently remained imprecise.

The availability of cloned repetitive sequences now makes it possible to investigate the nature of individual repeat families. The set of genomic DNA sequences which react with a given cloned repeat provides an objective definition of its repetitive sequence family. The object of the experiments described in this paper is to measure the number of members and sequence divergence of a series of individual repetitive sequence families.

* Present address: Department of Biology, University of California, San Diego, La Jolla, California 92093.
† Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125.

Results

Cloned Repetitive Sequences

Construction of the recombinant DNAs used for these experiments was described in detail by Scheller et al. (1977). The repetitive sequence clones were assembled from S. purpuratus DNA fragments as follows. The DNA was sheared to ~2000 nucleotides and renatured to Cot 40. The reaction mixture was then digested with S1 nuclease. The conditions used were such as to spare most of the renatured repetitive sequence duplex which can be detected by hydroxyapatite binding or by optical methods, even though this duplex fraction includes many internal mismatch sites (Chamberlin, Britten and Davidson, 1975; Britten et al., 1976). It is possible that the S1 nuclease-resistant fraction could lack severely mismatched base pairs, if they exist. It is also possible that some of the nuclease-resistant repeat duplexes terminate within the original sequence, due to strand scissions, rather than at the natural boundaries between the repetitive sequence and the flanking single-copy regions. Some of the cloned repeats could have terminated artificially at naturally occurring internal Eco RI sites, although this would be an infrequent occurrence. Finally, it should be noted that the cloning procedure itself could have selected a bizarre subfraction of the S1 nuclease-resistant duplex fragments, although there is no evidence that it did so.

To amplify the repetitive sequence fragments, synthetically constructed double-stranded DNA decamers including the Eco RI endonuclease recognition site were ligated to the termini of the repeat duplexes (Scheller et al., 1977, 1978). This preparation was treated with Eco RI endonuclease to free the sticky ends within the decamer, and cloned by ligation to the complementary Eco RI sites of the plasmid vector RSF2124 (So et al., 1975). An important point is that since the original renatured repetitive sequence duplexes included mismatched base pairs, the exact cloned sequences derived from them may never have existed in the sea urchin genome—that is, repair processes in the bacterial hosts would be expected to have generated daughter sequences which are most likely to represent a compromise between the two parental sequences. A repaired sequence of this nature, however, cannot be more divergent compared with all other sequences of its repeat family than were the parental sequences considered together. The cloned repeat fragments are therefore appropriate for use as probes for that repetitive sequence family in the genome to which the parental sequences belonged. Cloned repeat fragments were separated from plasmid vector DNA by Eco RI endonuclease cleavage followed by gel electrophoresis.

Lengths were determined for 26 repeat fragments excised from clones picked at random from plates. The data are listed in Table 1. The length distribution is close to that expected from the known sequence organization of sea urchin DNA. Twenty three of the 26 cloned sequences are only a few hundred nucleotides long, and the weight mean of these is ~300 nucleotides. Graham et al. (1974) and Eden et al. (1977) showed that the short interspersed repeats of the sea urchin genome average ~300 nucleotides. The three longer cloned repeats are ~1000 nucleotides in length. The genome of S. purpuratus is known to include approximately 30–40% of the repetitive sequence mass in S1 nuclease-resistant "long repeats." Clones 2097, 2100 and 2007 are apparently of this class. Together they constitute 34% of the sequence length of the 26 clones studied. These cloned sequences could have been terminated artificially, either by S1 nuclease cleavage or by the presence of an internal Eco RI site. Thus we do not know whether the genomic sequence elements from which the long clones were derived originally exceeded 1000 nucleotides. Some of the shorter clones could also have been derived from genomic sequences which were much longer than the cloned sequences. Such changes in length are probably not frequent, since the distribution of cloned insert lengths is consistent with the length distribution of S1 nuclease-treated repetitive DNA from the S. purpuratus genome.

Reiteration Frequency of the Cloned Repetitive Sequence Families

The number of occurrences in the genome of sequences homologous to each cloned repeat was next estimated from the kinetics of reactions between the cloned repeats, labeled with ^{32}P at the 5′ termini, and the genomic DNA. This number is taken to be the repeat family size, in those cases for which lowering the reaction criterion does not produce additional reaction. As shown in the following section, for five of the 26 determinations listed in Table 1, the number of family members is probably greater than our measurements indicate, since these five families include highly divergent relatives whose reaction was inhibited under the renaturation conditions used for the kinetic measurements. For seven other clones (see footnote c to Table 1), the result of lowering the criterion was not investigated, although the results presented below show that in about threefourths of these cases, no effect is to be expected. For the remaining 14 clones, additional family members cannot be detected when the reactions are carried out at a criterion lower than that used for the kinetic meas-

Table 1. Length of Cloned Fragments and Reiteration Frequency of Their Repetitive Sequence Families

Clone	Length (Nucleotide Pairs)	Rate Constant[a] (M^{-1} sec^{-1})	Approximate Reiteration Frequency[b] (Copies per Genome)
2005	240	1.5 ± 0.11	2,100
2007	1100	0.49 ± 0.036	400
2034	560	1.2 ± 0.12	1,000
2057A	325	1.1 ± 0.13	900
2057B	290	0.11 ± 0.013	90
2065	165	0.15 ± 0.026	130[c]
2066	420	0.0041 ± 0.00034	3
2090	220	0.17 ± 0.027	140[c]
2096	360	0.0073 ± 0.00060	6
2097	990	0.044 ± 0.0047	40
2099	235	0.10 ± 0.021	80
2100	1000	3.7 ± 0.54	3,100
2101	320	0.83 ± 0.12	700
2103	220	0.63 ± 0.19	530
2104	386	2.1 ± 0.18	1,800
2108	190	0.025 ± 0.0034	20
2109A	200	1.1 ± 0.21	900[c]
2109B	125	0.24 ± 0.026	200[c]
2110	260	1.9 ± 0.23	1,600
2111	155	15 ± 2.1	12,500
2112	220	0.54 ± 0.087	450
2125	180	0.062 ± 0.0066	90[c]
2131	210	0.97 ± 0.11	1,400
2133A	310	1.5 ± 0.21	2,100
2133B	310	0.042 ± 0.013	60
2137	190	0.37 ± 0.069	530

[a] Second-order rate constants were obtained by a least-squares method (Pearson et al., 1977). The errors shown are standard deviations calculated on the basis of the scatter of the data. These errors do not take into account any systematic problems which could have affected the determinations. Factors such as the minor differences in single-copy DNA, driver DNA and cloned DNA fragment lengths, thermal stability of the reaction products, salt concentrations and other unknown variables can affect the absolute rate of the reactions (see Bonner et al., 1973). The family size estimates are probably accurate to within a factor of ≤2, taking into account the probable systematic errors. Generally, the fragment lengths of the cloned tracers are comparable to those of the driver, and thus no length corrections are made on the observed rate constants, except for the three cases where the cloned fragment was ≥1000 nucleotides. The correction in these cases was obtained as follows: $K_{cL} = K_c(L_o/L_c)$, where K_{cL} is the length-corrected rate constant, K_c is the observed rate constant, L_c is the length of the cloned fragment and L_o is the length of the sea urchin DNA driver (Chamberlin et al., 1978).

[b] Reiteration frequency, S, is calculated as $S = K_c/K_{sc}$, where K_{sc} is the second-order rate constant for the reaction of the ^3H-DNA single-copy tracer with the same sea urchin DNA driver. The mean length of the single-copy tracer, 320 nucleotides, was too similar to that of the driver to warrant a length correction. For all samples except clones 2005, 2137, 2133A, 2133B, 2131 and 2125, the single-copy rate constant measured was $\sim 1.2 \times 10^{-3}$ M^{-1} sec^{-1}. For the latter clones, the rate constant was 0.7×10^{-3} M^{-1} sec^{-1}, probably due to a minor difference in salt concentration.

[c] The reiteration frequency for these cloned sequences represents minimum values calculated from 55°C reactions. Thermal stability measurements presented below show that for these families, the T_ms are dependent upon the reaction criterion. Thus as the reaction temperature is lowered to 45°C, the family appears to expand by including more divergent members (see text). For example, when clone 2109B is reacted with sea urchin DNA driver at 45°C, the rate constant increases by approximately a factor of 5, implying a reiteration frequency of at least 1000 for this repetitive family.

urements (see the following section). In these 14 cases, the renaturation kinetics should provide a reasonably complete estimate of copy number. The kinetics of some typical reactions are reproduced in Figure 1. The reactions were assayed by hydroxyapatite binding, and as expected, their kinetics closely approach second-order form. Reiteration frequencies for the 26 repeat families studied were calculated from the rate constants listed in Table 1.

The number of members in the families included

Table 2. Reiteration Frequencies for Sea Urchin Repetitive DNA Fractions and for the 26 Family Cloned Sample

Repetitive DNA[a]	Component	Rate Constant ($M^{-1}sec^{-1}$)	Component Size	Calculated Reiteration Frequency
"Total" Repeats[b]	1	3.3	0.36	2800
	2	0.097	0.64	80
"Short" Repeats[c]	1	2.6	0.46	2000
	2	0.046	0.54	40
26 Cloned Repeats	1	–	0.4	1600[d]
	2	–	0.8	90[d]

[a] Reactions with the "total" and "short" repeat ^{3}H-DNA preparations were carried out in the presence of a large excess of total sea urchin DNA at 50°C in 0.12 M phosphate buffer, 0.05% SDS or the equivalent criterion. The reactions were assayed by hydroxyapatite binding at 50°C. Rate constants and component sizes for these reactions were derived from a least-squares analysis of the data (Pearson et al., 1977).

[b] The preparation of this DNA fraction is described by Davidson et al. (1975). 1600 nucleotide long ^{3}H-DNA labeled in vivo was reacted to Cot 6×10^{-3}, and the single-stranded material was collected after fractionation on hydroxyapatite. The DNA was reacted to Cot 40; the duplex fraction was collected on hydroxyapatite and was then reacted again to Cot 20. The duplex fraction was collected on hydroxyapatite. This DNA was sheared to 300 nucleotides and renatured to Cot 100, and the duplex fraction ("total repeats") was collected on hydroxyapatite.

[c] The preparation of this DNA fraction is described by Costantini et al. (1978). 2700 nucleotide long ^{3}H-DNA labeled in vivo was reacted to Cot 40 and treated with S1 endonuclease. The S1-resistant duplex fraction was placed on a Sepharose CL 2B column, and the DNA in the included volume of the column (~300 nucleotides in length) was isolated. This DNA was again reacted to Cot 60, and the duplex fraction ("short repeats") was collected on hydroxyapatite.

[d] Calculated as $\log f = (\Sigma L_i \log f_i)/\Sigma L_i$, where f_i is the reiteration frequency of an individual repetitive sequence family, and f is the calculated reiteration frequency of the component.

renatured tracer. Finally, the tracer was reacted with the driver DNA at 45°C in 0.12 M phosphate buffer, and the heteroduplex thermal stability was measured as above.

Thermal Stability of the Renatured or Native Cloned Tracer

The thermal stability of the double-stranded tracers is affected by duplex length, base composition and perhaps the primary sequence. The T_m varied from as low as 72.6°C for clone 2007 to as high as 86.8°C for cloned repeat 2133A. These T_ms were calculated as the temperature at which the interpolated melting curve achieved 50% of its terminal value (details of the procedures used for data reduction are given in Experimental Procedures). The T_ms for the cloned tracers are listed in column 2 of Table 3. The major sources of difference in the T_ms of the native cloned fragments appear to be their diverse base composition and possibly their primary sequence. Thus as comparison with Table 1 shows, many of the fragments melted are approximately the same length, although they display quite different thermal stabilities. The lowest melting fragment, clone 2007, is also one of the longest, 1100 nucleotides; and the shortest fragment, clone 2109B, has a T_m which is among the highest observed, 83.4°C. Except for the few longer fragments, the differences in length among the cloned repeats could account for only a few degrees in thermal stability (see review by Britten, Graham and Neufeld, 1974). It follows that the cloned repeats are diverse with regard to base compositions. An approximate value for the overall base composition of each fragment can be calculated, and we estimate values ranging from as low as 7% G + C (clone 2007) to 48% G + C (clone 2108).

Divergence in the Total Short Repetitive Sequence Fraction of the Sea Urchin Genome

For purposes of comparison, we renatured and melted a short repetitive sequence fraction prepared by S1 nuclease digestion of partially reassociated sea urchin DNA, followed by criterion gel filtration to isolate repeat duplexes ~300 nucleotides long (Costantini et al., 1978). The derivative melting profiles of the repeat fractions renatured at 45 and 55°C are shown in Figure 3a. Very little difference is evident between these two samples—that is, only a minor fraction of the repetitive sequences is so divergent that it is unable to renature except at the lower criterion temperature. The overall mean divergence measured in this experiment is estimated by comparing the T_m of the total repeat fraction with that of the perfectly matched (that is, native) duplexes of equal length. As Table 3 shows, the T_m of the 45°C sample is 71.8°C, while the T_m of native 300 nucleotide fragments measured in the same manner is ~11°C higher. The difference, called ΔT_{Rep} is converted to percentage of sequence divergence by assuming that 1% sequence mismatch in the renatured strand pairs causes one degree decrease in T_m (reviewed by Britten et al., 1974). Thus the experiment of Figure

in our sample ranged from approximately 3–12,500 copies per haploid genome. The distribution of reiteration frequencies for these 26 families is plotted as a histogram in Figure 2. This sample displays a continuous unimodal distribution with a broad maximum around 10^2–10^3 copies per family. No evidence for discrete frequency components is to be found, at least in this small sample.

Is the reiteration frequency distribution for the 26 cloned families a representative sample? A reasonable statistical comparison is given in Table 2. A repetitive DNA tracer was prepared (Table 2, line 1) and further fractionated after S1 nuclease digestion on a cross-linked Sepharose column. The short repeat fraction (mean fragment length 300 nucleotides) was harvested (Table 2, line 2). Both tracers were then reacted with sea urchin DNA driver. The kinetics of these reactions were each resolved into the best combination of two kinetic components by means of a nonlinear least-squares analysis. Table 2 lists the rate constants and the calculated family sizes. Since the data in this paper show that there is a wide distribution of reiteration frequencies, the kinetic components shown in Table 2 represent the reassociation of many individual families. The solutions shown in Table 2 are a convenient and familiar way of describing the distribution of repetition frequencies in DNA. For sea urchin DNA, two components form an adequate description. The frequencies of the kinetic components are weighted geometric means of the frequencies of many individual repeat families.

Line 3 of Table 2 shows the two kinetic frequency components which have been calculated from the data for the 26 cloned repeat families shown in Table 1. To obtain these components, the set of cloned repeats was divided into two frequency classes on the assumption that the more highly repetitive component includes ~40% of the repetitive DNA while the other contains ~60%—that is, we assume the same mean component sizes as observed for the total repeats. The geometric means of the ten most repetitive families (~40%) and the 16 least repetitive families (~60%) are shown on the last line of Table 2. We conclude that the reiteration frequency distribution for the 26 cloned repeat families is quite similar to that of the total repetitive sequences of S. purpuratus DNA.

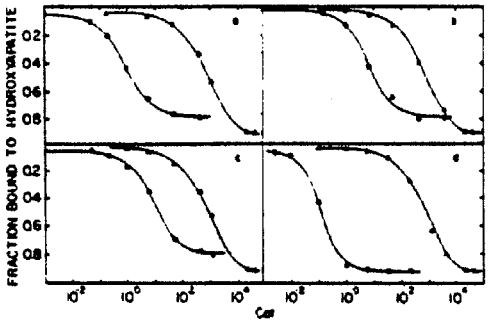


Figure 1. Reaction of Cloned Tracers with Genomic Sea Urchin DNA

Details of the manipulations and conditions used for the kinetic measurements are given in Experimental Procedures. Failure of complete reaction of the cloned tracer is probably due to minor contamination with free γ-^{32}P-ATP from the labeling reaction. Closed circles (●) show the reactions of the cloned ^{32}P-DNA with the driver DNA (the reassociation of which is not shown); closed triangles (▲) show the reactions of the single-copy ^{3}H-DNA tracer with the driver DNA. Points taken at Cot 25,000 and 45,000 indicate terminal reactions of the ^{3}H-DNA with the driver. These points were obtained in separate reactions. The curves show the least-squares solutions assuming second-order kinetics. The rate constants for the reactions shown are listed in Table 1. (a) Clone 2057A; (b) clone 2085; (c) clone 2057B; (d) clone 2100.

Divergence within Repetitive Sequence Families

Three separate thermal stability measurements were made with each of 15 cloned repetitive sequences, and partial measurements were carried out with several additional clones. First, the cloned tracer was renatured, and the thermal stability of the perfectly matched duplex fraction obtained was measured by stepwise thermal elution from hydroxyapatite. Second, the tracer was reacted at 55°C, 0.12 M phosphate buffer with sea urchin DNA driver. The reactions were routinely carried to termination. The stability of the driver-tracer heteroduplex population formed in the reaction was measured in exactly the same way as that of the

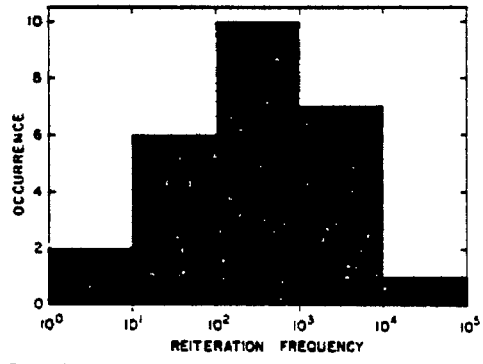


Figure 2. Number Distribution of Reiteration Frequencies of 26 Cloned Repetitive Sequence Families

Data are from Table 1, and details can be found in the footnotes to Table 1, the legend to Figure 1 and Experimental Procedures. The five families of Table 1 whose size is indeterminate (see text) are included in this analysis by using the minimum family size values shown in Table 1. The error involved is probably small on the scale shown here.

Cell
894

Table 3. Thermal Stability Data for the Cloned Repetitive Sequence Families[a]

Clone	Native T_m (°C)	T_m (45°C)	T_m (55°C)	ΔT_c (T_m Native − T_m 45°C)
2005	82.3	73.2	74.8	9.1 (II)
2007	72.7	69.1	70.1	3.6 (I)
2034	79.4	75.7	76.0	3.7 (I)
2057A	–	–	73.3	– (I or II)[e]
2057B	–	–	75.6	– (I or II)[e]
2065	79.6	59.7	64.5	>19.9 (III)
2068	80.7	–	–	– –
2090	83.1	62.6	64.3	>20.5 (III)
2100	–	81.8	82.4	– (I or II)[e]
2101	78.6	69.1	71.4	7.5 (II)
2104	83.6	74.3	75.2	9.3 (II)
2108	86.7	81.0	79.9	5.7 (II)
2109A	80.5	60.8	65.3	>19.7 (III)
2109B	83.4	58.1	–	>25.3 (III)
2110	78.6	72.5[b]	73.3[b]	6.1 (II)
2112	83.3	77.6	81.6	5.7 (II)
2125	81.7	60.2	64.6	>21.5 (III)
2131	86.2	75.4	77.0	10.8 (II)
2133A	86.5	83.0	83.5	3.8 (I)
				Average ΔT_c 10.2[c]

	T_m (45°C)	T_m (55°C)	ΔT_{seq}[d]
Short Repetitive DNA Fraction	71.8	72.7	11.0

[a] Native, 45°C and 55°C T_ms are calculated from thermal chromatography experiments such as those shown in Figure 3. The procedure used for deriving T_ms from the thermal elution profiles is discussed in Experimental Procedures.

[b] For clone 2110, the reactions with total sea urchin DNA were performed at 48 and 60°C, rather than at 45 and 55°C.

[c] Calculated as $\Sigma[(L_i)(\Delta T_{ci})]/\Sigma(L_i)$. All cloned repetitive sequence families for which ΔT_c was determined, except clone 2007, should be represented in the short repetitive DNA preparation obtained in the inclusion peak of the CL-Sepharose column after S1 nuclease digestion. For purposes of comparison to ΔT_{seq}, clone 2007 was therefore excluded from this calculation. Roman numerals indicate the degree of intrafamilial sequence divergence according to the arbitrary classification scheme described in the text.

[d] ΔT_{seq} for the short repetitive DNA fraction is calculated from the T_m (single-copy) corrected for length and for 4°C polymorphism of single-copy DNA sequence among sea urchin genomes (R.J. Britten et al., manuscript submitted). The single-copy ^{3}H-DNA was 320 nucleotides long and the sea urchin DNA driver was ~450 nucleotides long; the average duplex length is thus taken to be ~200 nucleotides. A perfect single-copy duplex of infinite length will have a T_m (measured by thermal chromatography) of 77.8° + 4° + 600/200 = 84.8°C. This number is in good agreement with the T_m measured by thermal elution on hydroxyapatite for long native DNA. The 45°C T_m of the short repetitive DNA (300 nucleotide duplex length), when corrected to infinite length, yields 71.8° + 600/300 = 73.8°C. Thus ΔT_{seq} = 84.8° − 73.8° = 11.0°C.

[e] Clones 2057A and B were not completely analyzed. From the absence of any low thermal stability fraction in the thermal elution profiles (data not shown), however, it is clear that the T_ms are not dependent upon the reaction criterion. Thus they must fall into class I or II. The native T_m was not determined for clone 2110, so ΔT_c could not be calculated. The high T_ms observed in the heteroduplex thermal stability measurements on this clone, however, and the correspondence between the 45°C T_m and the 55°C T_m, clearly demonstrate the lack of dependence on reaction criterion below 55°C for this family. Thus it also belongs to class I or II.

3a indicates that for the total short repeat fraction an average of 11 nucleotides are mismatched per 100 nucleotides of paired sequence.

Reactions of Cloned Tracers with Genomic DNA

When the cloned tracers are reacted with sea urchin DNA driver, the heteroduplex population formed should sample all the possible strand pairs between the cloned sequence and the homologous members of the repeat family in the genome. T_m values for the 45 and 55°C heteroduplex samples formed in reactions of 18 cloned repeats with sea urchin DNA are listed in columns 3 and 4 of Table 3, and representative examples of the thermal stability profiles are shown in Figure 3. These data are presented as elution histograms in Figures 3a and 3b, and as interpolated derivative melting curves in Figures 3c–3f. The melting profile of the renatured

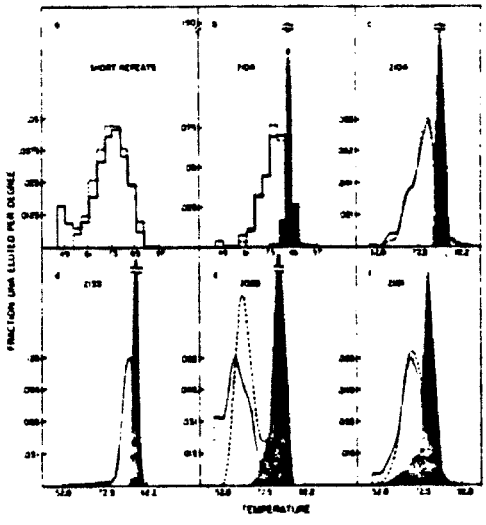


Figure 3. Thermal Stability Profiles for a Total Repetitive DNA Fraction and for Individual Repetitive Sequence Families

A short repetitive ^{3}H-DNA fraction was isolated from the S. purpuratus genome as described by Costantini et al. (1978). Reactions were carried out to kinetic termination at 45 or 55°C, and the duplex fraction was thermally eluted in 4°C intervals from hydroxyapatite columns. Shaded areas represent thermal elution of native or reannealed cloned ^{32}P-DNA. Solid lines represent elution profiles of heteroduplex preparations formed in reactions at 45°C. Dashed lines represent the elution of the heteroduplexes formed at 55°C. (a) Thermal elution profile presented as elution histogram for the short repetitive ^{3}H-DNA tracer reacted with sea urchin DNA. (b) Thermal elution profile presented as elution histogram for double-stranded clone 2104 DNA and for clone 2104-sea urchin DNA heteroduplexes. (c–f) Thermal elution profiles presented as smooth interpolated curves. (c) Double-stranded sea urchin DNA fragment from clone 2104 and clone 2104-sea urchin DNA heteroduplexes. These are the same data as those shown in (b). (d) Corresponding thermal elution profiles for the repetitive DNA of clone 2133; (e) same for clone 2085; (f) same for clone 2101.

or native cloned tracers is indicated as a shaded area in each of these figures.

Any differences between heteroduplex thermal stability profiles and that of the renatured cloned tracer depend upon the intrafamilial divergence within each repeat family. We first consider the case of families which contain very little sequence divergence. The thermal stability profiles of the 45 and 55°C heteroduplexes should be identical for such families, since in order for these melting profiles to differ, there must be sufficient intrafamilial divergence so that a measurable amount of heteroduplex forms at 45°C which cannot form at 55°C. Were there no sequence divergence at all within the repeat family, the heteroduplex melting profile could completely overlap the renatured tracer melting profile. It is possible, however, that sequence polymorphism could be present among individual sea urchin genomes—that is, those from which the driver DNA was prepared. R. J. Britten, A. Cetta and E. H. Davidson (manuscript submitted) have shown that the single-copy DNA of individual S. purpuratus genomes differs by ~4%. It is not known whether this is true of the repetitive sequences as well. If so (as seems probable), a repetitive sequence family with no divergence within a given genome could still produce a melting profile which, though partially overlapping that of the renatured cloned tracer, would peak 4°C below the latter. Families in which the observed sequence divergence is 4% or less—that is, in which the divergence *could* be accounted for completely by the known amount of (single-copy) sequence polymorphism—are referred to in the following discussion as *class I* families. An example is the repeat family to which clone 2133A belongs. Thermal stability measurements for this clone are shown in Figure 3d. Three of the 15 clones for which there are complete data fall into the low divergence class I category.

Those repeat families displaying moderate intrafamilial divergence are termed *class II* families. These are arbitrarily defined for purposes of discussion as families in which the T_m of the heteroduplexes between the cloned tracer and the genomic DNA is more than 4°C below that of the renatured tracer, and no significant difference is observed between the 45 and 55°C melting profiles. Of the 15 repeat families listed in Table 3 for which there are complete data, seven belong to class II. Examples are shown in Figures 3b and 3c (clone 2104) and 3f (clone 2101). The value of ΔT_c, the difference between the T_m of the heteroduplex population and the T_m of the double-stranded cloned fragment, lies between 5.7 and 10.8°C for the seven class II families (Table 3). The weight mean value of ΔT_c for these families is ~8°C.

Repetitive sequence families in which some members are too divergent to have formed stable strand pairs at 55°C in 0.12 M phosphate buffer are termed *class III* families. These families are characterized by a lower thermal stability profile when the cloned tracer is reacted with sea urchin DNA at 45°C rather than at 55°C. Since there is no reason to believe that all the family members are included even at 45°C, we assume the existence of nonreacted, more divergent related sequences in the genome. In other words, our measurements on class III families are criterion-limited. We can therefore, only estimate a minimum value for ΔT_c. Five examples are listed in Table 3, and Figure 3e illustrates the thermal stability profiles of a typical class III repeat family (clone 2085).

Further Analysis of Thermal Stability Measurements and Comparison of the Set of Cloned Families with Total Short Repeats

Our problem is to interpret the observed heteroduplex T_m in terms of the sequence divergence within the repeat family. To approach this question, we first consider a simple model in which the repetitive sequence family has diverged during evolution by accumulating substitutions in the different family members, all of which begin as the same ancestral or type sequence. A heuristic diagram is shown in Figure 4. Here the type sequence is at the origin, and each radial unit indicates an additional unit of divergence from the type sequence (for example, one nucleotide changed per sequence element). This diagram is an elementary representation of a very large number of possible base substitutions. Thus only radial displacements are significant. These displacements are all independent because there are only small numbers of substitutions. Each family member (symbolized by solid circles) is indicated by its own radius since its divergence, in this simple case, has occurred independently. The length of any radius, a_i, gives the number of deviations from the type sequence which exists in the i^{th} family member. The divergence between any two family members is the sum of their radii. The mean radius, $\hat{a}$, is the average amount of change between the type sequence and the family members. Thus in comparing all the family members with each other, the divergence measured would be $2\hat{a}$. For a total repeat fraction such as that shown in Figure 3a, $\Delta T_{m\,Rep}$ gives the weight average of the individual values of $2\hat{a}$ for the repeat families included in the preparation. We now consider the case of measurements with a clone fragment, as in Figures 3b–3f and Table 3. A cloned sequence could have any radius, C, and an example is indicated in Figure 4 by the diamond. The difference in T_m between the renatured, cloned fragment and the clone-genomic DNA heteroduplex populations is $\Delta T_c = C + \hat{a}$. When many different clones are studied, the average value of ΔT_c should, of course, approach the weight average value of $2\hat{a}$ for the population of repeat families.

Figure 4 considers a repeat family with a simple history—that is, a single multiplication event followed by divergence. It is also possible that several events of multiplication have occurred at different stages. If the most recent of these has provided virtually all the current family members, the earlier ones are, of course, irrelevant. It is also possible, however, that the current family consists of several more recently created subfamilies which occurred as the result of multiplication of already divergent descendants of an original type sequence. Here the mean divergence within the family includes all the inter-subfamily as well as intra-subfamily combinations. For this case, however, and in general for any given clone, 2 ΔT_c as measured in our experiments would provide a maximum limit for the value of the mean divergence within that repetitive sequence family. It follows that the weight average of the values of ΔT_c listed in the last column of Table 3 should approximate the mean intrafamilial divergence in the genome if the clones studied are a representative set.

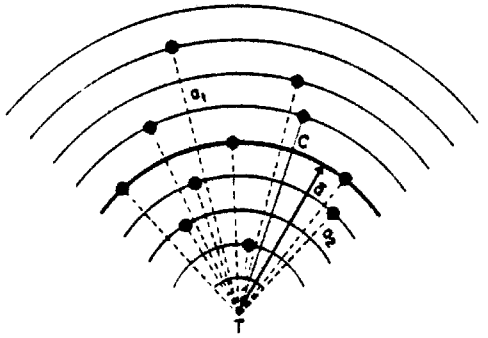


Figure 4. Simplest Case Heuristic Model for Divergence of a Repetitive Sequence Family

It is assumed that the present-day repeat family arose in the following way. A founder or "type" sequence, T, was multiplied in the genome, and the individual copies of this sequence then diverged independently and randomly. Each radial unit or arc in the diagram represents a unit of divergence from the type sequence, which is shown at the origin. The dimensions of these units are changes in nucleotides per sequence element. The solid circles represent individual family members. The mean divergence from the type sequence is signified by the thick arc. The dashed lines or radii represent the distance in units of divergence from a particular family member to the type sequence (for example, a_1 and a_2). The divergence between any two family members is the sum of their radii. The mean divergence is thus given by $\hat{a}$ (thick arrows). A cloned sequence is represented by the diamond and its divergence by the radius C.

The fact that the cloned sequence is itself derived as a hybrid of two different sequences does not affect this analysis, since the equal probability of either strand being corrected at mismatch sites means that in general, $C = {}^1/_2(a_{1,\ldots} + a_{2,\ldots})$, where $a_{1,\ldots}$ and $a_{2,\ldots}$ are the divergences from the type sequence of the two strands constituting the cloned duplex. The individual family members are arbitrarily placed on a given arc, and there is no significance to the lateral distance between individual elements (see text for discussion).

We can now compare the total set of cloned repeat families for which data have been obtained to the genomic short repetitive sequence fraction whose thermal stability profile is shown in Figure 3a. Although the T_m of the DNA reacting at 45 and 55°C is approximately the same (Table 3), there is a small low thermal stability component visible in the 45°C sample. This could represent the lower portion of the thermal stability profiles of class III repeat families in the genome. Just as in the cloned

sample, these appear to include only a minority of the repetitive DNA. Table 3 shows that the average of the ΔT_c values for the cloned repeat families is 10.2°C. As shown above, the weight mean for ΔT_c should approximate the measured value of ΔT_{Res}, which is 11.0°C. From the standpoint of their internal divergence, this agreement indicates that the 18 cloned repeat families are a reasonable sample of the total short repetitive sequence class, just as they appear to constitute a typical sample with respect to length and repetition frequency.

Discussion

The availability of cloned repetitive sequences has, for the first time, made it possible to examine the characteristics of individual repeat families. We show here that the sequence length, repetition frequency and intrafamilial divergence of the sample of cloned sea urchin DNA repeats which we have studied is fairly typical of the genomic repetitive sequence taken as a whole. Nonetheless, it remains possible that there are features of certain repetitive sequences which increase the probability of their being successfully cloned by the method which we have used. Were this true, the set of repeat families which we have investigated would be in some unknown respect nonrepresentative. The most probable problem a priori is the possibility that highly divergent repeat families were selected against. The close agreement between the mean intrafamilial divergence of the total genomic repeat fraction and the average intrafamilial divergence of the cloned families specifically argues against this. Both the cloned DNA, however, and the repeat fraction which was melted in the experiment of Figure 3a were exposed to S1 nuclease. We rely on our earlier studies which show that under appropriate conditions, S1 nuclease spares almost all of the mismatched duplexes formed when repetitive DNA is renatured (Britten et al., 1976). No experiments which have been carried out to date can eliminate the possibility that another class of very poorly related repeats exists in the genome, one so divergent that it can be detected only under lower criterion conditions not yet explored. Thus the properties of the cloned repeat families which we have studied are certainly not overtly bizarre and some of them are of special interest for other reasons. A randomly selected subset of nine of the cloned repeats included in the present study has been examined to determine whether numbers of their families are transcribed (Costantini et al., 1978; Scheller et al., 1978). These results (presented elsewhere) showed that hybridizing transcripts were found in every case, and that there is a wide range in the concentration of these transcripts in RNAs from different cell types.

One purpose of this investigation was to examine the frequency distribution of the repeat families in the S. purpuratus genome. The measurements reported in Table 1 provide the first unequivocal and precise demonstration that there is a broad range of repetition frequencies in the genome. The family size distribution is continuous, and displays no forbidden areas or discontinuities other than those expected from the relatively small sample size. It is clear that the data reduction methods used for analysis of DNA reassociation are indeed averaging procedures. The smooth shape of the frequency distribution contributes an important simplification for future calculations of complexity, family member, family size and other parameters ultimately needed for understanding the role and evolution of repetitive DNA.

A major focus of this investigation is measurement of the intrafamilial sequence divergence. Previously, divergence could be measured only as an average calculated over huge numbers of individual repeat families in experiments such as that shown in Figure 3a. We now see that individual repeat families differ sharply in the extent to which their members diverge from each other in sequence. Of the 18 cloned families for which we present data, 13 display a limited measurable divergence (class I + class II of Table 3). We have shown that these families lack detectable, distantly related members. The mean intrafamilial sequence divergence for these families is significantly <8%. This is a major result of the present study. A few of the repeat families, those of class I, are even less divergent, and it is possible that within any given genome, most copies of the repeats in class I families could be almost exact sequence replicas.

Repeat families displaying limited divergence are physically discrete sequence sets—that is, they have a certain number of members and do not extend outward indefinitely. Measurements made on the class I and II families are clearly not criterion-limited. This assurance provides a firm basis for calculations of repetitive sequence complexity and reiteration frequency. These statements are not true of the class III families. Five of the 18 families studied display a high degree of intrafamilial divergence and probably possess distant relatives not included in our measurements. The true size and sequence divergence within these families is not known and will be difficult to measure by conventional means. Nevertheless, clones 2109A, 2109B and 2090 all belong to class III families which are transcribed specifically (Costantini et al., 1978; Scheller et al., 1978). It will clearly be of interest to determine which members of these more divergent families are utilized in transcription, and which parts of their internal sequence are free to diverge and which are restrained.

The limitation in intrafamilial sequence divergence demonstrated here for the majority of the cloned repeat families should be considered in light of the evolutionary persistence of repetitive sequence families among sea urchin species and genera. Moore et al. (1978) report elsewhere that all of the cloned S. purpuratus repeats which they investigated have close homologues in a congener, S. franciscanus. The interspecies divergence of these cloned repeats is significantly less than that of the single-copy DNA. Moore et al. (1978) also found that many of the S. purpuratus cloned repeats react with sequences present in the genome of a very distantly related sea urchin species, Lytechinus pictus. For comparison, only 10–20% of total S. purpuratus single-copy DNA cross-reacts with the latter genome (Angerer, Davidson and Britten, 1976; Harpold and Craig, 1978). Studies on the evolution of total repeat fractions (Harpold and Craig, 1977) also indicate that repeat sequences disappear much more slowly (if at all) during sea urchin evolution than do single-copy sequences.

Restraint on repetitive sequence divergence, either within the repeat families of a given species, or over evolutionary time spanning the emergence of diverse species, could be due to selective pressures which prevent free sequence change in a large fraction of the repeat family members. Or perhaps repetitive sequences diverge as rapidly as do other sequences, but the type sequence of the family is preserved by frequent remultiplication of the "correct" surviving sequences. For example, the current members of class I and class II repeats could be mainly new evolutionary creations so recent in origin that there has not been enough time for a large amount of sequence divergence to occur. In this connection, it is significant that Moore et al. (1978) have found that some of the cloned S. purpuratus repeat families display several times more members in the S. purpuratus genome than do the closely homologous repeat families in the S. franciscanus genome. Intraspecies repeat family multiplication thus seems possible. Such "family correction" or remultiplication mechanisms are difficult to understand because most members of most of the cloned repeat families are not tandemly arranged, but are probably widely interspersed about the genome (Graham et al., 1974; Lee, Britten and Davidson, 1977; Scheller et al., 1978). In any case, the alternatives of continuous repeat family renewal and selective restraint on sequence changes due to functional requirements are not necessarily exclusive.

Experimental Procedures

Construction and Description of Recombinant Plasmids

Recombinant bacterial plasmids containing repetitive sea urchin DNA sequences were constructed as described by Scheller et al. (1977). A brief description is given in Results. All experiments with these recombinant DNAs were carried out under P2-EK1 conditions, as specified in the NIH Guidelines.

Preparation of Cloned Repetitive ^{32}P-DNA

Recombinant plasmid DNA was isolated essentially as described by Scheller et al. (1977). Plasmid-containing cells were grown, harvested, washed and lysed as described previously (Scheller et al., 1977). Superhelical DNA was purified from the supernatant by isopycnic centrifugation in a CsCl-ethidium bromide gradient. DNA was cleaved with >25 units of Eco RI per 25 μg of DNA. This digestion was carried out at 37°C for 30–60 min in 100 mM Tris-HCl (pH 7.6), 100 mM NaCl and 5 mM $MgCl_2$. The reaction mixtures were precipitated and redissolved in 10 mM Tris (pH 8.0) at a DNA concentration <50 μg/ml. 2–5 units of bacterial alkaline phosphatase were added per 50 μg of nucleic acid. The reaction was carried out at 37°C for 30–60 min. This mixture was then extracted twice with an equal volume of buffer-saturated phenol: Sevag solution (1:1) (Sevag solution is chloroform:isoamyl alcohol 24:1) and once with ether. Dephosphorylated Eco RI-cleaved DNA was labeled at the 5' termini with γ-^{32}P-ATP essentially as described by Maxam and Gilbert (1977). DNA was dissolved in 5 mM Tris (pH 9.5), 0.01 mM EDTA, 0.1 mM spermidine, 50 μg/ml DNA and denatured at 100°C for 3 min. The solution was then brought to 50 mM Tris (pH 9.5), 10 mM $MgCl_2$ and 5 mM dithiothreitol. γ-^{32}P-ATP was dried down, dissolved in H_2O and added to the reaction mixture at not less than 3 μM and at greater than a molar equivalence to the 5' termini. 10–20 units of polynucleotide kinase were added, and the reaction was incubated for 45 min at 37°C. This reaction was then extracted once with phenol:Sevag (1:1) and once with ether and ethanol-precipitated. Specific activities ranged from 5×10^6 to 2×10^7 cpm μg^{-1}, depending upon the length of the cloned fragment and the efficiency of the individual kinase reactions.

The 5'-^{32}P-labeled sea urchin DNA fragments were isolated from the plasmid DNA in one of two ways. For clones 2057A, 2057B, 2090 (preparation a), 2096, 2097, 2099, 2100, 2103, 2108 (preparation a), 2110 and 2111, the dried precipitates were dissolved in 0.1 M NaCl, 0.05 M Tris (pH 7.2), 1 mM EDTA and 0.1% SDS, heat-denatured, and renatured at 60°C for periods equivalent to >10 × $Cot_{1/2}$ for each fragment. The renatured DNA was loaded onto a 3% agarose gel and subjected to electrophoresis at 60 V for ~3.5 hr in 30 mM Tris, 20 mM sodium acetate, 10 mM EDTA (pH 7.8) at room temperature. The ^{32}P-DNA fragments were excised from the gels and eluted by dissolving the agarose in 5 M $NaClO_4$. 20 μg of sheared calf thymus DNA carrier were added; the DNA was then bound to hydroxyapatite at room temperature in 0.12 M phosphate buffer. After extensive washing, bound DNA was eluted from the hydroxyapatite with several aliquots of 0.5 M phosphate buffer, 0.1% SDS. For clones 2005, 2007, 2034, 2085, 2086, 2090 (preparation b), 2102, 2106 (preparation b), 2109A, 2109B, 2112, 2125, 2131, 2133A, 2133B and 2137, the labeled, dried precipitates were dissolved in 0.3 M NaOH, 10% glycerol, 1 mM EDTA and incubated at 37°C for 10 min. Samples were subjected to gel electrophoresis on different percentage acrylamide gels depending upon the known size of the inserted sea urchin DNA fragment. The majority of fragments were run on 4–6% gels at 60 V in 50 mM Tris-borate buffer (pH 8.3), 1 mM EDTA at 15°C for 16 hr. After electrophoresis, gels were stained with ethidium bromide and autoradiographed. The electrophoretic conditions used were essentially the strand separation conditions described by Maxam and Gilbert (1977). The polyacrylamide gels were eluted in the following manner. Gel slices were crushed with a siliconized glass rod in a 1.5 ml Eppendorff tube. 1 ml of 0.12 M phosphate buffer, 0.05% SDS, containing 10 μg of purified calf thymus DNA carrier, was added to the crushed gel. The mixture was incubated overnight at 37°C. The acrylamide was filtered through siliconized glass wool and rinsed with an additional 1 ml of 0.12 M phosphate buffer, 0.05% SDS. This gel eluate contained 80–95% of the radioactivity. In the case of clones 2090 (preparation b) and 2106 (preparation b), strand-separated DNAs were

used for the thermal stability measurements. The strand-separated DNA was concentrated 4 fold by extraction with 2-butanol to approximately 0.5 M phosphate buffer. The ^{32}P-DNA was then incubated for 10 hr at 55°C to allow any contaminating complementary strands to renature. The mixtures were then diluted to 0.12 M phosphate buffer and passed over 1 ml of hydroxyapatite at 50°C in 0.12 M phosphate buffer. The nonbound fraction contained single-stranded fragments. These fragments were contaminated from 0–5% with their complementary strands and up to 30% unincorporated γ-^{32}P-ATP. The labeled DNA was stored at −20°C after dialysis versus 0.12 M phosphate buffer, 0.1% SDS.

Length Measurements of Cloned Repetitive Sequence Elements

Lengths of cloned DNA fragments were determined by electrophoresis on 3% agarose gels, by comparison to DNA length standards. An Hae III digest of RSF2124 provided a set of length standards for calibration.

Preparation of Unlabeled DNA

DNA was extracted from fresh S. purpuratus sperm essentially as described by Britten et al. (1974). The DNA was sheared to low molecular weight in the Britten press at 50,000 psi (Britten et al., 1974). Three preparations of unlabeled DNA were used in these studies. Two preparations had weight average lengths of 450 nucleotides and one had a weight average length of 650 nucleotides as measured on alkaline sucrose gradients (Davidson et al., 1975).

Preparation of Single-Copy ^{3}H-DNA

^{3}H-DNA was extracted from sea urchin embryos grown in ^{3}H-thymidine and was used to prepare single-copy ^{3}H-DNA as described by Galau, Britten and Davidson (1974). The specific activity of the single-copy ^{3}H-DNA was 2×10^6 cpm μg^{-1} under our counting conditions.

Reactions with Cloned Repetitive ^{32}P-DNA

Reaction mixtures varied considerably with regard to amounts of cloned repetitive ^{32}P-DNA and sea urchin DNA. In general, 100–500 cpm of nonstrand-separated ^{32}P-DNA and 2–200 μg of total sea urchin DNA were reacted in 0.12 M phosphate buffer, 0.05% SDS in volumes ranging from 10–500 μl. Reaction mixtures also contained 1000–2000 cpm of single-copy ^{3}H-DNA. After denaturing at 97°C, the reactions were carried out in sealed glass ampules or in microcapillaries at 55°C. Samples were reacted to Cot values ranging from 10^{-3} to 10^4 M sec; reaction times were from 30 sec to 24 hr. In a few cases for high Cot reactions, 0.41 M phosphate buffer, 0.05% SDS were used. These Cot values and all other Cot values referred to in this work are *equivalent Cots* (that is, the Cot of the reaction corrected for the relative increase in rate due to salt concentrations above 0.18 M Na^+). Reactions were diluted into 2 ml of 0.12 M phosphate buffer, 0.05% SDS. This solution was applied to 1–2 ml of hydroxyapatite at 55°C and appropriate fractions were collected. The duplex DNA fraction was eluted from the hydroxyapatite by raising the temperature to 97°C. In all but four cases, the sequence ratios of driver to cloned ^{32}P-DNA exceeded 20, so reaction of the tracer with itself was negligible. Clones 2086, 2096, 2097 and 2108, due to their low reiteration frequencies, were driven by 2–4 fold sequence excess of sea urchin DNA. When corrections are applied for reaction of the ^{32}P-DNA with itself, however, no appreciable effect on the reiteration frequency is observed in these cases.

Thermal Chromatography

Reactions with Sea Urchin DNA

Reaction mixtures generally contained 200–2000 cpm of nonstrand-separated cloned repetitive ^{32}P-DNA and 200–600 μg of sea urchin DNA. In two cases (2108 and 2090), only one strand of the cloned ^{32}P-DNA was used. The reactions were all carried out to kinetic termination (>10 × the $Cot_{1/2}$). The mixtures were then diluted into 3 ml of 0.12 M phosphate buffer, 0.05% SDS. 1000–2000 cpm of renatured single-copy ^{3}H-DNA were added, and the sample was applied to 1–2 ml of hydroxyapatite at 45 or 53°C (for reaction temperatures of 45 and 55°C, respectively). The solution was allowed to equilibrate to column temperature before passage through the hydroxyapatite. Seven to ten additional 3 ml aliquots of 0.12 M phosphate buffer, 0.05% SDS were washed through the column, and the temperature of the column was then raised 4°C. Three 3 ml aliquots of 0.12 M phosphate buffer, 0.05% SDS were then collected. This process was repeated in 4°C intervals up to 97°C. The single-copy tracer provided an internal standard for the control of minor systematic variations in the operation of the hydroxyapatite columns. These included the rate of elution and the rate of temperature increase. The single-copy tracer melting profiles were relatively reproducible. The mean T_m for all of the single-copy standards was 77.8°C, with a standard deviation of only ±0.5°C. The T_ms measured for the single-copy standards were used to correct the experimental T_ms to a uniform standard melting condition. The magnitude of these corrections was usually only a few tenths of a degree C and never more than 1.3°C.

Thermal Chromatography of Native Cloned ^{32}P-DNA

For thermal chromatography of native ^{32}P-DNA fragments, the elution procedure followed was as described above. Where the cloned repetitive ^{32}P-DNA was renatured, 10^3–10^4 cpm of the ^{32}P-DNA were reacted in 0.5 M phosphate buffer, 0.05% SDS at 67°C to kinetic termination. The hydroxyapatite elution procedure was as described above.

Technical Controls to Assure That Hydroxyapatite Quantitatively Binds Poorly Mismatched Duplexes

Measurement of the stability of mismatched duplexes by thermal chromatography from hydroxyapatite could result in serious errors if very poorly matched duplexes (for example, 20–30% mismatched) fail to bind even though they are not single-stranded. To test whether poorly matched duplexes bind to hydroxyapatite, we reacted S. franciscanus DNA with S. purpuratus single-copy ^{3}H-DNA in 0.5 M phosphate buffer, 0.05% SDS at 50°C to Cot 20,000. This heterologous reaction is known to produce mismatched duplexes: the ΔT_m between the homologous S. purpuratus reaction and the heterologous S. franciscanus-S. purpuratus reaction is approximately 10°C (Angerer et al., 1976; Harpold and Craig, 1977; T. J. Hall, E. H. Davidson and R. J. Britten, unpublished observations). The reaction mixture, after dilution into 0.12 M phosphate buffer, 0.05% SDS was placed over a 1.5 ml bed of hydroxyapatite at 45°C. At this temperature, 0.29 of the single-copy ^{3}H-DNA passed through the column, approximately as expected for the heterologous reaction at this criterion and Cot. The bound ^{3}H-DNA displayed a broad thermal stability profile, clearly limited by the reaction criterion. The apparent T_m was ~84°C, and 30% of the ^{3}H-DNA bound originally had eluted by 57°C. Thus short, very poorly matched duplexes were bound to the column. To determine whether any double-stranded DNA was present in the 0.12 M phosphate buffer hydroxyapatite eluate at 45°C, the 0.29 unbound fraction was analyzed for duplex content by measuring the kinetics of its digestion with S1 nuclease. This experiment was carried out using increasing amounts of S1 nuclease and determining the fraction of ^{3}H-DNA excluded on a Sephadex G-100 column. The rate of S1 nuclease digestion of this DNA was found to be identical to that of a completely denatured single-copy ^{3}H-DNA sample consisting of the same single-copy ^{3}H-DNA plus S. franciscanus driver DNA. The digestion kinetics (Britten et al., 1976) provide a sensitive test for duplex structure and show that no significant fraction of the hydroxyapatite eluate was double-stranded. At high levels of S1 nuclease, 0.025 of the denatured ^{3}H-DNA and 0.075 of the ^{3}H-DNA not bound to hydroxyapatite were resistant to digestion. Thus the maximum fraction of the duplex present which was not bound on hydroxyapatite at 45°C was:

$$\frac{[(0.075) - (0.025)](0.29)}{[(0.075) - (0.025)](0.29) + 0.71} = 0.02$$

Thermal Stability Data Reduction Procedures

The values listed in Table 3 were obtained with the aid of a computer program which operated essentially as follows. The cumulative hydroxyapatite melting data were fit with an algorithm which calculated third-order polynomials through the data points and interpolated ten points within each 4°C temperature interval ("spline" algorithm). The derivatives of the polynomials were matched at each interpolated overlap point. The derivatives of the polynomials are shown, for example, in Figures 3c-3f. T_m from the polynomial curve were calculated as the temperature at which 50% of the DNA had melted.

Acknowledgments

This research was supported by a grant from the National Institute of General Medical Sciences. W.H.K. is the recipient of a Lievre Fellowship from the American Cancer Society, California Division, and T.L.T. is the recipient of an NIH postdoctoral fellowship. R.H.S. is a predoctoral fellow on an NIH training grant. The P2 facilities used in this work were equipped with the aid of funds from a Biomedical Research Support Grant.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received February 28, 1978; revised May 1, 1978

References

Angerer, R. C., Davidson, E. H. and Britten, R. J. (1976). Chromosoma *56*, 213–226.

Bonner, T. I., Brenner, D. J., Neufeld, B. R. and Britten, R. J. (1973). J. Mol. Biol. *81*, 123–135.

Britten, R. J. and Kohne, D. E. (1967). Carnegie Inst. Washington Yearbook *65*, 78–106.

Britten, R. J. and Kohne, D. E. (1968). Science *161*, 529–540.

Britten, R. J., Graham, D. E. and Neufeld, B. R. (1974). In Methods in Enzymology, *29E*, L. Grossman and K. Moldave, eds. (New York: Academic Press), pp. 363–406.

Britten, R. J., Graham, D. E., Eden, F. C., Painchaud, D. M. and Davidson, E. H. (1978). J. Mol. Evol. *9*, 1–23.

Chamberlin, M. E., Britten, R. J. and Davidson, E. H. (1975). J. Mol. Biol. *96*, 317–333.

Chamberlin, M. E., Galau, G. A., Britten, R. J. and Davidson, E. H. (1978). Nucl. Acids Res. *5*, 2073–2094.

Costantini, F. D., Scheller, R. H., Britten, R. J. and Davidson, E. H. (1978). Cell *15*, in press.

Davidson, E. H. and Britten, R. J. (1973). Quart. Rev. Biol. *48*, 565–613.

Davidson, E. H., Hough, B. R., Klein, W. H. and Britten, R. J. (1975). Cell *4*, 217–238.

Eden, F. C., Graham, D. E., Davidson, E. H. and Britten, R. J. (1977). Nucl. Acids Res. *4*, 1553–1567.

Galau, G. A., Britten, R. J. and Davidson, E. H. (1974). Cell *2*, 9–21.

Graham, D. E., Neufeld, B. R., Davidson, E. H. and Britten, R. J. (1974). Cell *1*, 127–137.

Harpold, M. M. and Craig, S. P. (1977). Nucl. Acids Res. *4*, 4425–4438.

Harpold, M. M. and Craig, S. P. (1978). Differentiation *10*, 7–11.

Lee, A. S., Britten, R. J. and Davidson, E. H. (1977). Science *196*, 189–192.

Maxam, A. M. and Gilbert, W. (1977). Proc. Nat. Acad. Sci. USA *74*, 560–564.

Mizuno, S. and Macgregor, H. C. (1974). Chromosoma *48*, 239–296.

Moore, G. P., Scheller, R. H., Davidson, E. H. and Britten, R. J. (1978). Cell *15*, in press.

Pearson, W. R., Davidson, E. H. and Britten, R. J. (1977). Nucl. Acids Res. *4*, 1727–1737.

Scheller, R. H., Thomas, T. L., Lee, A. S., Klein, W. H., Niles, W. D., Britten, R. J. and Davidson, E. H. (1977). Science *196*, 197–200.

Scheller, R. H., Costantini, F. D., Kozlowski, M. R., Britten, R. J. and Davidson, E. H. (1978). Cell *15*, in press.

So, M., Gill, R. and Falkow, S. (1975). Mol. Gen. Genet. *142*, 239–249.

Weinblum, D., Güngerich, U., Gesert, M. and Zahn, R. K. (1973). Biochim. Biophys. Acta *299*, 231–240.

CHAPTER 5

Cell, Vol. 15, 173-187, September 1978, Copyright © 1978 by MIT

Repetitive Sequence Transcripts in the Mature Sea Urchin Oocyte

Franklin D. Costantini, Richard H. Scheller,* Roy J. Britten† and Eric H. Davidson
Division of Biology
California Institute of Technology
Pasadena, California 91125

Summary

The expression of interspersed repetitive sequences in the RNA of mature sea urchin oocytes was investigated. ^{3}H-DNA tracers representing short interspersed repetitive sequences a few hundred nucleotides long, and long repetitive sequences approximately 2000 nucleotides long, were prepared from genomic DNA of the sea urchin, Strongylocentrotus purpuratus. These tracers were reacted with excess RNA from the mature oocyte. About 80% of the reactable short repeat tracer and 35% of the long repeat tracer hybridized. Thus most of the repetitive sequence families in the short repeat tracer are represented in oocyte RNA, and transcripts complementary to both strands of many repeat sequences are present. The kinetics of the reaction show that some transcripts are highly prevalent (>10^5 copies per oocyte), while others are rare (~10^3 copies per oocyte). Nine cloned repetitive sequences were labeled, strand-separated and reacted with the oocyte RNA. Transcripts of both strands of all nine repeats were found in the RNA. The prevalence of transcripts of the cloned repeat families varied from ~3000 to 100,000 copies per oocyte. Studies with both cloned and genomic tracers show that transcript prevalence is independent of the genomic reiteration frequency of the transcribed repetitive sequences. Most of the families represented by prevalent transcripts have fewer than 200 copies per haploid genome. The RNA molecules with which the cloned repeats react are at least 1000-2000 nucleotides in length. Other experiments show that a majority of the members of repeat families represented by prevalent transcripts in the oocyte RNA are interspersed among single-copy sequence elements in the genome.

Introduction

The mature oocyte contains a large and heterogeneous stockpile of RNA molecules. These are apparently destined for use in early development, although little precise information on this point is available. The best studied component of the heterogeneous RNA of the oocyte is maternal message (matmRNA). Most of the single-copy transcript in the RNA of the mature sea urchin oocyte appears to be matmRNA. Thus at least 73% of the single-copy sequence fraction which can be driven into RNA-DNA hybrids by oocyte RNA also reacts with the polysomal RNA of cleavage stage embryos (Hough-Evans et al., 1977). Humphreys (1971) and other investigators showed clearly that most of the mRNA molecules translated on cleavage stage polysomes are maternal in origin (reviewed by Davidson, 1976). The matmRNA represents ~1% of the mass of the RNA in the mature oocyte and its complexity is approximately 3.7×10^7 nucleotides, or ~6% of the total single-copy sequence in the genome (Anderson et al., 1976; Galau et al., 1976; Hough-Evans et al., 1977). Thus there are about 20,000 different mRNA sequences, each of which is present on the average in about 1600 copies per egg. Histone matmRNAs are also present in large quantities in sea urchin oocytes (Farquhar and McCarthy, 1973; Skoultchi and Gross, 1973; Gross et al., 1973). These are, of course, repetitive sequence transcripts, and they may well account for the results of some of the early filter hybridization measurements which suggested that oocytes contain repetititve sequence transcripts (reviewed by Davidson, 1976). Some of the competition hybridization experiments with oocyte RNA, however, suggest the presence of a greater variety of repetitive sequence transcripts than can be accounted for as histone mRNAs (for example, see Glisin, Glisin and Doty, 1966; Chetsanga et al., 1970). Hough and Davidson (1972) also showed that a significant fraction of an isolated repetitive sequence preparation from the Xenopus genome reacts with Xenopus oocyte RNAs. It can be inferred from the early hybridization literature that the maternal RNA of the oocyte contains a relatively diverse set of repetitive sequence transcripts aside from matmRNA.

This paper describes experiments which demonstrate that sea urchin oocyte RNA includes transcripts of some members of at least 80% of the repetitive sequence families in the genome. Different repeat families are represented to greatly different extents, so that the number of transcripts per oocyte varies from a few thousand to over 100,000. In addition, we report that both strands of most repeat families are represented in the oocyte RNA.

Results

Oocyte RNA Transcripts of Short and Long Repetitive Sequences

The repetitive sequences of the S. purpuratus genome fall into two categories with respect to sequence element length. Approximately 60% of the

* Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125.
† Also a staff member of the Carnegie Institute of Washington, Baltimore, Maryland 21210.

mass of the repetitive DNA consists of sequences only a few hundred nucleotides long, interspersed with single-copy DNA (Graham et al., 1974; Eden et al., 1977). As is characteristic of most interspersed genomes studied, the average length of these "short repeats" is close to 300 nucleotides (for example, Schmid and Deininger, 1975, Chamberlin, Britten and Davidson, 1975; Goldberg et al., 1975). The remaining 40% of the repetitive DNA mass is found in "long repeat" sequence elements, which in the sea urchin genome are ≥2000 nucleotides in length. The long and short repeat classes of repetitive sequence are at least partially distinct sequence sets. Thus renatured short repeats include more mismatched base pairs than do renatured long repeats (Britten et al., 1976), and cross-reactions between the long and short repeat fractions show that each sequence class comprises only a minor fraction (10–30%) of the mass of the other (Eden et al., 1977; our unpublished data).

Long and short repetitive sequence tracers were prepared for use in hybridization experiments with oocyte RNA, as described in detail in Experimental Procedures and Figure 1. The kinetics of the reaction of the short repeat tracer with whole sheared sea urchin DNA indicate that the distribution of repetitive sequence frequencies in this tracer is similar to that in the whole genome, as shown below. The same result was obtained by Eden et al. (1977). We also confirmed the conclusion of Eden and her co-workers that the sequences comprising the short repeat fraction are a minor component of the long repeat fraction, and vice versa. Thus an unlabeled short repeat DNA fraction prepared as in Figure 1 drives the long repeat tracer approximately 3–5 times more slowly than it does the short repeat tracer. A similar kinetic disparity is observed in the reciprocal reaction (data not shown).

A large excess of RNA extracted from mature oocytes was hybridized with the long and short repeat ^{3}H-DNA preparations. The reactions were performed under relatively low criterion conditions (0.41 M phosphate buffer, 55°C) to minimize the kinetic effects of mismatch in the repetitive sequence duplexes (see Klein et al., 1978). Hybridization was assayed by hydroxyapatite binding, and DNA tracer self-reaction was measured after low salt RNAase digestion (Galau, Britten and Davidson, 1974; Galau et al., 1976) as described in Experimental Procedures. The kinetics of these reactions are displayed in Figure 2. Here it can be seen that 73% of the short repeat tracer (open circles) is recovered in RNA-DNA hybrids at the highest RNA Cot (55,000 M sec), and there is no reason to believe that the reaction has terminated at this point. The short repeat tracer is approximately 93% reactive (measured with whole DNA

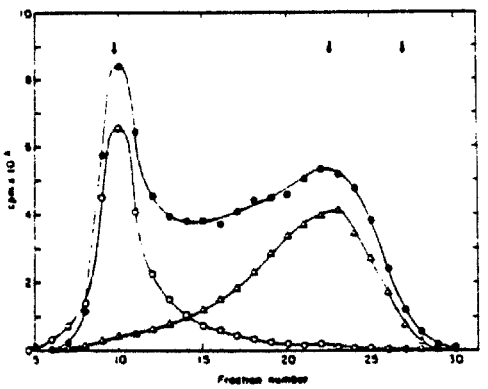


Figure 1. Sepharose CL-2B Fractionation of S1 Nuclease-Resistant Repetitive DNA

Total sea urchin ^{3}H-DNA was sheared to an average singe-strand length of 3300 nucleotides, renatured to Cot 40 and digested with S1 nuclease to remove single-stranded regions. The S1 nuclease-resistant duplex fraction (21%) was recovered on hydroxyapatite and chromatographed on a Sepharose CL-2B gel filtration column in 0.12 M PB at 60°C (●). Fractions were collected and an aliquot of each fraction was counted. DNA in the excluded peak (fractions 8–12) was pooled and rechromatographed (○), and again the excluded fractions (7–12) were pooled. This material constituted the long repetitive DNA fraction. Similarly, the included peak material (fractions 18–26) was pooled and rechromatographed (△), and the resulting peak (fractions 19–25) constituted the short repetitive DNA fraction. From left to right, the arrows mark the peak elution positions of long native DNA, 300 nucleotide long sheared native DNA and ^{32}P-orthophosphate. The k_{ss} of 300 nucleotide long native DNA on this column is 0.74.

driver), and by Cot 55,000, tracer self-reaction has reduced the ^{3}H-DNA fraction available for hybridization by about 5%. Using these numbers, the experiment in Figure 2 shows that at least 83% of the short repeat tracer is capable of hybridizing with oocyte RNA—that is, almost all of the short repetitive DNA sequence is represented in oocyte RNA. This does not necessarily imply that most of the short repeats in the genome have been transcribed, but that at least one member of most short repetitive sequence families has been transcribed. These data also indicate that both complementary strands of most short repeat sequence families are represented in the RNA. Were this not true, a maximum of only 50% of the tracer could have reacted.

The kinetics of hybridization of the short repeat tracer with oocyte RNA are very heterogeneous. It follows that the concentration of different repetitive sequence transcripts in oocyte RNA must vary greatly. For comparison, the pseudo-first-order hybridization kinetics for the reaction of a single-copy ^{3}H-DNA tracer with excess oocyte RNA are also indicated in Figure 2 (dashed line) (Galau et al., 1976; Hough-Evans et al., 1977). The short repeat

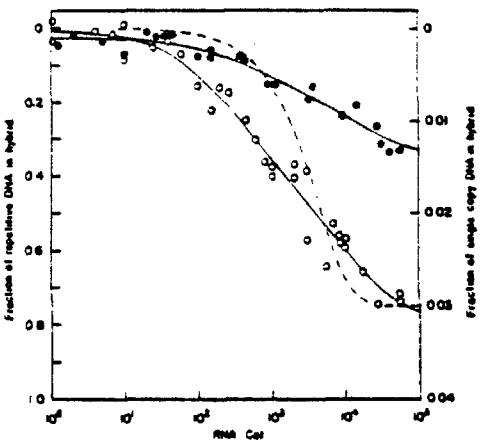


Figure 2. Hybridization of Short and Long Repeat ^{3}H-DNA with Excess Oocyte RNA

The short repeat ^{3}H-DNA (○) and long repeat ^{3}H-DNA (●) tracers were prepared as described in Figure 1 and the text, and hybridized with excess oocyte RNA at 55°C in 0.41 M PB, 0.2% SDS, 0.005 M EDTA. The fraction of ^{3}H-DNA in RNA-DNA hybrids (ordinate) is shown as a function of RNA Cot. This fraction was measured by binding to hydroxyapatite at 50°C in 0.12 M PB, 0.05% SDS. The small amount of tracer binding due to DNA-DNA duplex formation (<5% at the highest RNA Cot) was measured as described in Experimental Procedures, and has been subtracted from the total binding to yield the values shown. The RNA/^{3}H-DNA mass ratio was at least 10^4 in all reactions; in reactions to high RNA Cot, a 10^5 fold excess of RNA was used to prevent tracer self-reaction. The data include hybridizations using several different preparations of oocyte RNA. To exclude the possibility that the lower hybridization of the long repeat tracer could be due to incomplete denaturation of long sequences in the RNA prior to hybridization, the long repeat ^{3}H-DNA was hybridized with a sample of oocyte RNA that had been denatured in 80% dimethyl sulfoxide at 55°C prior to hybridization (▦). These data are indistinguishable from the data obtained after the usual aqueous thermal denaturation (see Experimental Procedures). The solid line through the short repeat data shows a least-squares solution assuming three second-order kinetic components. This analysis suggests that 11% of the short repetitive DNA hybridizes with a second-order rate constant (k_3) of 1.08×10^{-5} M^{-1} sec^{-1}, 31% hybridizes with a k_2 of 2.36×10^{-3} M^{-1} sec^{-1} and 37% hybridizes with a k_1 of 9.96×10^{-2} M^{-1} sec^{-1}. The long repeat tracer reaction was best fit with two second-order components (solid line). 13% of the long repeat ^{3}H-DNA appears to hybridize with a k_2 of 2.45×10^{-3} M^{-1} sec^{-1}, and 19% with a k_1 of 7.69×10^{-2} M^{-1} sec^{-1}. For these reactions, it was assumed that the hybridizing RNAs were significantly longer than the ^{3}H-DNA tracers (see Figure 7), and a minimum length correction factor (f_L) of 2 was applied as described in equation (1) of Table 1. The dashed line represents the kinetics of hybridization of a single-copy ^{3}H-DNA tracer with excess oocyte RNA, fit with a single pseudo-first-order function with a rate constant (k_1) of 2.3×10^{-4} M^{-1} sec^{-1} (Hough-Evans et al., 1977). The scale for this reaction is shown on the right-hand ordinate.

hybridization data were reduced by a least-squares procedure, assuming three components, each reacting with second-order kinetics. The assumption of second-order rather than pseudo-first-order kinetics is suggested by the apparent presence of both complements of most repeat sequences in the RNA. The kinetic parameters derived from the least-squares solution are given in the legend to Figure 2, and the solution itself is shown by the solid line drawn through the open circles. The rate constants obtained for the short repeat hybridization reaction provide an estimate of the approximate range of concentrations of repetitive sequence transcripts in the oocyte RNA [equation (1) of Table 1]. Hough-Evans et al. (1977) calculated that one S. purpuratus oocyte contains on the average 1600 copies of each single-copy transcript. From the rate constants listed in the legend to Figure 2, we estimate that 11% of the short repeat ^{3}H-DNA reacts with RNA sequences present in an average of 1.5×10^5 copies per oocyte, 31% reacts with RNA sequences present in an average of 3.3×10^4 copies per oocyte, and 35% reacts with RNA sequences present in an average of 1.4×10^3 copies per oocyte. The prevalence of short repetitive sequence transcripts in the oocyte RNA may well form a broad unimodal distribution, and the average prevalence classes cited here merely provide a useful description of the data. Whatever the actual transcript frequency distribution, it is clear that some repetitive sequence families are represented in oocyte RNA in about the same number of copies as an average single-copy sequence transcript, while other repetitive sequences are present in up to 100 times this number of RNA trascripts per oocyte.

Figure 2 shows that the long repeat tracer (closed circles) hybridizes significantly less than does the short repeat tracer. The RNA/^{3}H-DNA mass ratio used in the high RNA Cot reactions was the same for both tracers. Were each transcript present in only 100 copies per oocyte (less than one tenth the prevalence of single-copy transcripts), this ratio would be sufficient to provide a 100 fold RNA sequence excess for a transcript 2000 nucleotides long which is complementary to a repeat occurring 100 times per genome. The failure of the long repeat tracer to react more than 33% at RNA Cot 55,000 therefore suggests that the oocyte RNA does not contain transcripts complementary to much of the long repeat sequence, unless their concentrations are less than a few percent of the concentration of single-copy transcripts. It cannot be determined from this experiment whether the long repeat DNA sequence which does react is represented asymmetrically in the RNA. In any case, the high prevalence component suggested by the kinetics of the short repeat reaction is not evident in the long repeat tracer reaction. The least-squares solution indicated by the solid line through the closed circles in Figure 2 suggests that

13% of the long repeat tracer hybridizes with RNA transcripts present in an average of 3×10^4 copies per oocyte, and 19% hybridizes with RNA transcripts present in an average of 1100 copies per oocyte.

Several repetitive genes are known to be represented in oocyte RNAs, such as the histone genes and the ribosomal RNA genes. These genes would be included in the long repeat DNA fraction. The aggregate complexity of the known repetitive gene sets, however, is much lower than that of the long repeat fraction. Eden et al. (1977) estimated the complexity of the latter to be at least 6×10^5 nucleotide pairs. Highly prevalent oocyte RNA transcripts such as histone mRNA and rRNA clearly cannot account for more than a few percent of the long repeat tracer reaction because the observed kinetics are several orders of magnitude too slow. Furthermore, much of all of the long repetitive sequence hybridization could be due to contaminating short repeats. Whatever the nature of the hybridized fragments in the long repeat tracer, the clear differences between the long and short repeat tracer reactions with oocyte RNA provide additional evidence that these are at least partially distinct repetitive sequence sets.

Hybridization of Cloned Repetitive Sequences with Oocyte RNA

The availability of cloned repetitive sequences offers an independent opportunity to confirm the main conclusions drawn from the experiments of Figure 2. Each cloned repeat represents a single repetitive sequence family. According to the data shown in Figure 2, different repeat families should be represented in oocyte RNA by transcripts differing widely in prevalence. Furthermore, both complementary strands of many (or all) of the short repeat families should be represented in the RNA.

Construction of the repetitive sequence clones used for these experiments has been described previously (Scheller et al., 1977). In brief, S. purpuratus DNA was renatured and the repetitive sequence duplexes were isolated after S1 nuclease treatment. These fragments were ligated into the Eco R1 endonuclease recognition sites of plasmid RSF2124 with the aid of synthetic double-stranded "linkers" containing Eco R1 sites. Most of the cloned repetitive sequence inserts are a few hundred nucleotides in length, as are the majority of the repetitive sequences in the genome, while a few (including an 1100 nucleotide sequence used in the present experiments) are significantly longer (Klein et al., 1978). A set of nine of these cloned repeats was used in the present experiments, and the representation of the same nine sequences in sea urchin nuclear RNAs is the subject of the accompanying paper by Scheller et al. (1978). The tracer derived from each clone was reacted with oocyte RNA to determine the prevalence of the complementary transcripts.

Two procedures were used to measure transcript prevalence. First, the strand-separated repeat fragments were reacted with excess oocyte RNA, and the prevalence of complementary transcripts was calculated from the kinetics of the hybridization reactions. These calculations were carried out according to equation (1) in the legend to Table 1. The rate constants for the reactions of the cloned tracers with the oocyte RNA and the calculated number of transcripts per oocyte are listed in Table 1. A second method was to react excess strand-separated tracer with increasing amounts of oocyte RNA. These reactions were carried out to more than 10 times $Cot_{1/2}$ with respect to the ^{32}P-DNA tracer fragments, thus ensuring termination of the reactions. With this method, referred to here as "titration," the fraction of oocyte RNA complementary to each cloned tracer fragment is measured by the fraction of the tracer reacting as the RNA/^{32}P-DNA ratio is increased. The calculation of transcript prevalence from titration data was carried out as described in the accompanying paper by Scheller et al. (1978). Thus the number of copies of RNA transcripts complementary to each cloned sequence in the oocyte was estimated with the aid of least-squares methods by application of their equations (2), (3) and (4). The numerical parameters used in these calculations are given in the legend to Table 1. Scheller et al. (1978) demonstrated that the kinetic and titration methods generally agree within a factor of 2-3, which, as they discuss, is within expectation considering the various errors to which each method is subject. In the present study, we applied the titration method mainly to rare RNA transcripts, thereby avoiding the difficulty of achieving the RNA sequence excess needed for kinetic determinations.

RNA excess hybridization kinetics are shown for three of the cloned repeats in Figure 3. The repeat sequence represented by the most highly prevalent transcripts is that carried in clone 2109B (open circles), since its reaction occurs most rapidly. On the basis of the rate of the reaction of the clone 2109B upper strand repeat fragment, we estimate that there are $\sim 8.3 \times 10^4$ complementary transcripts per oocyte. The lower strand reacts with similar kinetics (Table 1). The complementary strands of the clone 2090 repeat (Figure 3, open and solid triangles) also react at approximately the same rate as each other. The kinetics of these reactions, however, show that the prevalence of

Table 1. Number of Transcripts Complementary to Cloned Repetitive Sequences in Oocyte RNA

				Number of Transcripts per Oocyte			
				RNA Excess Kinetics		Titrations	
Clone	Strand[a]	Length[b] (NTP)	Genomic Repetition Frequency[c]	Second-Order Rate Constant[d] (k_2)	Transcripts per Oocyte[e]	RNA Fraction[f]	Transcripts per Oocyte[e]
2007	U	1100	400				(7000)[a]
	L					1.5×10^{-6}	7000
2034	U	560	1000			3.6×10^{-7}	3300
	L					5.8×10^{-7}	5300
2090	U	220	140	1.7×10^{-3}*	35,000		
	L			1.7×10^{-3}*	35,000	2.5×10^{-6}‡	57,000
2101	U	320	700			5.2×10^{-7}†	8400
	L					1.9×10^{-7}†	3000
2108	U	190	20	3.8×10^{-3}	79,000	5.8×10^{-6}†	160,000
	L			1.3×10^{-3}	27,000	3.2×10^{-6}†	90,000
2109A	U	200	900	1.6×10^{-4}	2,500		
	L			2.0×10^{-4}	2,800		
2109B	U	125	1000	6.0×10^{-3}*	83,000		
	L			6.0×10^{-3}	83,000		
2133B	U	310	60			4.1×10^{-7}	5800
	L					3.4×10^{-7}	5800
2137	U	190	530	9.0×10^{-4}	19,000	6.5×10^{-7}‡	18,000
	L			7.8×10^{-4}	16,000		

* Data shown in Figure 3.
† Data shown in Figure 4.
‡ Data shown in Figure 7.

[a] The two complementary strands of each cloned repetitive sequence were designated "upper" and "lower," according to their electrophoretic mobility on polyacrylamide gels after denaturation (Scheller et al., 1978). The lower strand is the strand which migrates more rapidly.

[b] The length of each cloned repetitive sequence was estimated from electrophoretic mobility of the duplex fragment on agarose gels, relative to reference fragments of known molecular weight (Klein et al., 1978).

[c] Genomic repetition frequency was estimated by Klein et al. (1978) from the kinetics of reassociation of each cloned sequence with excess sea urchin DNA. The numbers shown are the occurrences of each sequence per haploid genome.

[d] Each strand was hybridized with excess oocyte RNA as described in the text. Second-order rate constants, in units of M^{-1} sec^{-1}, were extracted from the kinetic data by the least-squares method described by Pearson, Davidson and Britten (1977).

[e] The number of RNA copies (or RNA transcripts complementary to a cloned DNA sequence) per oocyte was calculated by comparing the observed hybridization kinetics with the kinetics of hybridization of single-copy ^{3}H–DNA with excess oocyte RNA, as measured by Hough-Evans et al. (1977). This calculation was carried out by means of equation (1) in the accompanying paper by Scheller et al. (1978). Insertion of the appropriate numerical values yields:

$$\text{transcripts per oocyte} = \frac{k_2}{2.3 \times 10^{-4}}(1600)(f_L) \qquad (1)$$

where k_2 is the observed rate constant for the hybridization of the cloned DNA sequence with excess oocyte RNA from column 5 of this table; 2.3×10^{-4} M^{-1} sec^{-1} is the first-order rate constant for the hybridization of 250 nucleotide single-copy DNA with excess oocyte RNA (Hough-Evans et al., 1977); and 1600 is the average number of RNA transcripts of each single-copy sequence per oocyte, determined by Hough-Evans et al. (1977). f_L is a factor which corrects for the rate retardation observed when the length of the driver (the RNA in this case) exceeds the length of the tracer. f_L is approximately equal to $(L_{driver}/L_{tracer})^{1/2}$ (Chamberlin et al., 1978). We have found (data not shown) that the amount of RNA degradation during incubations of several hours (corresponding to an RNA Cot of several thousand) is minimal. From the experiments of Figure 7, it appears that the lengths of RNA molecules reacting with four of the repetitive clones (2101, 2108, 2090 and 2137) average ~2000 nucleotides. Thus the RNA driver length in these hybridizations exceeds the DNA tracer length by a factor of ~10, and a value for f_L of 3.0 has been applied. For sequences whose complementary transcript length has not been measured, we use a conservative value of $f_L = 2$. An additional minor kinetic uncertainty derives from the sequence mismatch with characterizes the duplexes formed by reacting the cloned repeats with other genomic (or RNA) copies of the sequence. This uncertainty is discussed in more detail in the accompanying paper by Scheller et al. (1978). The intrafamilial sequence divergence of these cloned repetitive sequence families was measured by Klein et al. (1978), and is listed in Table 2 of Scheller et al. (1978).

The transcript copy number per oocyte can also be calculated directly, without reference to the single-copy sequence concentration, by the formula

$$\text{transcripts per oocyte} = \frac{k_{2\,observed}}{k_{2\,estimated}} \frac{(3 \times 10^{-9})(6.02 \times 10^{23})}{(350)(L)} \qquad (2)$$

Table 1 —*continued*

where $k_{t\,equivalent}$ is the rate constant expected if the RNA consisted only of the sequence in question, calculated as

$$k_{t\,equivalent} = 118\ M^{-1}\ sec^{-1}\ \frac{5375}{L}\ \frac{1}{f_L} \qquad (3)$$

In these equations, L is the fragment length or complexity of the tracer, 3×10^{-9} is the number of grams of total RNA per oocyte, 350 is the average molecular weight of a ribonucleotide and 118 M^{-1} sec^{-1} is the rate constant measured (Galau et al., 1977) for the second-order reaction of φX174 RF DNA, the complexity of which is 5375 nucleotides. Since this is essentially the method used to calculate the number of single-copy sequence transcripts per oocyte (Galau et al., 1974, 1976; Hough-Evans et al., 1977), calculations by equation (1) and by equations (2) and (3) are equivalent.

[f] The RNA fraction, or the mass fraction of oocyte RNA complementary to a particular cloned DNA sequence, was measured by titration as illustrated in Figure 4. The RNA fraction is equal to the initial slope of the titration curve (that is, the mass of ^{32}P-DNA hybridized divided by the mass of RNA in the reaction), or can be extracted from the function shown in equation (2) of Scheller et al. (1978). RNA fractions shown were obtained by least-squares methods as described by Scheller et al. (1978).

[g] The number of RNA copies (or RNA transcripts complementary to a particular cloned DNA sequence) per oocyte is calculated from the RNA fraction as:

$$\text{transcripts per oocyte} = \frac{(\text{RNA fraction})(3 \times 10^{-9})(6.02 \times 10^{23})}{(L)(350)} \qquad (4)$$

where 3×10^{-9} is the total number of grams of RNA per oocyte, L is the length of the cloned sequence in nucleotides and 350 is the average molecular weight of a ribonucleotide residue. This is essentially equation (4) of Scheller et al. (1978).

[h] The clone 2007 repeat, upper strand, was not titrated with oocyte RNA. Kinetics of incomplete RNA excess reactions with both strands of the clone 2007 fragment indicate that oocyte RNA contains a similar concentration of transcripts complementary to each strand.

transcripts complementary to the clone 2090 repeat is about 3 fold lower than that of transcripts complementary to the clone 2109B repeat. In addition, Figure 3 indicates that transcripts complementary to the upper strand (open squares) and lower strand (solid squares) of the clone 2101 repeat sequence are present at somewhat different concentrations in the oocyte RNA. These transcripts are so rare in the oocyte that the RNA/^{32}P-DNA ratio used (5×10^{0}) was insufficient to permit complete reaction.

Titration curves for the clone 2101 and 2108 repeats are shown in Figure 4. From these data, we estimate that the upper strand (open circles) of the clone 2108 repetitive sequence is represented in RNA by $\sim 1.6 \times 10^{3}$ complementary transcripts per oocyte, while the lower strand (closed circles) is about 2 fold less prevalent (Table 1). The titration data shown in Figure 4B indicate that transcripts complementary to the clone 2101 repeat are represented only a few thousand times per oocyte, as summarized in Table 1. This result is consistent with the incomplete termination of the RNA excess hybridization reaction shown for this repeat fragment (Figure 3). While the titration experiments of Figure 4B show that both strands of the clone 2101 repeat are indeed represented in the oocyte RNA, it again appears that the complementary transcripts are present at concentrations which differ 2-3 fold. As a demonstration that this is not due to differential reactability of the upper and lower strand tracers, both tracers were titrated with increasing quantities of genomic DNA. This experiment is shown in Figure 4C. Data for the two tracers are essentially identical, and it follows that the differences in the titration curves shown in Figure 4B are due to differences in RNA transcript prevalence.

Table 1 shows that the number of transcripts per oocyte complementary to nine individual repetitive sequences varies from a few thousand to about 10^{5} — that is, the different repetitive sequence families appear to be expressed to very different extents as suggested by the experiments with the genomic repeat tracers (Figure 2). The oocyte contains different numbers of transcripts complementary to the two strands of some repetitive sequences (clones 2108, 2101 and 2034), but approximately equal amounts of transcripts complementary to the two strands of other repetitive sequences (clones 2109A, 2109B, 2090, 2137, 2133B and 2007). An important point is that neither strand of any of the nine repetitive sequences examined was unrepresented in oocyte RNA. Transcripts of the two longest cloned repetitive sequences (clones 2007 and 2034) are relatively rare in oocyte RNA. Were these clones typical members of the long repeat fraction studied above, however, our results would imply that both strands of long repetitive sequences are represented in the oocyte RNA, as are both strands of short repetitive sequences. This conclusion should be regarded as tentative until additional measurements on more long repetitive sequences are available.

No relation is apparent in Table 1 between the genomic repetition frequency and the transcript prevalence in oocyte RNA. For example, the clone 2108 sequence is present in only about 20 copies per haploid genome, although its complementary transcripts are highly prevalent in oocyte RNA; on the other hand, the clone 2101 sequence has a genomic repetition frequency of about 700, while its complementary transcripts are rare in oocyte

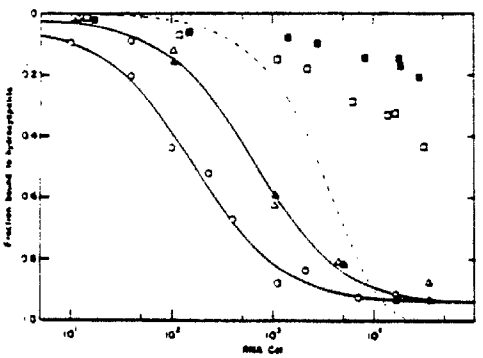


Figure 3. Kinetics of the Reactions of Cloned Repetitive DNA Sequences with Excess Oocyte RNA

The separated strands of cloned repetitive sequences were hybridized with a 1–5 × 10^6 mass excess of oocyte RNA at 55°C in 0.41 M PB, 0.2% SDS, 0.005 M EDTA. The cloned tracers were terminally labeled with ^{32}P. Hybridization was assayed by binding to hydroxyapatite at the indicated values of RNA Cot. The observed hydroxyapatite binding was ≥95% sensitive to low salt RNAase treatment (Galau et al., 1974) and therefore represents RNA-DNA hybrid formation rather than DNA renaturation. The solid lines are second-order functions fit to the data by a least-squares procedure (Pearson et al., 1977), assuming a single kinetic component in each case. For comparison, the dashed line represents the kinetics of hybridization of single-copy DNA with excess oocyte RNA normalized to 100% reaction. This reaction is pseudo-first-order in form, and the rate constant (k_1) is 2.3 × 10^{-4} M^{-1} sec^{-1} (Hough-Evans et al., 1977). Clone 2109B upper strand (○) hybridizes with a second-order rate constant (k_2) of 6 × 10^{-3} M^{-1} sec^{-1}; for clone 2090, upper strand (△) and lower strand (▲), k_2 = 1.7 × 10^{-2} M^{-1} sec^{-1}. Clone 2101 upper strand (□) and lower strand (■) hybridize incompletely at the oocyte RNA/^{32}P-DNA mass ratios attainable in these experiments because the tracer is in sequence excess (see text). The observed fraction of ^{32}P-DNA bound to hydroxyapatite was normalized by the fraction of ^{32}P in each cloned DNA preparation which is capable of reassociating with excess sea urchin DNA (that is, the tracer reactivity). This fraction ranged from 0.6 to 0.8 for the tracers shown. The nonreactive ^{32}P was mainly γ-^{32}P-ATP persisting from the polynucleotide kinase labeling reactions.

RNA. This observation suggests that the variation in oocyte RNA prevalence among different repetitive sequences results not simply from the variation in their genomic repetition frequencies, but from the greater expression of certain repetitive sequence families. To examine this question further, the fraction of the short repetitive ^3H-DNA tracer which is represented by highly prevalent oocyte RNA transcripts was isolated, and the genomic repetition frequencies of these sequences were measured.

Genomic Repetition Frequency of Short Repeats Whose Transcripts Are Prevalent in Oocyte RNA

A subfraction of the short repeat tracer that is complementary to the more prevalent RNA transcripts was isolated by two cycles of hybridization with excess oocyte RNA at Cot 500 (see Figure 2). The procedure used and the yields at each step are given in the legend to Figure 5. The selected ^3H-DNA was 19% of the starting repeat tracer. The kinetics of the reaction of this selected set of repeats with excess oocyte RNA is shown in Figure 5 (open circles). For comparison, the dashed line in Figure 5 shows the reaction of the unfractionated short repeat tracer with oocyte RNA, reproduced from Figure 2. The selected tracer clearly consists of sequences whose complements are relatively prevalent in the RNA. A least-squares analysis of the data indicates that the selected repeats are represented by an average of ~1.1 × 10^6 complementary transcripts of each sequence per oocyte. Comparison with the data in Figure 2 indicates that this transcript prevalence is consistent with expectation, if the selected tracer indeed consists of the most highly represented 19% of the short repeat ^3H-DNA. The kinetics of the reaction shown in Figure 5 indicate that in contrast to the starting tracer, very little of the selected repeat ^3H-DNA is complementary to RNAs of lower prevalence.

The selected repeat tracer was reacted with excess sheared sea urchin DNA, as shown in Figure 6 (open circles). For comparison, the reassociation of the starting short repeat tracer with excess sea urchin DNA is also shown (closed circles). The reaction of the selected short repeat fraction is only slightly faster than the reaction of the original unfractionated short repeat preparation. Were prevalence in oocyte RNA directly proportional to genomic repetition frequency, the selected ^3H-DNA fraction would consist of the most highly reiterated 19% of the short repeats—that is, of sequences occurring several thousand times per haploid genome. The kinetics of the reaction shown in Figure 6, however, indicate a different result. The data are consistent with a distribution of genomic repeat frequencies such that of the 96% of the selected repeat tracer which is reactive, 19% consists of repeats occurring only 20 times per haploid genome, and 33% consists of repeats occurring an average of 250 times per genome. Of the *different* repetitive sequence families represented in the oocyte by prevalent RNA transcripts, the vast majority belong to the lower repetition frequency classes, since there can be very few *different* highly repetitive sequences. In other words, most of the highly expressed repetitive sequences families are of relatively small size. There are approximately 5 × 10^3 of these families in the S. purpuratus genome. Since the selected tracer contains slightly less than 20% of the small repeat families, we estimate that several hundred to a thousand of

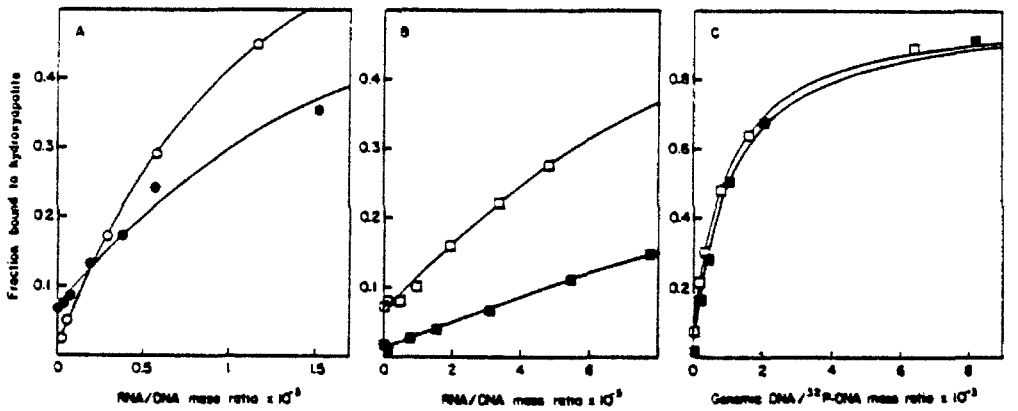


Figure 4. Titration of Cloned Repetitive Sequences with Sea Urchin Oocyte RNA or Sea Urchin DNA

(A) The ^{32}P-DNA repeat tracer of clone 2108, upper strand (○) or lower strand (●), was hybridized with increasing amounts of total oocyte RNA. Reactions were carried out to ≥10 × $Cot_{1/2}$ with respect to the cloned ^{32}P-DNA tracer. Since only one of the two complementary strands of the clone 2108 sequence is present in each hybridization mixture, DNA renaturation cannot occur. The fraction of ^{32}P-DNA hybridized with oocyte RNA was assayed by binding to hydroxyapatite. The small fraction of ^{32}P-DNA bound to hydroxyapatite in the absence of RNA, indicated by the ordinate intercept, is due to minor contamination of the strand-separated DNA preparation with the complementary strand (see the accompanying paper by Scheller et al. (1978) for data regarding the purity of the strand-separated tracers). The increase in hydroxyapatite binding with increasing RNA/DNA ratio is due to RNA-DNA hybridization, and is completely sensitive to RNAase digestion under low salt conditions. The initial slopes of the curves shown are proportional to the fraction of oocyte RNA capable of hybridizing with each DNA strand. The solid lines are fit to the data according to equation (2) of Scheller et al. (1978).

(B) ^{32}P-DNA repeat tracer of clone 2101, upper strand (□) or lower strand (■), was hybridized with varying amounts of oocyte RNA, as in (A).

(C) The repeat tracer of clone 2101, upper strand (□) or lower strand (■), was renatured with increasing amounts of unfractionated sea urchin DNA sheared to a weight mean of single-strand fragment length of 650 nucleotides. The fraction of ^{32}P-DNA binding to hydroxyapatite is plotted as a function of the sea urchin DNA/^{32}P-DNA mass ratio. Taking into account the effects of the driver DNA fragment length distribution relative to the length of the reacting sequence (Moore et al., 1978), the reiteration frequency measured in this experiment is in satisfactory agreement with the value of 700 copies per haploid genome reported by Klein et al. (1978).

them are represented by prevalent transcripts in the oocyte RNA.

Size of RNA Molecules Containing Short Repetitive Sequence Transcripts

Oocyte RNA was fractionated according to length, and the concentrations of transcripts complementary to several of the cloned repetitive sequences were measured in three RNA size fractions. The RNA was denatured by treatment with 80% (v/v) dimethyl sulfoxide (DMSO) at 55°C and sedimented through sucrose gradients containing 60% DMSO at 25°C. These conditions are sufficient to disrupt most inter- or intramolecular base pairing in the RNA (Strauss, Kelly and Sinsheimer, 1968; Bantle and Hahn, 1976). The sedimentation profile of the oocyte RNA in the DMSO gradient is shown in Figure 7A. Most of the absorbance pattern is due to ribosomal RNA. Fractions were pooled into three size classes, as indicated: size class I contains RNA sedimenting between 19.5S and approximately 36S; RNA in size class II sediments between 11S and 19.5S; RNA in size class III sediments between approximately 2S and 11S. Four of the cloned repeat fragments included in the experiments summarized in Table 1 were used for these experiments. Figures 7B–7E show measurements by the titration method of the relative concentration of the transcripts complementary to these repeats in each RNA size class. As the legend to Figure 7 indicates, the total numbers of transcripts in the three size classes agree reasonably well with the measurements listed in Table 1. The relative distribution of transcripts complementary to each sequence among the three RNA size classes is represented by the inset histograms in the figure.

Figure 7 shows that each of the four cloned repetitive sequences examined hybridizes the most with RNA of size class II, less with size class I and still less with size class III. Thus most of the oocyte RNA molecules containing these short repetitive sequence transcripts appear to fall in the 11S–19.5S fraction – that is, to be ~1000 to 2500 nucleotides in length. The same result is obtained with a rare set of transcripts, those complementary to the clone 2101 repeat, as with prevalent transcripts, such as those complementary to the clone 2108 repeat. The sedimentation profile of the ribosomal RNA in Figure 7A suggests that the oocyte RNA is slightly degraded. Our measurements may there-

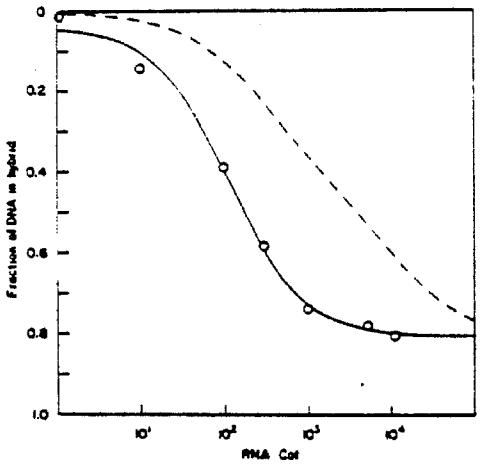


Figure 5. Hybridization of a Selected Fraction of Short Repetitive ^{3}H-DNA with Excess Oocyte RNA

The short repeat ^{3}H-DNA tracer was hybridized with a 3 × 10^6 fold excess of oocyte RNA to RNA Cot 500 and fractionated by binding to hydroxyapatite. 37% of the ^{3}H-DNA was bound, 30% was in RNA-DNA hybrid and 7% was renatured DNA-DNA duplex. The bound material was eluted, hybridized with additional oocyte RNA to RNA Cot 500 and again fractionated on hydroxyapatite, yielding 57.5% bound. This second bound fraction was eluted, and the ^{3}H-DNA was purified of RNA by alkaline hydrolysis (0.1 N NaOH at 37°C for 1 hr) followed by chromatography on a Sephadex G-100 column. The fraction binding to hydroxyapatite at Cot 10^{-4} (12%) was removed, and the remainder of the ^{3}H-DNA, representing 19% of the starting short repeat tracer, constituted the selected short repeat fraction. This ^{3}H-DNA was hybridized with a 10^6 fold mass excess of oocyte RNA to the indicated values of RNA Cot. The fraction of the ^{3}H-DNA tracer in hybrids (○) was assayed by binding to hydroxyapatite and corrected for a small amount (0–5%) of DNA-DNA renaturation, as described in Experimental Procedures. The solid line shows the least-squares solution assuming a single second-order component. The second-order rate constant is 7.9 × 10^{-3} M^{-1} sec^{-1}. The dashed line represents the kinetics of hybridization of the starting short repeat tracer with excess oocyte RNA, reproduced from Figure 2.

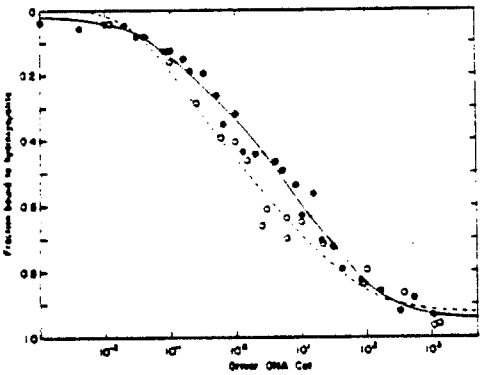


Figure 6. Reassociation of Total Short Repetitive ^{3}H-DNA and Selected Short Repetitive ^{3}H-DNA with Excess Whole Sea Urchin DNA

The total short repeat ^{3}H-DNA preparation (●), or that fraction of the short repeat ^{3}H-DNA tracer selected by low Cot hybridization to oocyte RNA (○) (see Figure 5), was reacted with a 1–5 × 10^6 fold mass excess of total sea urchin DNA that was sheared to a weight mean single-strand fragment length of 650 nucleotides. The ordinates show the fraction of ^{3}H-DNA binding to hydroxyapatite at 50°C in 0.12 M PB, 0.05% SDS. The Cot is calculated with respect to the concentration of the genomic driver DNA. Included as an internal kinetic standard in each reaction mixture was a ^{32}P-labeled, cloned repetitive sequence (clone 2034; data not shown). The rate of reassociation of this ^{32}P-DNA sequence with the driver DNA was for all reaction mixtures consistent with the 1000 fold reiteration frequency measured previously for this sequence (Klein et al., 1978). The solid line shows a least-squares solution for the reaction of the unfractionated short repeat tracer, assuming three second-order kinetic components. This solution uses the three rate constants measured by Graham et al. (1974) for the repetitive components of whole S. purpuratus DNA: 8.2 M^{-1} sec^{-1} for the fastest component (6500 occurrences per haploid genome), 0.3 M^{-1} sec^{-1} for the middle component (250 copies per genome) and 0.023 M^{-1} sec^{-1} for the slowest component (20 copies per genome). The fractions of the short repeat tracer in the three components are, respectively, 0.26, 0.35 and 0.31 (6% of the ^{3}H-cpm are nonreactive and 2% bind at Cot <10^{-5}). The dashed line is the least-squares solution for the kinetics of the reaction of the selected short repeat fraction represented by prevalent oocyte transcripts using the same three second-order rate constants. The fractions of the selected short repeat tracer included in the three components are 0.44, 0.32 and 0.18, respectively (4% of the ^{3}H-cpm are nonreactive).

fore underestimate slightly the true size of the oocyte RNA molecules bearing the repetitive sequence transcripts. In any case, it is clear that few RNA molecules containing repetitive sequence transcripts sediment at less than 11S. We conclude that the short repeat transcripts are present on oocyte RNA molecules which are many times the length of the cloned repetitive sequence elements. Since the messenger RNAs of the oocyte are about 2000 nucleotides long (reviewed in Davidson, 1976), the observations in Figure 7 raise the possibility that the short repeat transcripts might be covalently linked to maternal mRNA molecules.

Genomic Sequence Organization around Repeat Families Represented by Prevalent RNA Transcripts

While many of the repetitive sequences in the sea urchin genome are interspersed with single-copy DNA sequences, some occur in long blocks uninterrupted by single-copy sequence (Graham et al., 1974; Eden et al., 1977). The purpose of the following experiments was to determine whether repetitive sequences represented by high abundance oocyte transcripts were interspersed with single-copy DNA sequences in the genome. Alternatively, they could be organized as repetitive sequence clusters. Sea urchin ^{3}H-DNA of starting length 3300 nucleotides was stripped of the "foldback" fraction by hydroxyapatite binding at Cot 5 × 10^{-4}. After this procedure, its weight mean single-strand fragment length was 2600 nucleotides. The tracer was

Cell
182

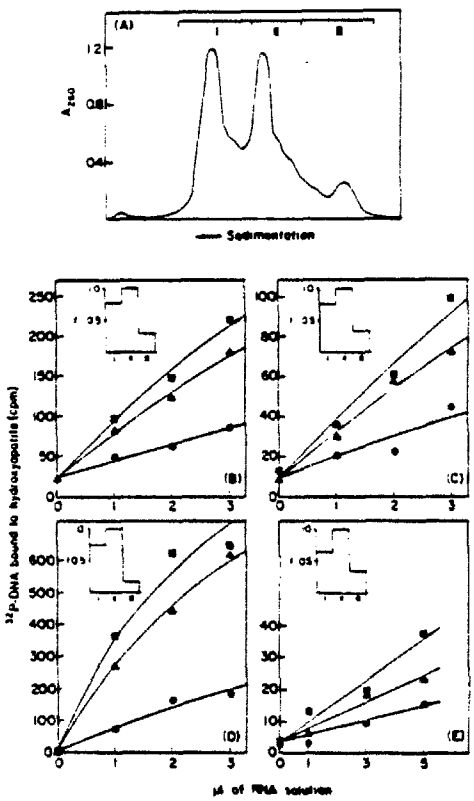


Figure 7. Size of Oocyte RNA Molecules Hybridizing with Cloned Repetitive Sequences

(A) Fractionation of oocyte RNA by sedimentation in denaturing DMSO-sucrose gradients. Total oocyte RNA was dissolved at 1.25 mg/ml in 80% (v/v) dimethyl sulfoxide (DMSO), 0.1 M LiCl, 0.01 M Tris–HCl (pH 6.5), 0.005 M EDTA, 0.2% SDS, and heated at 55°C for 5 min. It was then sedimented through 4–20% exponential sucrose gradients containing 60% DMSO at 25°C, 35,000 rpm for 66 hr in the Beckman SW41 Ti rotor. Absorbance at 260 nm is plotted as a function of distance, after subtraction of background absorbance due to DMSO and EDTA in the gradient. The three major ribosomal RNA peaks, from right to left, contain 5S and 5.8S RNA, 18S RNA and 28S RNA, respectively. Fractions were pooled into three size classes, I, II and III, as indicated. Size class I contains RNA sedimenting between 19.5S and about 36S; size class II contains RNA sedimenting between 11S and 19.5S; size class III contains RNA sedimenting between about 2S and 11S.

(B) Titration of the clone 2090 repeat fragment, lower strand, with the three oocyte RNA size classes. RNA in each of the three size classes described in (A) was concentrated to a volume of 0.1 ml. 0, 1, 2 or 3 μl of fraction I, II or III were mixed with 2000 cpm of the lower strand fragment of the clone 2090 repeat (^{32}P-DNA spec. act. 8.8 × 10^8 cpm/μg). The hybridization mixtures contained 0.41 M PB, 0.2% SDS and 0.005 M EDTA. They were heated at 105°C for 45 sec and incubated at 55°C to a terminal DNA Cot. The amount of ^{32}P-DNA in hybrid was assayed by binding to hydroxyapatite and is plotted as a function of the volume of RNA used in each hybridization mixture [(▲) size class I; (■) size class II; (●) size class III]. The small amount of ^{32}P-DNA binding to hydroxyapatite in the absence of added RNA is due to minor contamination of the lower strand of the clone 2090 repeat fragment with the complementary upper strand. The increase in ^{32}P-DNA binding to hydroxyapatite with added RNA is due to the formation of oocyte RNA–^{32}P-DNA hybrids and is labile to low salt RNAase treatment. The graph shows the relation between the amount of the ^{32}P-DNA hybridized and the amount of each RNA size class in the reactions. The number of RNA transcripts complementary to the L strand of the clone 2090 repeat in each size class is calculated from the mass of ^{32}P-DNA hybridized by a given volume of RNA solution in the initial linear region of the curve (that is, when the ^{32}P-DNA is in large excess). The inset shows the relative number of complementary transcripts in each RNA size class as a fraction (f) of the number of transcripts in size class II. The total number of complementary transcripts per oocyte can be estimated from the total amount of hybridization with the three RNA size classes, taking into account the quantity of RNA contained in the denaturing gradients (785 μg), the fraction of each size class used in the experiment and the amount of RNA per oocyte (3 × 10^{-9} g). The number obtained is 57,000 complementary transcripts per oocyte. (C) Titration of the upper strand of the clone 2137 repeat with the three size classes of oocyte RNA. The clone 2137 upper strand tracer was titrated with the three RNA size classes as in (B), and the data are similarly displayed. The ^{32}P-DNA had a specific activity of 8.4 × 10^8 cpm/μg, and 1500 cpm were used in each hybridization mixture. The number of transcripts complementary to this sequence per oocyte, estimated by summing the three size classes, is 18,000.

(D) Titration of the upper strand of the clone 2108 repeat with the three size classes of oocyte RNA. The titration was performed as described above. 1700 cpm of ^{32}P-DNA (spec. act. 8 × 10^8 cpm/μg) were used in each hybridization reaction. The data indicate that the total number of transcripts complementary to the clone 2108 repeat upper strand is 320,000 per oocyte.

(E) Titration of the upper strand of the clone 2101 repeat with the three size classes of oocyte RNA. The titration was performed as above, except that 0, 1, 3 or 5 μl of each RNA size class was used. Each hybridization contained 900 cpm of ^{32}P-DNA (spec. act. 3.2 × 10^8 cpm/μg). The total number per oocyte of transcripts complementary to the clone 2101 upper strand tracer was 9000.

then reacted with excess oocyte RNA. The kinetics of this reaction are shown in Figure 8 (closed circles). For comparison, the dashed curve in Figure 8 shows the fraction of total DNA consisting of hybridized repetitive sequences at various RNA Cots, according to Figure 2. At each RNA Cot, the fraction of the 2600 nucleotide DNA fragments bound to hydroxyapatite greatly exceeds the fraction of the DNA included in hybridizing repetitive sequences (dashed line). On the other hand, only 3% of single-copy DNA sequences can hybridize with oocyte RNA, and these sequences react with an RNA $Cot_{1/2}$ of 3000 (Hough-Evans et al., 1977). Essentially all of the reaction seen in Figure 8 below RNA Cot 1000 must therefore be due to repetitive sequence hybridization. It follows that the hybridizing repetitive sequences on most of the bound 2600 nucleotide fragments are interspersed with other, nonhybridized sequences. The large fraction of the tracer involved in the reaction shown in Figure 8 requires that much of the interspersed flanking sequence be single-copy.

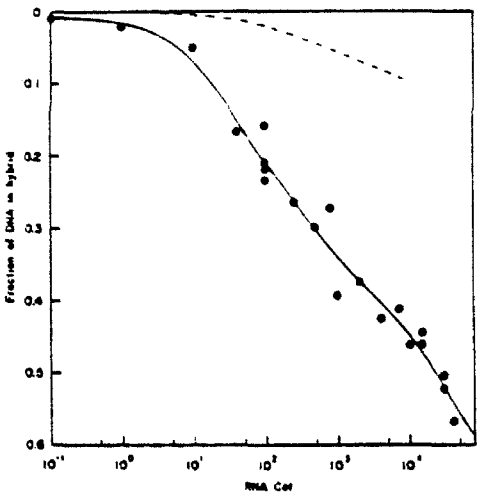


Figure 8. Hybridization of 2600 Nucleotide Long ^{3}H-DNA with Excess Total Oocyte RNA

A ^{3}H-DNA tracer of weight average single-strand fragment length 2600 nucleotides, from which the fraction binding to hydroxyapatite at Cot 5 × 10^{-4} (20%) had been removed, was hybridized with a 10^6–10^8 fold excess of oocyte RNA. The fraction of ^{3}H-DNA bound to hydroxyapatite has been corrected, as described in Experimental Procedures, for any self-reaction of the DNA (≤5%). The solid line is a least-squares solution to the hybridization data assuming three second-order kinetic components. The dashed line represents the hybridization of repetitive DNA with the oocyte RNA, expressed here as the fraction of whole sea urchin DNA. This quantity, H_R, was calculated at various values of RNA Cot from the data in Figure 2, using the relation $H_R = (H_S \cdot F_S) + (H_L \cdot F_L)$, where H_S and H_L are the fractions of short and long repetitive DNA hybridizing with oocyte RNA at a given RNA Cot, and F_S and F_L are the fractions of whole sea urchin DNA consisting of short and long repetitive sequences (0.13 and 0.08, respectively).

The sequence organization of the hybridized 2600 nucleotide tracer fragments was further investigated as follows. That fraction of the 2600 nucleotide tracer reacting with RNA by Cot 100 was bound to hydroxyapatite and eluted without denaturation in 0.5 M phosphate buffer. The average fraction of each fragment in RNA-DNA hybrid was 10%, as measured by S1 nuclease resistance. This corresponds to about one hybridized 260 nucleotide sequence per 2600 nucleotide fragment. The S1-resistant ^{3}H-DNA fraction was reassociated with excess total sea urchin DNA (see Figure 9, solid triangles), and as expected, it consists entirely of repetitive sequence. In addition, an aliquot of the 2600 nucleotide DNA which had reacted at RNA Cot 100 was reassociated with excess sheared sea urchin driver DNA. The kinetics of this reaction were assayed by S1 nuclease digestion to measure the amount of the 2600 nucleotide ^{3}H-DNA fragments in duplex regions as the reaction proceeded (Figure 9, open circles). A similar determination was carried out with the starting 2600 nucleotide ^{3}H-DNA tracer (Figure 9, closed circles). This experiment demonstrates that the hybridized and the unselected 2600 nucleotide tracers are essentially *the same in their content of repetitive and single-copy* sequence. Thus Figure 9 shows that an average of 25–30% of each fragment is repetitive sequence which reacts by driver DNA Cot 100, and approximately 70–75% is single-copy sequence. Since each selected tracer fragment includes a repetitive sequence complementary to prevalent RNA transcripts, it can be concluded that most members of the highly expressed repeat families are interspersed with single-copy sequences in the genome. The RNA transcripts could conceivably derive from only a few of the family members, however, and this demonstration does not necessarily mean that the particular family members which are transcribed are interspersed repetitive sequences.

Discussion

This investigation provides direct evidence that transcripts of at least some members of many interspersed short repetitive sequence families are accumulated during oogenesis. An unexpected conclusion is that all or almost all of the short repeat families in the genome are represented by complementary transcripts in the oocyte. We find, however, that different repeat families are represented to very diverse extents. Transcripts of certain repetitive sequence families are as much as 100 times more abundant than transcripts of other *repetitive sequence families*. This result is supported by data obtained with both cloned repetitive sequence tracers (Table 1) and with genomic repetitive DNA fractions (Figure 2), and by two independent methods of measurement, tracer excess titrations and RNA excess hybridization kinetics. There can be little doubt that there are highly prevalent transcripts in the oocyte which represent a minor fraction, perhaps 20%, of the repetitive sequence families in the genome, as well as other repeat transcripts about as rare as single-copy transcripts. Figure 6 shows that most of the different highly represented repeat families are not large. They typically include about 20 to about 200 copies per haploid genome. It is clear that the differences in repeat transcript prevalence are not a simple consequence of the genomic reiteration frequencies. There is, in other words, a specific set of repetitive sequence families which ends up being *highly expressed* in the maternal RNA. The mechanism leading to the specific patterns of transcript accumulation which we observe could include differences in transcription rate among di-

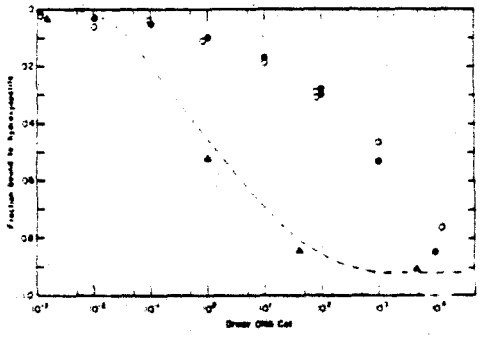


Figure 9. Single-Copy Sequence Content of 2600 Nucleotide Long ^{3}H-DNA Hybridizing with Oocyte RNA at RNA Cot 100

The 2600 nucleotide ^{3}H-DNA tracer was hybridized with excess oocyte RNA (as in Figure 8) to RNA Cot 100 and passed over hydroxyapatite. The bound fraction (25.6%) was eluted without denaturation in 0.5 M PB. A portion was treated with RNAase under low salt conditions (see Experimental Procedures) to digest the RNA from RNA-DNA hybrids and then passed over hydroxyapatite. The DNA which did not bind to this second hydroxyapatite column (84%, or 21.5% of the starting DNA) contains DNA sequences which had been in hybrids, plus covalently linked flanking sequences. This DNA fraction retained a weight mean fragment length of 2600 nucleotides according to alkaline sucrose velocity sedimentation measurements. The hybridized DNA fraction (○), or the starting 2600 nucleotide tracer (●), was reassociated with excess 450 nucleotide long sea urchin DNA. Each reaction mixture was treated with S1 nuclease to digest single-stranded regions of DNA, while sparing duplex regions, and fractionated by hydroxyapatite chromatography. The fraction of ^{3}H-DNA bound to hydroxyapatite at each Cot value represents the fraction of the 2600 nucleotide long ^{3}H-DNA that was in driver DNA-^{3}H-DNA duplex regions at that Cot. Also shown is the reassociation with excess whole sea urchin DNA of ^{3}H-DNA isolated from the hybrid regions formed by reacting the 2600 nucleotide ^{3}H-DNA and oocyte RNA to RNA Cot 100 (▲). This material was isolated by hybridizing 2600 nucleotide tracer with excess oocyte RNA to RNA Cot 100, digesting the mixture with S1 nuclease and binding the resistant fraction (2%) to hydroxyapatite. The bound ^{3}H-DNA was eluted with 0.5 M PB, digested with RNAase in low salt and purified of RNAase-resistant DNA-DNA duplexes by hydroxyapatite binding. The unbound tracer fraction was then reacted with excess whole sea urchin DNA and assayed by hydroxyapatite binding without S1 nuclease treatment as shown. The dashed line, for comparison, is the reaction of the selected short repetitive ^{3}H-DNA fraction with excess sea urchin DNA, reproduced from Figure 6.

verse repeat families, differences in the number of copies transcribed per family or differences in the transcript turnover rate. Oogenesis is lengthy and complex process, and little is known about the patterns of accumulation of any of the heterogeneous RNA species stockpiled in the mature oocyte.

It is interesting that both strands of each repeat sequence are represented in the oocyte RNA. In the accompanying paper, Scheller et al. (1978) demonstrate that complementary repeat transcripts also exist in nuclear RNAs. They also observe a similar specificity in the pattern of repetitive sequence representation in sea urchin nuclear RNAs, and show that a different set of repeats is highly represented in two different nuclear RNAs. Thus the findings we report here are not peculiar to the maternal RNA of the oocyte. The presence of both complementary strands of each repeat in the RNA raises the possibility that sequence-specific RNA-RNA or RNA-DNA interactions could occur. Since only a minor fraction of repeat families is highly represented in the oocyte RNA, the regions of the RNA or DNA affected by such interactions might be restricted in an interesting way. The possible regulatory significance of such interactions is discussed by Scheller and colleagues in the accompanying paper, and the reader is referred to their paper for speculations on this subject. The possibility that the oocyte RNA contains transcripts which participate in regulatory interactions is, of course, a fascinating one. The mechanism by which stored components from oogenesis act to set up the differential patterns of gene expression in the early embryo remains a conceptually challenging, and unsolved, problem of developmental biology (see review by Davidson, 1976).

It appears probable, although it is not demonstrated in this study, that the oocyte RNA molecules containing the repetitive sequence transcripts also contain interspersed single-copy sequence transcripts. We know from the data shown in Figure 7 that these RNAs are at least 1000–2000 nucleotides in length, significantly larger than the repeat sequences themselves. Furthermore, the experiment of Figure 9 shows that most members of the highly represented repeat families are short sequence elements interspersed among single-copy sequences in the genome. We lack direct evidence, however, that the *transcribed* sequences are interspersed; as pointed out earlier, the short repeat sequence set is represented in long repeat DNA preparations as well. It is not known whether this is due to homology between short and long repetitive sequences or merely to mutual contamination of the long and short repeat preparations. The RNA molecules hybridizing with the short repeats could conceivably be transcribed from long blocks of repetitive sequence which include one or more members of the interspersed repeat family. In this case, the hybridizing RNAs would consist mainly of repetitive sequence transcript, and the single-copy transcripts of the oocyte would exist on separate molecules. As yet there is no evidence to exclude this or other more complex models. If the repeat transcripts are linked covalently with single-copy transcripts, they are likely to be part of the stored matmRNA molecules. From the data of Figures 2 and 6, we estimate that the number of short repeat

transcripts in the oocyte is roughly equivalent to the number of single-copy transcripts in the oocyte. The functional polysomal mRNAs synthesized in the sea urchin embryo nuclei lack detectable covalently linked repetitive sequence transcripts (Goldberg et al., 1973), as is also the case for other organisms (for example, Campo and Bishop, 1974; Klein et al., 1974; Legler and Cohen, 1976). This raises the possibility that the matmRNA undergoes some form of processing during which repetitive sequence transcripts are removed prior to assembly of the message in embryo polysomes. Alternatively, some of the putative interspersed RNA molecules bearing both matmRNA sequences and repeat transcripts might perform some other function in the embryo.

The developmental role of the repetitive sequence transcripts in the oocyte is now a matter of great interest. Direct investigation of the origin, fate and sequence organization of these transcripts should begin to elucidate their significance.

Experimental Procedures

In Vivo Labeling and Isolation of Sea Urchin DNA

Sea urchin DNA was labeled in vivo by a modification of the method of Kleene and Humphreys (1977). Strongylocentrotus purpuratus embryos were cultured at 15°C in Millipore-filtered seawater containing penicillin and streptomycin (Hinegardner, 1967; Smith et al., 1974). At the 8 cell stage (~4.5 hr after fertilization), embryos were settled and resuspended at 7 × 10^6 embryos per ml with 50 μCi/ml ^{3}H-thymidine (90 Ci/mmole), 50 μCi/ml 2-^{3}H-adenosine (12.3 Ci/mmole) and 26 μCi/ml 5,6-^{3}H-uridine (35.4 Ci/mmole). At 12 hr after fertilization, the same amount of each isotope was again added to the culture. Embryos were grown to the hatched blastula stage and then harvested by centrifugation. They were washed once in seawater and once in SEDTA [0.1 M NaCl, 0.05 M EDTA (pH 8.0)] and then resuspended in a few milliliters of SEDTA. The resulting slurry was dripped from a pasteur pipette onto a bed of powdered dry ice, so that the drops froze immediately. The embryo-dry ice mixture was blended in a Waring blender to a fine homogeneous powder, allowed to thaw partially and then suspended in SEDTA (~1 ml per 10^8 orginal embryos). 0.5% SDS was added with stirring to complete lysis, and DNA was isolated by standard procedures including phenol-chloroform-isoamyl alcohol extraction, winding and RNAase A and pronase digestions (Graham et al., 1974; Angerer, Davidson and Britten, 1975). The DNA obtained had a specific radioactivity of 1.15 × 10^6 cpm/μg.

Preparation of Long and Short Repetitive DNA Fractions

Sea urchin ^{3}H-DNA in 0.2 M Na acetate was sheared in a Virtis 60 homogenizer (Britten, Graham and Neufeld, 1974) for 20 min at 10,500 rpm, 0°C. It was then passed through Chelex 100 (Bio-Rad) to remove any contaminating heavy metal ions. Following denaturation with 0.1 N NaOH for 10 min at 25°C, the DNA was reassociated in 0.41 M sodium phosphate buffer (PB), 1 mM EDTA (pH 6.8) at 60°C to Cot 40 (Britten et al., 1974). The DNA was dialyzed extensively to remove phosphate, heated briefly to 60°C in 0.3 M NaCl, 0.01 M PIPES (pH 6.7) and digested for 45 min at 37°C with S1 nuclease (Vogt, 1973) (10 μl/mg of DNA) in 0.15 M NaCl, 0.025 M Na acetate, 0.005 M PIPES, 0.1 mM $ZnSO_4$, 0.002 M β-mercaptoethanol (pH 4.4). The S1 nuclease conditions correspond to a DIG of 0.8, as defined by Britten et al. (1976). This amount of digestion is sufficient to remove single-stranded regions without destroying most repetitive sequence duplexes (Britten et al., 1976; Eden et al., 1977). Phosphate buffer was added to 0.12 M, and the DNA was extracted once with IAC (24:1 chloroform-isoamyl alcohol), diluted 2 fold with 0.12 M PB, 0.2% SDS and passed over hydroxyapatite at 60°C. The bound fraction was eluted with 0.5 M PB and constituted 21% of the input DNA.

This material was chromatographed on a column of Sepharose CL-2B (Pharmacia) in 0.12 M PB at 60°C. Excluded and included peak fractions were pooled, concentrated by extraction with sec-butanol (Stafford and Bieber, 1975), dialyzed into 0.12 M PB and rechromatographed as indicated in Figure 1. The resulting peak fractions were similarly pooled and concentrated. The long repeat fraction was sonicated with a Branson Model S125 sonifier and microtip for 6 × 20 sec at 2 Amp in 0.12 M PB. The resulting fragments averaged 570 nucleotides in length, according to measurements made by velocity sedimentation in alkaline sucrose gradients. To remove any contaminating single-copy DNA and unreactable DNA fragments, the DNA was denatured and renatured in 0.41 M PB, 5 mM EDTA, 0.2% SDS at 60°C to Cot 60, and passed over hydroxyapatite at 50°C in 0.12 M PB, 0.05% SDS. The bound fraction (60%) was eluted with 0.5 M PB and constitutes the long repetitive ^{3}H-DNA tracer used for the hybridization experiments. The DNA from the short duplex peak, consisting of short repetitive sequences averaging 305 nucleotides in length, was similarly renatured to Cot 20, and the fraction bound to hydroxyapatite (80%) was isolated (since the short repeats comprise about 13% of the DNA, this is comparable to a whole DNA Cot of about 150). Long and short repeat fractions were isolated from unlabeled sea urchin DNA in the same fashion.

DNA-DNA Reassociation

All reassociations were in 0.12 M PB (pH 6.8) at 50°C, or in 0.41 M PB at 55°C, unless otherwise noted. Reassociation mixtures also contained 2–5 mM EDTA and 0.1–0.2% SDS, and some included calf thymus DNA as carrier. The mixtures were sealed in siliconized glass capillaries, boiled at 99°C (0.12 M PB) or 104°C (0.41 M PB) for 30–60 sec, incubated for the desired time and frozen in dry ice-acetone. For assay, the reassociation mixtures were thawed, diluted into a 100–200 fold excess of 0.12 M PB, 0.05% SDS and passed over hydroxyapatite columns in 0.12 M PB, 0.05% SDS at 50°C. Bound fragments were melted from the column at 96°C, and fractions were counted by liquid scintillation in Handifluor (Mallinckrodt) or assayed for absorbance at 260 nm. Bio-Gel HTP DNA-grade hydroxyapatite, lot #15535 (Bio-Rad), was used throughout this study. All values of DNA or RNA Cot (DNA or total RNA concentration × time, in units of moles of nucleotides per liter × seconds) cited in this paper are *equivalent* Cot—that is, if a reaction is carried out at a salt concentration other than 0.12 M PB (0.18 M Na^+), the Cot is corrected by a factor which takes into account the increase in the renaturation rate constant with increasing Na^+ concentration (Britten et al., 1974).

RNA-DNA Hybridizations

Incubation conditions and assays for RNA-DNA hybridizations were essentially the same as for DNA-DNA renaturation. When necessary, self-reaction of the DNA tracer was monitored, essentially by the method of Galau et al. (1974), as follows. The hybridization mixture, incubated to the desired RNA Cot, was diluted into a 100 fold excess of 0.02 M PB, and this sample was divided into two aliquots. One aliquot was adjusted to 0.12 M PB, 0.05% SDS and fractionated on hydroxyapatite as usual. The fraction of ^{3}H-DNA binding to hydroxyapatite, f_{max}, is the sum of f_s, the fraction of ^{3}H-DNA containing an RNA-DNA hybrid, and f_0, the fraction of ^{3}H-DNA containing a DNA-DNA duplex. To determine f_0, the second aliquot in 0.02 M PB was treated with 50 μg/ml RNAase A at 37°C for 1 hr, a condition sufficient to digest the RNA in RNA-DNA hybrids, leaving only the DNA-DNA duplexes intact. This aliquot was adjusted to 0.12 M PB, 0.05% SDS,

extracted once with IAC and fractionated on hydroxyapatite. The fraction of ^{3}H-DNA binding to hydroxyapatite after the RNAase digestion is f_0, which was usually between 0.01 and 0.1. The fraction of ^{3}H-DNA containing a RNA-DNA hybrid, f_R, was calculated as $f_R = (f_{total} - f_0)/(1 - f_{0_R})$, where f_{0_R} is the fraction of tracer binding to hydroxyapatite at very low Cot.

Unlabeled Whole Sea Urchin DNA

Unlabeled whole sea urchin DNA was extracted from sea urchin sperm and sheared as previously described (Britten et al., 1974).

Total Oocyte RNA

Total RNA was isolated from mature sea urchin oocytes as previously described (Galau et al., 1976).

Cloned ^{32}P-Labeled Repetitive DNA Sequences

The purification, labeling and strand separation of cloned repetitive sea urchin DNA sequences are described in the accompanying paper by Scheller et al. (1978).

RNA Fractionation of DMSO-Sucrose Gradients

Total oocyte RNA (765 μg) was dissolved in 0.6 ml of 90% (v/v) DMSO (dimethyl sulfoxide, Eastman spectro grade) containing 0.1 M LiCl, 0.005 M EDTA, 0.2% SDS and 0.01 M Tris-HCl (pH 6.5). This sample was heated at 55°C for 5 min, then diluted to 0.9 ml with 0.1 M LiCl, 0.005 M EDTA, 0.01 M Tris-HCl (pH 6.5). One third of this sample was layered onto each of three sucrose gradients (4–20%) containing 60% (v/v) DMSO, 0.1 M LiCl, 0.01 M Tris-HCl (pH 6.5), 0.005 M EDTA and 0.05% SDS. After centrifugation for 65 hr in the SW41 Ti rotor at 35,000 rpm, 25°C, gradients were pumped through an ISCO recording spectrophotometer and fractions were collected. Fractions were pooled into RNA size classes as described in Figure 7, dialyzed extensively against 0.02 M Na acetate (pH 6.8) at 4°C and concentrated by repeated extraction with sec-butanol. RNA fractions were further concentrated by ethanol precipitation, resuspended in 0.005 M Na acetate (pH 6.8) and stored at −70°C.

Acknowledgments

This research was supported by a grant from the National Institute of Child Health and Human Development. F.D.C. and R.H.S. were supported by NIH training grants to the Division of Biology and the Division of Chemistry.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 13, 1978; revised June 8, 1978

References

Anderson, D. M., Galau, G. A., Britten, R. J. and Davidson, E. H. (1976). Sequence complexity of the RNA accumulated in oocytes of Arbacia punctulata. Dev. Biol. *51*, 138–145.

Angerer, R. C., Davidson, E. H. and Britten, R. J. (1975). DNA sequence organization in the mollusc Aplysia californica. Cell *6*, 29–40.

Bantle, J. A. and Hahn, W. E. (1976). Complexity and characterization of polyadenylated RNA in the mouse brain. Cell *8*, 139–150.

Britten, R. J., Graham, D. E. and Neufeld, B. R. (1974). Analysis of repeating DNA sequences by reassociation. In Methods in Enzymology, *29E*, L. Grossman and K. Moldave, eds. (New York: Academic Press), p. 363–406.

Britten, R. J., Graham, D. E., Eden, F. C., Painchaud, D. M. and Davidson, E. H. (1976). Evolutionary divergence and length of repetitive sequences in sea urchin DNA. J. Mol. Evol. *9*, 1–23.

Campo, M. S. and Bishop, J. O. (1974). Two classes of messenger RNA in cultured rat cells: repetitive sequences transcripts and unique sequence transcripts. J. Mol. Biol. *90*, 649–663.

Chamberlin, M. E., Britten, R. J. and Davidson, E. H. (1975). Sequence organization in Xenopus DNA studied by the electron microscope. J. Mol. Biol. *96*, 317–333.

Chamberlin, M. E., Galau, G. A., Britten, R. J. and Davidson, E. H. (1978). Studies on nucleic acid reassociation kinetics: renaturation of DNA fragments of unequal length. Nucl. Acids Res. *5*, 2073–2094.

Chetsanga, C. J., Poccia, D. L., Hill, R. J. and Doty, P. (1970). Stage-specific RNA transcription in developing sea urchins and their chromatins. Cold Spring Harbor Symp. Quant. Biol. *35*, 629–634.

Davidson, E. H. (1976). Gene Activity in Early Development, second edition (New York: Academic Press).

Eden, F. C., Graham, D. E., Davidson, E. H. and Britten, R. J. (1977). Exploration of long and short repetitive sequence relationships in the sea urchin genome. Nucl. Acids Res. *4*, 1553–1567.

Farquhar, M. N. and McCarthy, B. J. (1973). Histone mRNA in eggs and embryos of Strongylocentrotus purpuratus. Biochem. Biophys. Res. Commun. *53*, 515–522.

Galau, G. A., Britten, R. J. and Davidson, E. H. (1974). A measurement of the sequence complexity of polysomal messenger RNA in sea urchin embryos. Cell *2*, 9–21.

Galau, G. A., Klein, W. H., Davis, M. M., Wold, B. J., Britten, R. J. and Davidson, E. H. (1976). Structural gene sets active in embryos and adult tissues of the sea urchin. Cell *7*, 487–505.

Galau, G. A., Lipson, E. D., Britten, R. J. and Davidson, E. H. (1977). Synthesis and turnover of polysomal mRNAs in sea urchin embryos. Cell *10*, 415–432.

Glisin, V. R., Glisin, M. V. and Doty, P. (1966). The nature of messenger RNA in the early stages of sea urchin development. Proc. Nat. Acad. Sci. USA *56*, 285–289.

Goldberg, R. B., Galau, G. A., Britten, R. J. and Davidson, E. H. (1973). Nonrepetitive DNA sequence representation in sea urchin embryo messenger RNA. Proc. Nat. Acad. Sci. USA *70*, 3516–3520.

Goldberg, R. B., Crain, W. R., Ruderman, J. V., Moore, G. P., Barnett, T. R., Higgins, R. C., Gelfand, R. A., Galau, G. A., Britten, R. J. and Davidson, E. H. (1975). DNA sequence organization in the genomes of five marine invertebrates. Chromosoma *51*, 225–251.

Graham, D. E., Neufeld, B. R., Davidson, E. H. and Britten, R. J. (1974). Interspersion of repetitive and nonrepetitive DNA sequences in the sea urchin genome. Cell *1*, 127–137.

Gross, K., Ruderman, J., Jacobs-Lorena, M., Baglioni, C. and Gross, P. R. (1973). Cell-free synthesis of histones directed by messenger RNA from sea urchin embryos. Nature New Biol. *241*, 272–274.

Hinegardner, R. T. (1967). Echinoderms. In Methods in Developmental Biology, F. H. Wilt and N. K. Wessells, eds. (New York: Thomas Y. Crowell), pp. 139–155.

Hough, B. R. and Davidson, E. H. (1972). Studies on the repetitive sequence transcripts of Xenopus oocytes. J. Mol. Biol. *70*, 491–509.

Hough-Evans, B. R., Wold, B. J., Ernst, S. G. Britten, R. J. and Davidson, E. H. (1977). Appearance and persistence of maternal RNA sequences in sea urchin development. Dev. Biol. *60*, 258–277.

Humphreys, T. (1971). Measurements of messenger RNA entering polysomes upon fertilization of sea urchin eggs. Dev. Biol. *26*, 201–208.

Kleene, K. C. and Humphreys, T. (1977). Similarity of hnRNA sequences in blastula and pluteus stage sea urchin embryos. Cell *12*, 143–155.

Klein, W. H., Murphy, W., Attardi, G., Britten R. J. and Davidson,

E. H. (1974). Distribution of repetitive and nonrepetitive sequence transcripts in HeLa mRNA. Proc. Nat. Acad. Sci. USA *71*, 1785–1789.

Klein, W. H., Thomas, T. L., Lai, C., Scheller, R. H., Britten, R. J. and Davidson, E. H. (1978). Characteristics of individual repetitive sequence families in the sea urchin genome studied with cloned repeats. Cell *14*, 889–900.

Legler, M. K. and Cohen, E. P. (1978). Estimation of the number of nucleotide sequences in mouse DNA complementary to messenger RNAs specifying a complete mouse immunoglobulin. Biochemistry *15*, 4390–4399.

Moore, G. P., Scheller, R. H., Davidson, E. H. and Britten, R. J. (1978). Evolutionary change in repetition frequency of sea urchin DNA sequences. Cell, in press.

Pearson, W. R., Davidson, E. H. and Britten, R. J. (1977). A program for least squares analysis of reassociation and hybridization data. Nucl. Acids Res. *4*, 1727–1737.

Scheller, R. H., Thomas, T. L., Lee, A. S., Klein, W. H., Niles, W. D., Britten, R. J. and Davidson, E. H. (1977). Clones of individual repetitive sequences from sea urchin DNA constructed with synthetic EcoR1 sites. Science *196*, 197–200.

Scheller, R. H., Costantini, F. D., Kozlowski, M. R., Britten, R. J. and Davidson, E. H. (1978). Specific representation of cloned repetitive DNA sequences in sea urchin RNAs. Cell *15*, 189–203.

Schmid, C. W. and Deininger, P. L. (1975). Sequence organization of the human genome. Cell *6*, 345–358.

Skoultchi, A. and Gross, P. R. (1973). Maternal histone messenger RNA: detection by molecular hybridization. Proc. Nat. Acad. Sci. USA *70*, 2840–2844.

Smith, M. J., Chamberlin, M. E., Hough, B. R. and Davidson, E. H. (1974). Repetitive and non-repetitive sequence in sea urchin hnRNA. J. Mol. Biol. *85*, 103–126.

Stafford, D. W. and Bieber, D. (1975). Concentration of DNA solutions by extraction with 2-butanol. Biochim. Biophys. Acta *378*, 18–21.

Strauss, J. H., Kelly, R. B. and Sinsheimer, R. L. (1968). Denaturation of RNA with dimethyl sulfoxide. Biopolymers *6*, 793–807.

Vogt, V. M. (1973). Purification and further properties of single-strand-specific nuclease from Aspergillus oryzae. Eur. J. Biochem. *33*, 192–200.

CHAPTER 6

Cell, Vol. 15, 189-203, September 1978, Copyright © 1978 by MIT

Specific Representation of Cloned Repetitive DNA Sequences in Sea Urchin RNAs

Richard H. Scheller
Division of Chemistry
California Institute of Technology
Pasadena, California 91125
Franklin D. Costantini, Michael R. Kozlowski,*
Roy J. Britten† and Eric H. Davidson
Division of Biology
California Institute of Technology
Pasadena, California 91125

Summary

Nine cloned repetitive sequences were labeled, strand-separated and individually hybridized with RNA extracted from the nuclei of gastrula stage sea urchin embryos and of adult sea urchin intestine cells. The concentration of transcripts complementary to each cloned sequence was measured by RNA excess hybridization kinetics and by a DNA excess titration method. Transcripts of certain of the repeat families are present at over 100 times the concentration of transcripts of other families in each RNA. The set of repetitive sequence families highly represented in intestine nuclear RNA is different from that highly represented in gastrula nuclear RNA. Together with the results obtained with mature oocyte RNA and presented in the accompanying paper by Costantini et al. (1978), these findings show that quantitative patterns of repetitive sequence representation in RNA are specific to each cell type. Both strands of all of the nine cloned repeats are represented at some level in all the RNAs studied. Usually, though not always, the concentrations of transcripts complementary to the two strands of each repeat do not differ by more than a factor of two. The cloned tracers do not react with polysomal messenger RNA, and the nuclear RNA molecules with which they hybridize are many times larger than the repetitive sequences themselves.

Introduction

The presence of a diverse set of repetitive sequence families is an ubiquitous feature of animal genomes. In the DNA of most groups such as mammals, amphibians, echinoderms and molluscs, a majority of the individual repeat sequences are only a few hundred nucleotides long and are interspersed among single-copy sequences (Davidson et al., 1974; Schmid and Deininger, 1975; Goldberg et al., 1975; Wu et al., 1977). At least in sea urchin DNA, single-copy structural genes are located nonrandomly close to the interspersed repeats (Davidson et al., 1975). The arrangement of repetitive sequence elements in the genome, and the finding that evolutionary change in these sequences is relatively restricted, provide indirect arguments that they must perform some generally important genomic function (for example, see Britten and Davidson, 1971; Davidson and Britten, 1973; Davidson et al., 1975; Harpold and Craig, 1977; Davidson, Klein and Britten, 1977; Klein et al., 1978; Moore et al., 1978). In their sequence length, repetition frequencies and other characteristics, most of the repeats do not appear to resemble structural genes. No general function has yet been identified for repetitive sequence families, and the basic reason for their existence remains unknown.

A key question is whether repetitive sequences are transcribed, and if so, whether the patterns of transcription have anything to do with the state of cell differentiation. Early hybridization studies identified repetitive sequence transcripts and suggested that the patterns of repeat transcription are indeed tissue-specific (for reviews, see Davidson and Britten, 1973; Davidson, 1976). Most of the data generated in these studies, are only partially interpretable due to the inadequacy of the methods used for kinetic measurements. This led to difficulty in distinguishing between quantitative changes in repeat transcript sequence concentration, qualitative changes in the sets of repeats represented in various RNAs and differences in the genomic reiteration frequencies of the repeats that are transcribed. These questions can be answered in a precise way by studying the hybridization with cellular RNAs of individual cloned repetitive sequences, each representing a specific repeat family. This paper describes experiments which demonstrate that the transcripts of cloned repetitive sequence families are present at very different concentrations in the nuclear RNA of sea urchin embryos, as compared with the nuclear RNA of adult intestine cells. The set of cloned repeats which is highly represented in gastrula nuclear RNA is different from the set which is highly represented in intestine nuclear RNA. The accompanying report of Costantini et al. (1978) shows that these cloned repetitive sequences display a still distinct pattern of representation in mature oocyte RNA. Taken together, these data lead to the general conclusion that repetitive sequence representation in RNA is highly tissue-specific. The repeat transcripts appear not to be associated with polysomal mRNA, and they differ from the latter in that

* Present address: Department of Neurobiology, University of California, Irvine, California 92717.
† Also a staff member of the Carnegie Institution of Washington, Baltimore, Maryland 21210.

both complementary strands of each repeat sequence are always represented in the RNA.

Results

Strand-Separated, Cloned Repetitive Sequence Tracers

The object of the experiments described below was to measure the concentration of transcripts complementary to individual repetitive sequence families in the nuclear RNAs of two different tissues. Each family was represented by a cloned DNA fragment (Scheller et al., 1977). For the purpose of the hybridization experiments and in order to determine whether repetitive sequence representation in the RNA is symmetrical, it was necessary to separate the two complementary strands of each cloned DNA fragment. This was accomplished by electrophoresis in neutral polyacrylamide gels loaded with alkali-denatured DNA fragments (see Experimental Procedures). Figure 1 shows that each of the separated strands is essentially free of its complement, since no reaction whatsoever is observed when it is incubated alone.

Table 1 lists some characteristics of the nine cloned repetitive sequences studied in this paper and the repeat families to which they belong. The data shown are reproduced from the measurements of Klein et al. (1978). Table 1 indicates the genomic reiteration frequencies, the cloned sea urchin DNA fragment length, the approximate base composition and an estimate of the internal sequence divergence for each family. The latter is the average fraction of mispaired bases in the strand pairs formed when complementary members of the repeat family react with each other. It is calculated by comparing the thermal stability of the native cloned repeat with the thermal stability of heteroduplexes formed between a cloned DNA fragment and the complementary sequences in the genomic DNA. The cloned repeats chosen for the present study are heterogeneous with respect to all of the parameters listed in Table 1. Some, such as those carried in clones 2007 and 2034, belong to families displaying very little intrafamilial sequence divergence. The repeat families represented by clones 2090, 2109A and 2109B, on the other hand, are relatively divergent. Although seven of the nine repeat fragments fall in the 100–400 nucleotide range, those carried in clones 2034 and 2007 are longer. Reiteration frequency for this set of repeat families varies from approximately 20 copies per haploid genome to approximately 1000. With regard to the properties considered in Table 1, the nine cloned repeats chosen for this study are fairly representative of the variety of repetitive sequences found in the S. purpuratus genome (Klein et al., 1978).

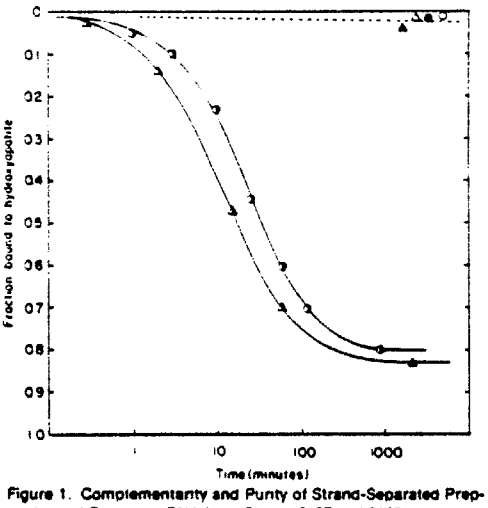


Figure 1. Complementarity and Purity of Strand-Separated Preparations of Repetitive DNA from Clones 2137 and 2108

Equal aliquots of radioactive lower and upper strand DNA from clones 2137 and 2108 were mixed, heated at 105°C and incubated at 60°C in 0.5 M phosphate buffer. The fraction of ^{32}P-DNA reassociated at each time point was assayed by hydroxyapatite binding. The solid lines are least-squares solutions assuming second-order kinetics. (—○—) clone 2137, (—△—) clone 2108. The half-time for the reaction of the clone 2137 DNA was 24.2 min, and that for the reaction of the clone 2108 DNA was 11.2 min. From the amount of radioactivity (2.54×10^6 cpm/ml) and the fragment lengths (Table 1), the specific activities were calculated to be approximately 1.7×10^7 cpm/μg for the 2137 fragment and 8.2×10^6 cpm/μg for the 2108 fragment. As a kinetic standard for this calculation, we used the second-order renaturation of ϕX174RF DNA studied by Galau, Britten and Davidson (1977). When either 2137 upper strand (○), 2108 upper strand (△), 2137 lower strand (●) or 2108 lower strand (▲) were incubated in the absence of their respective complementary strands, no detectable reaction occurred (—).

Hybridization Reactions with Sea Urchin Nuclear RNAs

The nuclear RNAs chosen for this study derive from dissimilar sources—specifically, gastrula stage embryos and adult intestine cells. Hough et al. (1975) showed that the complexity of gastrula hnRNA is $\sim 2 \times 10^8$ nucleotides, or approximately a third of the total single-copy sequence length in the genome, and that the average single-copy RNA sequence is present in 0.5–1 copy per nucleus. This conclusion was confirmed recently in our laboratory by independent measurements of the kinetics of reaction of a single-copy ^{3}H-DNA tracer with whole embryo RNA. Wold et al. (1978) found that the complexity of intestine RNA is at least equal to, if not greater than, that of the gastrula hnRNA, but

Table 1. Some Characteristics of Nine Cloned Repetitive Sequences and Their Genomic Families

Clone[a]	Genomic Reiteration Frequency[b]	Length[c] (NTP)	Approximate Base Composition[d]	Estimated Intrafamilial Divergence[e] (%)
2007	400	1100	7	3.6
2034	1000	560	25	3.7
2090	140	220	38	>20.5
2101	700	320	20	7.5
2108	20	190	48	5.7
2109A	900	200	32	>19.7
2109B	1000	125	44	>25.3
2133B	60	310		
2137	530	190		

[a] Clones of this series constructed in this laboratory by blunt end ligation of renatured repeats are designated CS2000–CS3999. For convenience, the CS prefix is omitted in this paper.
[b] Measured by the renaturation kinetics of reactions between the labeled cloned tracer and excess genomic DNA (Klein et al., 1978).
[c] Measured by gel electrophoresis in the presence of standards of known length.
[d] Percentage of (G + C) calculated from the thermal stability of the native cloned tracer, taking into account the duplex length. Since primary sequence may affect this calculation, the values shown are only approximate. Such errors are not likely to be very large, however, and in the case where the primary sequence is known (clone 2109B; unpublished data), agreement is quite satisfactory. See Klein et al. (1978).
[e] Values for the mean intrafamilial sequence divergence are shown, as measured by Klein et al. (1978). This parameter is the mean number of mispaired bases per 100 nucleotides in strand pairs between complementary members of the same repeat family. It is calculated assuming that 1% sequence mismatch results in a 1°C decrease in heteroduplex thermal stability relative to a perfectly matched duplex (Britten, Graham and Neufeld, 1974). The data used for this calculation are the Tm of the native cloned fragment and the Tm of the heteroduplexes formed by reacting the cloned fragment with genomic DNA at 45°C in 0.12 M phosphate buffer.

the steady state quantity of complex RNA per nucleus is less. The data of Wold et al. (1978) show that there is only about 0.1 copy of each sequence per average intestine nucleus. This nuclear RNA was chosen for the present experiments because the pattern of structural gene expression in adult intestine is very different from that in the gastrula. Galau et al. (1976) and Wold et al. (1978) showed that the complexity of intestine polysomal mRNA is only $\sim 6 \times 10^8$ nucleotides as compared with 17×10^8 nucleotides for gastrula. About half of the intestine mRNA sequence set is also represented in gastrula or blastula polysomal mRNA.

The kinetics of reactions between excess gastrula and intestine nuclear RNAs and the strand-separated, cloned tracers were measured. From these kinetics, it is possible to calculate the approximate concentration in the RNA of transcripts of those specific repetitive sequence families to which the cloned tracers belong. This calculation is carried out by comparing the rate of the hybridization reaction to the rate of the reaction of the same RNA with single-copy DNA. The number of RNA transcripts per nucleus complementary to a given strand of a cloned repeat, T_c, is given by:

$$T_c = \frac{k_c}{k_{sc}} T_{sc} f_L \qquad (1)$$

In this equation, T_{sc} is the number of transcripts of an average single-copy sequence per nucleus, cited above for each nuclear RNA; f_L is a correction factor for the effect on the kinetics of disparity in tracer and driver length (Chamberlin et al., 1978); k_{sc} is the pseudo-first-order single-copy hybridization rate constant; and k_c is the rate constant for the reaction of the cloned tracer with the nuclear RNA. Numerical values of these parameters used in the calculation of T_c are listed in the legend to Table 2. The intrafamilial sequence divergence and the tracer base composition vary for each (Table 1), and this could affect the hybridization rate constants used for calculations with equation (1). The hybridization conditions which we used (0.5 M phosphate buffer, 55°C) provide a reaction criterion significantly below the Tm of the RNA–^{32}P-DNA hybrids formed in these reactions (data not shown). We know from the small intrafamilial divergence in most cases (Klein et al., 1978) that the kinetic effects of divergence within the cloned repeat families cannot be large (Bonner et al., 1973). For the highly divergent cloned repeat families termed class III by Klein et al. (1978) (that is, clones 2090, 2109A and 2109B), however, the hybridization kinetics could yield underestimates of transcript concentration by several fold. We might expect the kinetic estimates of transcript number to be low for these cases, both because of kinetic retardation and because there may be some family members and transcripts whose sequences are too divergent to react at all under our conditions (Klein et al., 1978). An additional effect on the reaction kinetics derives from the disparity in length between the hybridizing RNA molecules and some of the cloned tracer DNA fragments. As shown by Costantini et al. (1978) and below, most of the reactive RNAs are much longer than the tracers. A length correction of several fold on the kinetics of the RNA hybridization reactions is indicated by the study of Chamberlin et al. (1978), who showed that long driver-short tracer reactions are retarded. Where the average length of the reacting RNA species is known by direct measurement, we have corrected the measured rate constant according to equation (5) of Chamberlin and colleagues. For other cases,

Table 2. Prevalence of Cloned Repetitive Sequence Transcripts in Three RNAs

		Gastrula Nuclear RNA					Intestine Nuclear RNA					Total Oocyte RNA[f]				
Clone	Strand	Rate Constant[a]	Transcripts per Nucleus by Kinetics[b]	RNA Fraction[c]	Transcripts per Nucleus by Titration[d]	Representation[e] (%)	Rate Constant[a]	Transcripts per Nucleus by Kinetics[b]	RNA Fraction[c]	Transcripts per Nucleus by Titration[d]	Representation[e] (%)	Rate Constant[a]	Transcripts per Cell by Kinetics[b]	RNA Fraction[c]	Transcripts per Cell by Titration[d]	Representation[e] (%)
2007	U	1.4×10^{-3}	13			3.2	3.9×10^{-3}	31			39				7,000	1.1
	L	1.2×10^{-3}	11			2.7	4.2×10^{-3}	34			43			1.5×10^{-6}	7,000	1.1
2034	U	1.1×10^{-3}	20	1.5×10^{-6}	9	0.9	6.1×10^{-3}	48			48			3.6×10^{-7}	3,300	0.20
	L	5.0×10^{-4}	10			1.0	3.9×10^{-3}	31			31			5.8×10^{-7}	5,300	0.33
2090	U	3.3×10^{-3}	120	3.0×10^{-6}	47	34	9.2×10^{-3}	74			53	1.7×10^{-3}	35,000			18
	L	2.1×10^{-3}	76	4.0×10^{-6}	63	45	1.5×10^{-3}	12			86	1.7×10^{-3}	35,000			18
2101	U	5.4×10^{-4}	10	1.1×10^{-6}	12	1.7	6.4×10^{-3}	51			73			5.6×10^{-7}	8,400	0.86
	L	4.8×10^{-4}	8			1.1	3.8×10^{-3}	30			43			1.9×10^{-7}	3,000	0.29
2108	U	1.2×10^{-2}	22	1.1×10^{-6}	19	95			1.2×10^{-6}	1.1	55	3.8×10^{-3}	79,000	5.8×10^{-6}	160,000	590
	L			5.3×10^{-7}	9.6	48			5.9×10^{-7}	0.53	27	1.3×10^{-3}	27,000	3.2×10^{-6}	90,000	270
2109A	U	5.8×10^{-4}	10			1.0	2.0×10^{-4}	16	2.3×10^{-4}	20	22	1.8×10^{-4}	2,500			0.30
	L	5.8×10^{-4}	10	5.3×10^{-7}	9.1	1.0	2.5×10^{-4}	2			3	2.0×10^{-4}	2,800			0.33
2109B	U	3.5×10^{-2}	640			64	2.4×10^{-4}	1.8			1.8	6×10^{-4}	83,000			5
	L	3.5×10^{-2}	640			64	1.1×10^{-4}	0.88			0.9	6×10^{-4}	83,000			5
2133B	U			1.4×10^{-7}	1.6	2.7			8.7×10^{-7}	0.48	8			4.1×10^{-7}	6,800	7.0
	L			5.5×10^{-8}	0.6	1.0			5.3×10^{-7}	0.29	4.8			3.4×10^{-7}	5,600	5.9
2137	U	2.7×10^{-3}	50	6.0×10^{-7}	11	4.2	7.4×10^{-3}	59			11	9×10^{-4}	19,000			2.8
	L	2.9×10^{-3}	52	1.2×10^{-6}	22	8.4	5.3×10^{-3}	4.2			7.9	7.8×10^{-4}	16,000			2.4

[a] The second-order rate constants – that is, k_c in equation (1) (in units of M^{-1} sec $^{-1}$) – are obtained from nonlinear least-squares solutions to the RNA excess kinetics (Pearson et al., 1977) (see text).

[b] The number of transcripts per nucleus or per cell is calculated by application of equation (1) in the text. The constants used in these calculations are as follows: gastrula $k_{ss} = 1.1 \times 10^{-4}$, $T_{ss} = 1$; intestine $k_{ss} = 2.5 \times 10^{-4}$, $T_{ss} = 0.1$; oocyte $k_{ss} = 2.3 \times 10^{-4}$, $T_{ss} = 1600$. The value of f_t calculated from the data of Figure 6 for the clone 2090 fragment reacting with gastrula nuclear RNA was 4; for the clone 2137 fragment reacting with gastrula nuclear RNA the value of f_t was 2; for all other reactions with gastrula and intestine nuclear RNA except those with the clone 2007 fragment, a minimum esumata of 2 was applied – that is, assuming that the hybridizing RNA molecules are at least 4 times larger than the cloned fragment. Chamberlin et al. (1978) found that the rate of long driver-short tracer reactions is retarded by a factor approximately equal to $(L_T/L_D)^{1/2}$, where L_T and L_D are tracer and driver fragment lengths. For the long clone 2007 fragment, f_t was taken as 1. Values for f_t for oocyte reactions are those found in Table 1 of the accompanying paper by Costantini et al. (1978).

[c] The RNA fraction is derived from the titration curves by application of equation (2) or, in cases of large asymmetry of representation, by equation (3). The values shown are the inverse of the least-squares solution for (a) in these expressions.

[d] The number of transcripts per nucleus or per cell by titration is calculated using equation (4) in text. Values of L, the length of each cloned fragment, are from Table 1. Values of Q, the amount of RNA per nucleus or per cell in the preparations used for these calculations are as follows: for gastrula nuclear RNA, $Q = 2 \times 10^{-12}$ g; for intestine nuclear RNA, $Q = 1 \times 10^{-12}$ g; for oocyte RNA, $Q = 3 \times 10^{-9}$ g (Whiteley, 1949; Hough et al., 1975; Wold et al. 1978; also our unpublished data).

[e] Percentage of representation is calculated according to equation (5), using titration data where possible.

[f] Kinetic and titration data for oocyte RNA are from Costantini et al. (1978).

a minimum reasonable correction has been applied, as indicated in the legend to Table 2. Despite these various uncertainties, the transcript concentrations calculated from the rate constants using equation (1) agree within a factor of 2–3 with measurements of transcript concentrations made by the totally independent titration method, as discussed below.

Kinetic data were obtained by measuring hydroxyapatite binding of the cloned tracer fragments as a function of RNA Cot. Examples are shown in Figure 2. In each panel of Figure 2, the two right-hand curves are kinetic standards presented for purposes of comparison. The right-hand solid line indicates the kinetics of the reaction of a single-copy ^{3}H–DNA tracer with excess intestine nuclear RNA under the same conditions, and the right-hand dashed line shows the reaction of the single-copy ^{3}H–DNA tracer with excess gastrula nuclear RNA. Each of the cloned tracers reacts at a particular rate with each nuclear RNA. Figure 2A shows that the upper strand of the repeat fragment from clone 2034 reacts much faster with intestine nuclear RNA (circles) than with gastrula nuclear RNA (triangles). The opposite case is seen in Figure 2B. Here the clone 2109B upper strand is found to react more rapidly with gastrula nuclear RNA than with intestine nuclear RNA. A third situation is illustrated in Figure 2C. In this case, the upper strand fragment of clone 2090 is shown to react at about the same rate with either of the nuclear RNAs. There is no possibility that the RNA Cot-dependent hybridization observed could be due to DNA contaminating the nuclear RNA preparations. Prior treatment of hybridization samples with ribonuclease under low salt conditions (in which RNA-DNA hybrids are destroyed though DNA duplexes are not affected) completely eliminates all hydroxyapatite binding of the strand-separated tracer fragments.

A result of major interest in these experiments is that *both* strands of each cloned repetitive sequence react with the nuclear RNAs. Usually, but not always, the transcript concentration is within a factor of two for the two complementary strands. Since both strands are more or less equally repre-

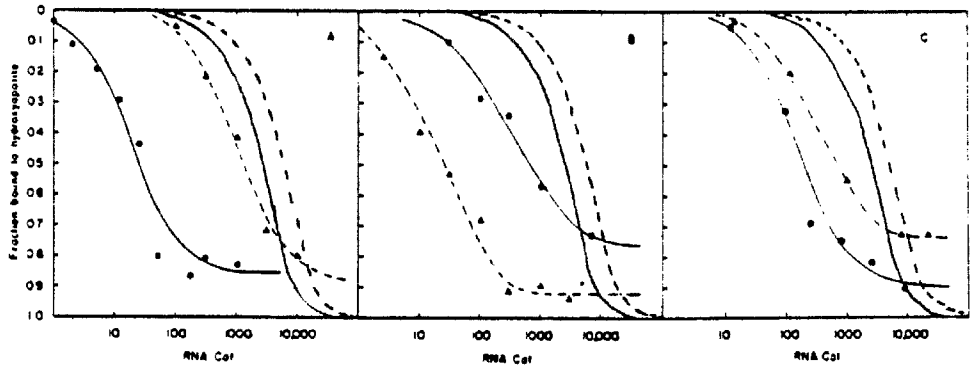


Figure 2. Kinetics of Reaction between Excess Nuclear RNA and Strand-Separated Cloned Repeat Tracers

Each panel displays the kinetics of reaction between a strand-separated cloned tracer and intestine (●) and gastrula (▲) nuclear RNAs. The reactions were carried out in 0.5 M phosphate buffer at 55°C and assayed by hydroxyapatite binding. The possible presence of DNA-DNA duplex was routinely monitored by measuring the binding to hydroxyapatite of samples exposed to a low salt RNAase treatment. The methods used are described in Experimental Procedures. Data plotted are normalized to the reactivity of the individual tracers, indicated below. The lines drawn through these data show best least-squares solutions to the data, assuming second-order kinetics (see text). Rate constants are listed in Table 2. The kinetics of the reactions of single-copy ^{3}H–DNA tracer with gastrula nuclear RNA (data from Hough et al., 1975; also our unpublished data) and with intestine nuclear RNA (data from Wold et al., 1978) are indicated in each panel. Pseudo-first-order rate constants for these reactions are: for intestine nuclear RNA, 2.5×10^{-6} M^{-1} sec^{-1} (solid line), and for gastrula nuclear RNA, 1.1×10^{-4} M^{-1} sec^{-1} (dashed line).

(A) Reactions with the upper strand of the sea urchin DNA insert of clone 2034. Reactivity of the clone 2034 U fragment was 77%. The "nonreactive" ^{32}P radioactivity in the annealing mixtures is mainly γ-^{32}P-ATP persisting from the kinase reaction used for labeling the fragments. The intestine nuclear RNA was present at a 7×10^6 fold mass excess, and the gastrula nuclear RNA at a 4×10^6 fold mass excess with respect to the tracer.

(B) Reactions with the upper strand of the sea urchin DNA insert of clone 2109B. Reactivity of this tracer was 63%. The intestine nuclear RNA was present in a 1×10^7 fold mass excess, and the gastrula nuclear RNA in a 5×10^6 fold mass excess with respect to the ^{32}P–DNA tracer. The lower termination with the intestine RNA is due to the fact that the sequence excess for this RNA was only about 7 fold, due to its relatively low prevalence in intestine nuclear RNA and to competition for the upper strand fragment by a slight excess of RNAs of the same complementarity as the upper strand.

(C) Reaction with the upper strand of the sea urchin DNA insert of clone 2090. The reactivity of this tracer was 61%. The intestine nuclear RNA was present in a 3×10^6 fold mass excess, and the gastrula nuclear RNA in a 2.5×10^6 fold mass excess with respect to the tracer DNA. The slight decrease in the terminal value in the gastrula reaction is of the same cause as indicated in (B) for the intestine RNA reaction.

sented in the nuclear RNA, the kinetics of most of the reactions with the cloned tracers are generally closer in form to second-order than to pseudo-first-order. The solid (intestine) and dashed (gastrula) lines shown in Figure 2 for the cloned tracer reactions are the least-squares solutions assuming second-order kinetics. Second-order rate constants and values of T_c calculated by application of equation (1) are shown in Table 2.

Measurements of the concentration in nuclear RNAs of transcripts complementary to each of the two strands are shown for several clones in Figure 3. The method here is titration of *excess* strand-

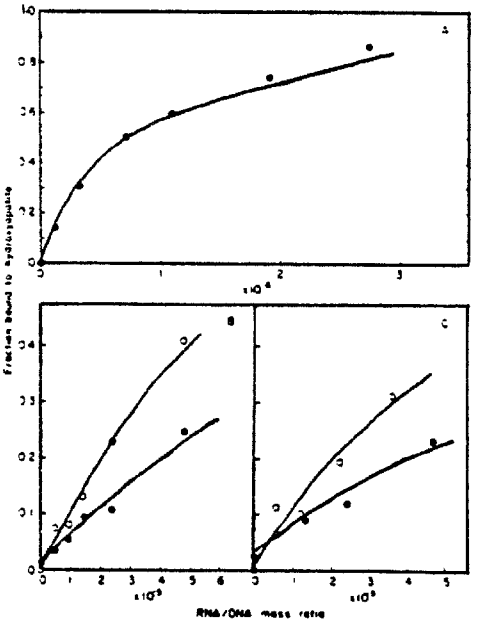


Figure 3. Titration of Strand-Separated Cloned Repeat Fragments with Gastrula and Intestine Nuclear RNAs

Strand-separated cloned tracers were reacted in excess with varying amounts of intestine or gastrula nuclear RNA to Cot values (with respect to the excess ^{32}P-DNA) sufficient to terminate the reactions. The abscissa shows RNA/DNA mass ratio × 10^{-4} for clone 2034; × 10^{-5} for clone 2108. The fraction (1/a) of RNA consisting of transcripts complementary to each cloned tracer fragment was calculated as described in the text. Solid lines are the functions described by equation (2). The number of complementary transcripts per nucleus (T_c) was calculated from these data using equation (4). Values of 1/a and T_c are listed in Table 2.

(A) The upper strand of the clone 2034 fragment was reacted with varying amounts of gastrula nuclear RNA (●). The ^{32}P-DNA concentration was measured by tracer renaturation kinetics, as in Figure 1. The DNA specific activity was 6.5 × 10^6 cpm/μg and its reactivity was 80%.

(B) The upper strand (○) and lower strand (●) of the clone 2108 fragment were reacted with varying amounts of intestine nuclear RNA. The tracer specific activity was 3 × 10^7 cpm/μg and its reactivity was 78%.

(C) The same tracers as in (B) were titrated with gastrula nuclear RNA.

separated ^{32}P-DNA tracer with increasing amounts of RNA. As more RNA is added, a greater amount of the ^{32}P-DNA is hybridized and can be bound to hydroxyapatite. The reactions were run to kinetic termination (that is, ≥10 × $Cot_{\frac{1}{2}}$) with respect to the excess partner—in this case, the ^{32}P-DNA. Calculations of transcript prevalence carried out by this method are not subject to the particular uncertainties which affect the kinetic estimates. Thus titration estimates are not as sensitive to kinetic retardation since the reactions are carried to completion. If, however, a significant number of RNA molecules are present which are broken within the repeat transcript sequence, the titration estimates will yield values which are high, since all the resulting RNA fragments can cause binding of DNA tracer molecules to hydroxyapatite. This should not cause a severe problem, given the high average ratio between the length of the hybridizing RNA molecules and the cloned repeats (see Costantini et al., 1978; also the data presented below). Note that the accuracy of titration measurements depends upon the accuracy of the tracer-specific activity determinations. A practical advantage of the titration over the kinetic method is that it requires smaller quantities of nuclear RNA. This is important in the use of rare transcripts, for which it is difficult to obtain RNA sequence excess.

When both strands of the sequence are present equally in the unlabeled nucleic acid (here the RNA), the form of the titration curve can be described as follows (Wallace, Dube and Bonner, 1977; Moore et al., 1978):

$$\frac{T}{T_0} = \frac{1}{1 + a/R} \tag{2}$$

In this equation, T/T_0 is the fraction of the strand-separated ^{32}P-DNA fragment that can be bound to hydroxyapatite at kinetic termination in a reaction carried out at an RNA/^{32}P-DNA ratio of R, and 1/a is the fraction of the RNA complementary to the tracer fragment. For low RNA/DNA ratios—that is, in the initial part of the titration curve when only a small amount of the tracer can react,

$$\frac{T}{t_0} = \left(\frac{1}{a}\right) R. \tag{3}$$

Thus 1/a is the initial slope of the titration curve. If only the one strand which is complementary to the cloned tracer were present in the RNA, equation (3) would provide a description for the titration curve which is exact beyond its early phase. In fact, the concentration of RNAs complementary to the two strands of the cloned fragments is usually not just the same (see below), and for this reason equation (2) does not always provide an exact form for our present purposes. In the large majority of cases, our data were obtained at sufficiently low RNA/

DNA ratios so that the difference between the value of T/T_u calculated by equation (2) or equation (3) was insignificant. The value of (a) was obtained by least-squares methods (Pearson, Davidson and Britten, 1977). Values of a^{-1} are shown in Table 2. Given an estimate of (a), the number of complementary transcripts for each cloned tracer strand per nucleus, T_c, is calculated in the following manner:

$$T_c = \frac{1}{a} Q \left(\frac{1}{350L}\right) N \qquad (4)$$

In this equation, Q is the mass of RNA per nucleus in the RNA preparation (in grams; see legend to Table 2 for numerical values), a is as defined above in equations (2) and (3), N is Avogadro's number and L is the length of the cloned fragments (from Table 1) in nucleotides. The values of T_c calculated by application of equation (4) to titration data are listed in Table 2, where they can be compared to the kinetic estimates of T_c.

Figure 3A illustrates the titration with gastrula nuclear RNA of the upper strand of the clone 2034 fragment. The form of the complete titration curve, fit with equation (2), can be seen in this panel. The lower strand of the clone 2034 fragment is represented about equally in gastrula nuclear RNA (data not shown). Approximately nine transcripts complementary to the 2034U fragment are present per gastrula nucleus, while according to the kinetic experiment shown in Figure 2A, there are approximately 20 transcripts per nucleus. Figures 3B and 3C show intestine nuclear RNA and gastrula nuclear RNA titrations carried out with the upper and lower strands of the clone 2108 fragment. There are 10–20 times more transcripts complementary to both the upper and lower strands of the 2108 fragment in gastrula nuclear RNA than in intestine nuclear RNA. The titration curves, however, show that concentrations of transcripts complementary to the upper and lower strands are only a factor of about two apart in each nuclear RNA.

Figure 4 shows the most extreme example of asymmetric strand representation which we encountered. Figure 4A shows the reaction of the upper and lower strands of the clone 2109A fragment with each other, and demonstrates the purity of each strand by the method used in Figure 1. Figure 4B shows the kinetics of the reaction of these two strands with gastrula nuclear RNA. Both strands are represented by approximately 10 complementary transcripts per gastrula nucleus. In

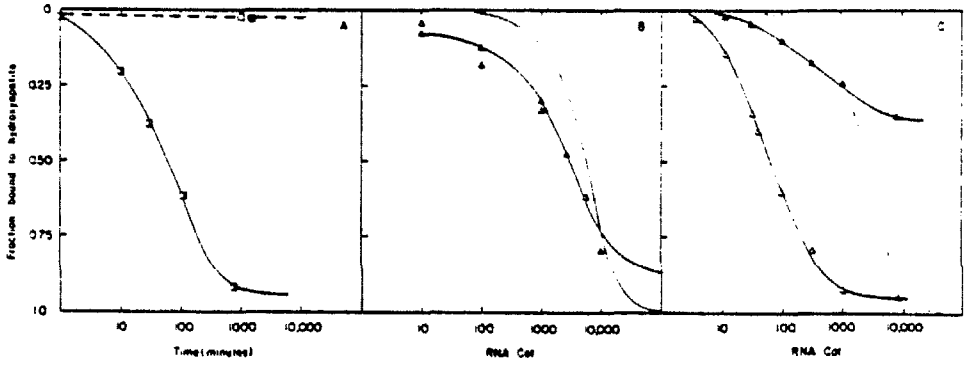


Figure 4. Kinetics of Reactions of the Upper and Lower Strands of the Clone 2109A Fragment with Nuclear RNAs

(A) Renaturation of the upper and lower strands of the clone 2109A fragment, as in Figure 1 (□). Failure of separated strands to react with themselves is also shown: 2109A U strand (□) and 2109A L strand (■). Data have been normalized for a tracer reactivity of 62%, measured by reaction with excess total sea urchin DNA. The renaturation follows second-order kinetics (solid line). Specific activity of these tracers was calculated to be 9×10^6 cpm/μg. This experiment was carried out to ensure that the 2109A U and L tracer preparations used in (B) and (C) behaved appropriately.

(B) Reaction of the 2109A U (△) and L (▲) tracers with gastrula nuclear RNA. Both sets of data are consistent with the same second-order kinetic function (solid line). The rate constant obtained appears in Table 2. The mass ratio of gastrula nuclear RNA to the cloned tracers was 9×10^6 in this experiment. The solid line indicates the kinetics of the reaction of gastrula nuclear RNA with a single-copy ^{3}H–DNA tracer (data from Hough et al., 1975).

(C) Reaction of the same 2109A U (△) and L (▲) tracers with intestine nuclear RNA. Transcripts complementary to the two strands differ in concentration by at least 3 fold. The failure of termination of the L strand reaction is due mainly to competition from the excess upper strand complement in the RNA. The mass ratio of RNA to tracer was 5×10^6. Estimated second-order rate constants for these reactions are listed in Table 2. Due to the unusual strand asymmetry, however, the kinetics of these reactions should deviate from second-order form, and the solutions are only approximate—that is, the upper strand reaction should tend toward pseudo-first-order kinetics. Other factors, such as presence of small nucleic acid fragments, may be attenuating the latter portion of the reaction, thus yielding the overall second-order form observed (solid line). The maximum difference in the number of complementary transcripts per nucleus due to uncertainties in the form of the kinetics will not exceed a factor of two. The dashed line shows the kinetics of the reaction of intestine nuclear RNA with single-copy ^{3}H–DNA tracer (data from Wold et al., 1978).

intestine nuclear RNA, however, the upper and lower strand of the clone 2109A fragment are very differently represented. The lower strand reaction fails to terminate, essentially because of competition from the RNA sequence present in excess (that is, the sequence complementary to the 2109A upper strand). The upper strand reacts completely. We estimate that there are 16–20 transcripts complementary to the upper strand of the 2109A fragment per intestine nucleus, but only about 2–6 transcripts complementary to the 2109A lower strand per nucleus.

The following qualitative conclusions can be drawn from the series of kinetic and titration measurements summarized in Table 2. First, all the cloned repeat families are represented at some level in both gastrula and intestine nuclear RNAs. Second, each repetitive sequence family displays a particular pattern of representation with regard to the number of complementary RNA transcripts per nucleus—that is, some are represented preferentially in intestine nuclear RNA, some in gastrula nuclear RNA and some are about equally represented in the two RNAs. Finally, both strands of each clone are always represented in the nuclear RNAs and are usually, but not always, found at similar concentrations.

Approximate Length of Nuclear RNA Molecules Hybridizing with the Cloned Repeat Fragments

Since the cloned repetitive sequences are only a few hundred nucleotides long, the question arises whether they might react with nuclear RNA molecules of a similar size. Figure 5 shows an experiment in which the gastrula nuclear RNA preparation was separated into size classes (I, II and III of Figure 5A) by velocity sedimentation in a denaturing DMSO sucrose gradient. The hnRNA in the gastrula nuclear RNA preparations is probably slightly degraded. Thus the size distribution shown in Figure 5A may not represent the original in vivo size distribution. Fractions I, II and III were reacted separately with the upper strand of the clone 2090 repeat fragment (Figure 5B) and with the upper strand of the clone 2137 fragment (Figure 5C). Equal *volumes* of each RNA size fraction were used to titrate excess quantities of the cloned tracer. Thus the amount of a cloned tracer capable of hybridizing with a given volume of each size class solution indicates the relative concentration of complementary transcripts in that RNA size class. Our results are summarized in the inset histograms of Figures 5B and 5C. The experiment shows that the clone 2090 fragment reacts primarily with RNAs of the largest size class (that is, >19.5S), while the clone 2137 fragment reacts mainly with the middle size class (that is, 11S–19.5S). A similar result was

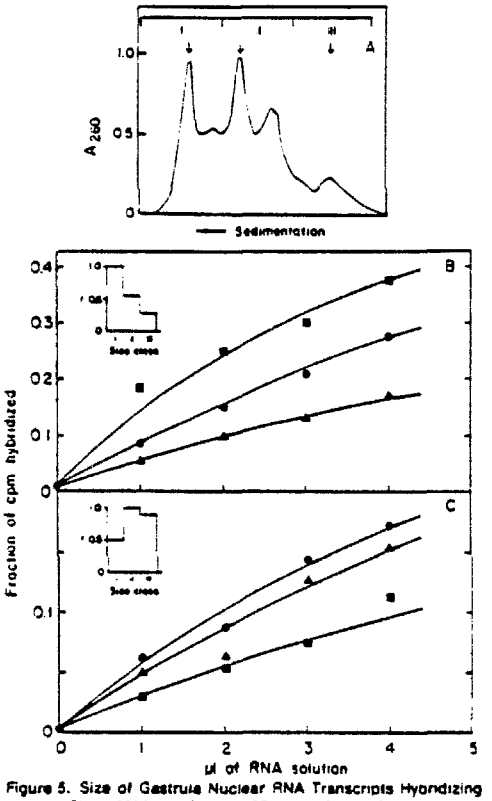


Figure 5. Size of Gastrula Nuclear RNA Transcripts Hybridizing with the Clone 2090 and Clone 2137 Upper Strand Fragments

(A) The gastrula nuclear RNA preparation was denatured in 80% DMSO at 60°C and fractionated by velocity sedimentation in 60% DMSO sucrose gradients at 25°C exactly as described by Costantini et al. (1978). Three size fractions were collected using the positions of the ribosomal RNA subunits as size markers. Fraction I is >19.5S, II is 11S–19.5S and III is 2S–11S. These fractions were dialyzed, precipitated and suspended in equal volumes of 0.5 mM Na acetate (pH 6.5).

(B and C) The volume of each RNA size fraction (I, II or III) indicated on the abscissa was reacted with constant quantities of the cloned fragments: RNA fraction I (■), fraction II (●) and fraction III (▲). The titration curves were analyzed by least-squares methods according to equation (2) (see text). The inset histograms show the relative number of transcripts for each RNA size class as a fraction (f) of the number in the size class with the largest number of transcripts. (B) Upper strand of the clone 2090 fragment. (C) Upper strand of the clone 2137 fragment.

obtained by Costantini et al. (1978) in reactions between several cloned repeat tracers and mature oocyte RNA. Considering that some degradation of the RNA is probable, these data would appear to exclude the possibility that the complementary transcripts exist primarily as small molecules of about the same dimensions as the interspersed repetitive sequences.

Quantitative Pattern of Representation of Cloned Repeats

As described in the accompanying paper, Costantini et al. (1978) have measured the number of transcripts per mature oocyte which are complementary to the same nine cloned repeats as were used in the present experiments. Measurements are described here for two different nuclear RNAs. The nine cloned fragments belong to repetitive sequence families differing greatly from each other with respect to genomic reiteration frequency. To be able to compare the extent to which these repeat families are represented in the different RNAs, we define a parameter termed "representation:"

$$\% \text{ representation} = \frac{T_c}{F_c T_{sc}} \times 100 \qquad (5)$$

where, as above, T_c is the number of transcripts complementary to the cloned tracer per oocyte or per gastrula or intestine nucleus, T_{sc} is the number of copies of a typical single-copy transcript in the oocyte or the nuclear RNAs and F_c is the genomic reiteration frequency of the repeat family to which a given clone belongs. The rationale for equation (5) is that if all the members of a repeat family were being transcribed at the same rate as the average single-copy sequence, and the repeat transcripts were processed, turned over and accumulated with the same kinetics as the single-copy transcripts, each repeat family would produce transcripts per cell or nucleus in proportion to its frequency in the genome. Thus dividing T_c by F_c normalizes for the different repeat families according to their size, and dividing by T_{sc} normalizes for the different steady state levels of single-copy transcripts accumulated in the three RNAs studied. In the oocyte RNA, the number of copies of each typical single-copy sequence (T_{sc}) is approximately 1600 (Galau et al., 1976; Hough-Evans et al., 1977), while as stated above, in the gastrula nuclear RNA, T_{sc} is 1 or less, and in intestine nuclear RNA, T_{sc} is about 0.1. Were the repeat transcripts which we observe known to be interspersed among covalently linked single-copy transcripts which are typical with respect to turnover and prevalence, the representation as calculated by equation (5) would directly indicate the percentage of the repeat family members being transcribed. As yet, however we have no direct knowledge of either the sequence organization or the synthesis and turnover kinetics of these transcripts. Thus we cannot interpret the representation values obtained by equation (5) in a simple or unique way. Calculation of the representation parameter nonetheless yields an interesting and significant set of comparisons.

Table 2 shows that the numbers of copies of transcripts complementary to the cloned repeats vary from about 3000–100,000 for the oocyte (average ~20,000), from about 1–600 for the gastrula nucleus (average ~100) and from about 1–50 for an intestine nucleus (average ~10). There are two important conclusions to be drawn from these values. First, in each RNA, the number of transcripts of individual repeat families per nucleus (or oocyte) varies by two orders of magnitude or more. Second, comparing the three RNAs, the average numbers of transcripts for given repeat families per nucleus or per oocyte are in roughly the same proportion as the numbers of single-copy transcripts. Thus there are an average of about 10 times more single-copy transcripts of each sequence per gastrula nucleus than per intestine nucleus, and about 10^3 times more single-copy transcripts of each sequence per oocyte than per gastrula nucleus. This rough correspondence provides some empirical justification for the presence of T_{sc} in the denominator of equation (5).

The representation values calculated for each cloned sequence in the oocyte RNA, gastrula nuclear RNA and intestine nuclear RNA are shown in Table 2. The lowest representation value in each RNA is on the order of 1%, and the highest value is just under 100% for the intestine and gastrula nuclear RNAs and several hundred percent for the oocyte RNAs. In other words, even when the concentration of transcripts from each repeat family is normalized for the family reiteration frequency, the striking differences in the extent to which each family is represented in a given RNA are not decreased. The large variations in the representation of different repeat families therefore demonstrate differences in the transcriptional expression of these families. Transcript concentration per cell or nucleus does not depend simply upon whether the family is large or small.

The pattern of representation among the nine repeat families is displayed in Figure 6 as a series of histograms which show the representation of each strand of each cloned tracer. In a given RNA, the differences in representation between the various repeat families are generally much greater than the differences between the complementary strands of each family. Changes in the pattern of representation when the different RNAs are compared are best illustrated in Figure 7. Here the representations observed for the nine clones are normalized for comparison by setting the sum of the three observations for each equal to unity. It is evident that the representation of the clone 2034 fragment is high in intestine nuclear RNA and very low in the other RNAs (Figure 7A); the clone 2109B fragment is highly represented in gastrula nuclear RNA but relatively little represented in the other

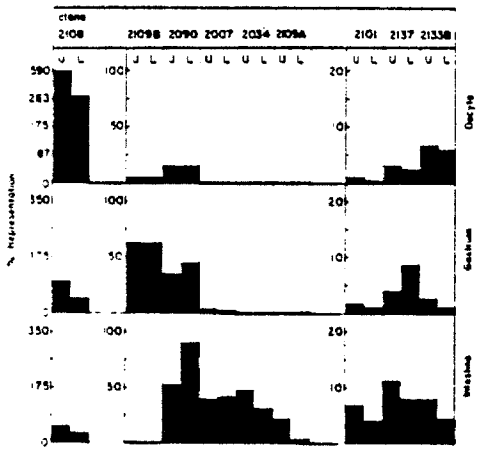


Figure 6. Representation of Cloned Repeat Families

Representation values (ordinates) are from Table 2 and were calculated as indicated in equation (5). The top row of histograms shows the representation of each strand of the nine clones in oocyte RNA. These measurements are from Costantini et al. (1978). The second row of histograms indicates the representation of each strand of the nine clones in gastrula nuclear RNA, and the third row shows their representation in intestine nuclear RNA.

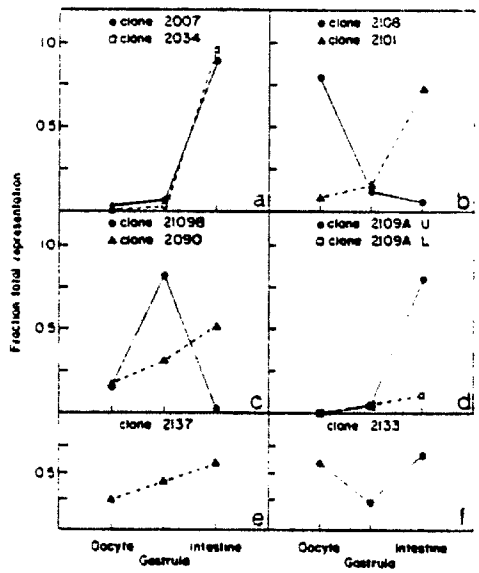


Figure 7 Representation Patterns of Cloned Repetitive Sequence Families in Different RNAs

Representation data for each cloned repeat family are from Table 2. The data have been normalized by dividing the percentage of representations for each RNA by the sum of the percentage of representation for all three RNAs. Upper strand (U) representation was used in each case, except for (d): (a) (—●—) 2007 U, (- -□- -) 2034 U; (b) (—●—) 2108 U, (- -▲- -) 2101 U; (c) (—●—) 21098 U, (- -▲- -) 2090 U; (d) (—●—) 2109A U, (- -□- -) 2109A L; (e) (- -▲- -) 2137 U; (f) (—●—) 2133 U.

RNAs (Figure 7C); the clone 2108 sequence is represented to the greatest extent in oocyte RNA (Figure 7B) and so forth.

To summarize, these experiments show that the transcripts of individual repetitive sequence families are present to very different extents in different cell types. The concentrations of specific repeat transcripts within each RNA differ by as much as two orders of magnitude. The factors which could control representation include the rates of transcription per repeat element in the genome, the fraction of the genomic copies being transcribed and the transcript turnover rates. It is clear that representation is not simply determined by the genomic reiteration frequency. Whatever the mechanism, the data demonstrate that the concentration in RNA of the specific repetitive sequence transcripts is a function of the state of differentiation of the cell.

Discussion

These experiments show that different sets of repetitive sequence families are highly represented in the RNA of each of three cell types. Transcripts of highly represented repeat families may be present at over 100 times greater concentration than those of relatively nonrepresented families in a given RNA. Yet it seems important that at least some transcripts of all nine cloned repeat families were found in all the RNAs. The least prevalent repetitive sequence transcripts are found at concentrations close to those of typical single-copy transcripts (Table 2). The experiments of Costantini et al. (1978) described in the accompanying paper support these conclusions independently. Their study shows that a repetitive sequence tracer prepared from genomic DNA reacts almost completely with excess mature oocyte RNA at high RNA Cot. A minor subfraction of this tracer reacts much more rapidly, however, due to the high concentration of its transcripts. Thus observations made with both genomic and cloned repetitive sequence tracers and relying on several independent methods of measurement show that repeat transcripts have a specific pattern of concentration. This complex situation exceeds the resolving power of the early hybridization experiments mentioned in the Introduction. Nonetheless, the general import of those experiments—that repetitive sequence representation changes with state of differentiation—seems confirmed by our present results.

A result which is not predictable from previous data is that both complements of every repeat

family are represented in the RNA. No exceptions to this rule have been discovered, and usually the two strands are present in roughly equivalent concentrations. This is summarized graphically in Figure 6. The greatest asymmetry in strand representation which we observed is 3–10 fold. Since the representation of different repeat families may vary by two orders of magnitude or more, both strands of each family seem to participate in the cell type-specific expression of that family. The simplest, though certainly not the only, explanation for the expression of both complements is that multiple members of each repeat family are being transcribed asymmetrically in separate transcription units. These members are likely to be interspersed in distant regions of the genome and would therefore be independently oriented. If both orientations occur frequently with respect to the direction of transcription, the two strands would be represented about equally in the RNA, as is generally observed. This explanation is consistent with the interpretation that many members of each highly represented repeat family are being transcribed. An alternative explanation which cannot at present be excluded is that repetitive sequence transcription is actually symmetrical at most if not all sites. Hough et al. (1975) showed that the single-copy sequences of hnRNA are asymmetrically represented in gastrula nuclei. Thus symmetric representation of repeats in the nuclear RNA could not be explained by symmetric transcription at given sites if the RNA repeats are covalently associated with typical single-copy transcripts. Unfortunately, the nature of the RNA sequences flanking the repeat transcripts is not yet known, and this question must remain unanswered until more information is obtained.

Measurements made by reacting single-copy tracers with the same RNAs as those used in these investigations show that approximately 6% of single-copy sequence is represented in oocyte RNA, and perhaps 30% in the two nuclear RNAs (Hough et al., 1975; Anderson et al., 1976; Galau et al., 1976; Kleene and Humphreys, 1977; Hough-Evans et al., 1977). According to these studies, overlaps in the single-copy sequence sets among cell types are large. Thus only a minor fraction of the total genomic single-copy sequence complement can so far be accounted for in RNA transcripts. In contrast, *all* of the repetitive sequence families appear to be represented at some level in each of the three RNAs studied here (see also Costantini et al., 1978). Furthermore, six of the nine cloned repeat families are represented to an extent >50% in one or another of these three RNAs. It does not seem improbable that every middle repetitive sequence family in the genome may be represented in RNA in a cell type-specific way. Perhaps all the repeat families are used in the sense of transcription of some members, while only a minor portion of the single-copy sequence is transcribed.

Repetitive sequences could be included in polysomal mRNA as short RNA sequence elements covalently linked to single-copy message sequences, or they could be present on separate molecules as transcripts of repetitive structural genes. An example of the latter is the histone message set. Previous studies have shown that repeat transcripts are not covalently associated with single-copy sequence transcripts in polysomal mRNAs of sea urchin embryos (Goldberg et al., 1973), HeLa cells (Klein et al., 1974) or rodent cells (Campo and Bishop, 1974; Rabbitts and Milstein, 1975; Legler and Cohen, 1976), although Dina, Meza and Crippa (1974) claimed such an organization for the messenger RNAs of Xenopus embryos. Were transcripts of the cloned repetitive sequences covalently associated with single-copy polysomal message, this association should be easily demonstrable. The alternative possibility that the cloned repeats are derived from repetitive structural genes seems somewhat improbable, a priori, given their short length, their sequence divergence (Klein et al., 1978) and the observation of Costantini et al. (1978) that both strands of the cloned repeat tracers are represented in the RNA of the mature oocyte. Like other mRNAs, the bona fide maternal messages of oocyte RNA are present asymmetrically. This has been shown most recently in experiments carried out with cloned structural genes for sea urchin oocyte maternal messages (T. L. Thomas, R. J. Britten and E. H. Davidson, unpublished observations).

To determine whether their transcripts are associated to a significant extent with messenger RNA, we reacted several of the cloned repetitive sequences with polysomal RNA from blastula stage embryos. In a representative experiment, the separated strands of the clone 2109B repeat fragment were reacted with a 5×10^6 fold mass excess of polysomal RNA. Complementary transcripts proved to be sufficiently rare in the polysomal RNA that this RNA/DNA ratio resulted in a tracer DNA sequence excess for both strands. Using equations (3) and (4), we calculated from the data obtained in these reactions that about 2.5 transcripts complementary to the upper strand and 4.1 transcripts complementary to the lower strand are present in the polysomal RNA of each typical blastula cell. These numbers are comparable to the numbers per cell of rare or complex class mRNAs transcribed from single-copy genes (Galau et al., 1974), although the clone 2109B sequence occurs in the DNA approximately a thousand times per haploid genome. We believe that the small amount of clone 2109B representation observed in polysomal RNA

is due to contamination either with nuclear or with nonpolysomal maternal RNAs persisting from the oocyte, although other explanations cannot be excluded. Table 2 shows that the clone 2109B transcript appears in several hundred copies per gastrula nucleus and in almost 83,000 copies per oocyte. In no case have we observed a high representation of a cloned repeat in the blastula polysomal RNA compared with its representation in oocyte RNA or gastrula nuclear RNA.

What is the biological meaning of the specific patterns of repeat transcript concentration? It is highly improbable that the short repetitive sequences are structural genes, given their characteristics and their symmetrical representation in oocyte RNA. Furthermore, one of the specifically represented cloned repeats studied here, 2109B, contains no less than 18 translational termination signals, including both orientations and all possible reading frames (J. W. Posakony, R. J. Britten and E. H. Davidson, unpublished observations on primary sequence data). Although the interspersed repeat transcripts do not seem to have the characteristics of polysomal message, they are clearly associated with nuclear RNA, as shown by this and earlier work (for example, Darnell and Balint, 1970; Jelinek et al., 1973; Smith et al., 1974). A hypothesis which may be relevant to our present results is that hnRNA has an intranuclear regulatory function mediated by the interspersed repetitive sequence transcripts. In this extension of our earlier gene regulation models (Davidson et al., 1977), we argued that the *concentration* of transcripts of different repeat families could control the sequence-specific interactions required for each cell to express an appropriate set of structural genes. We predicted that repeat transcript *concentrations* should vary greatly in nuclear RNAs, comparing different repetitive sequence families and different cell types. This proposition, in fact, led us to undertake the present experiments, and as far as they go, the prediction is confirmed. We remain completely ignorant of the way in which the RNA repeat transcripts are actually used, if at all. Nonetheless, it is interesting to consider the observations summarized in Table 2 and Figure 6 from the point of view of these regulatory hypotheses. The repeat transcripts of the hnRNA could perform sequence-specific regulatory interactions in any of several ways. One possibility is that they interact with complementary repetitive sequences in the DNA and, as previously proposed, mediate structural gene transcription (Britten and Davidson, 1969; Davidson et al., 1977). In this case, the presence of transcripts of both strands would suggest that promoters of both orientations are used. A closely related alternative proposition is that sequence-specific regulatory interactions take place between complementary hnRNA molecules rather than with the genomic DNA. Federoff and Wall (1976) and Federoff, Wallauer and Wall (1977) showed that hnRNA of HeLa cells can partially renature to form RNA-RNA duplexes. Their experiments also indicated that at least a large part of these duplexes hybridize with repetitive DNA sequences. Our present findings demonstrate that the set of such duplexes which could be formed in a given hnRNA would be cell type-specific, since this process, like any renaturation, should depend upon the sequence concentration of the reacting strands. RNA-RNA duplexes could serve as processing sites for mRNA precursors, as also suggested by other investigators (for example, Federoff et al., 1977). In terms of the kinetics of the putative repetitive sequence reactions, the genomic and hnRNA sequence organizations required and the cell type specificity of repeat transcript concentrations, the RNA-DNA and RNA-RNA regulation models are very similar. This is shown diagramatically in Figure 8. There is, of course, no evidence that either RNA-RNA or RNA-DNA complexes exist in the nucleus, nor is there any convincing reason to rule out the possibility of either. It is interesting to consider a purely hypothetical calculation in which the real time for completion of either type of reaction is estimated from the tran-

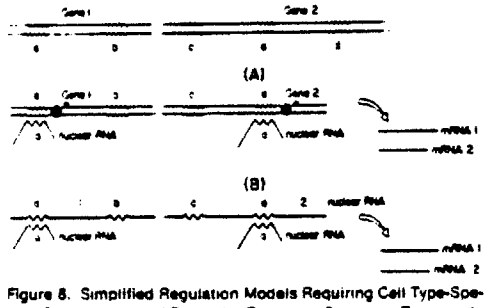


Figure 8. Simplified Regulation Models Requiring Cell Type-Specific Concentration of Particular Repetitive Sequence Transcripts

The top line represents a region of genomic DNA containing two structural genes and various short repetitive sequences (a–d), indicated by double wavy lines, interspersed with single-copy sequences, indicated by double straight lines. In (A), an RNA-DNA interaction scheme is indicated. Transcription of genes (1) and (2) on the DNA occurs when the contiguous repetitive sequence "a" reacts with an hnRNA molecule containing an homologous repeat element, "a'" (single wavy line). An R loop-like triplex or some other sequence-specific structure could be formed. The tailed solid circle indicates a polymerase transcription complex. In (B), the sequence-specific interaction is at the nuclear RNA level. Key as above. The RNA-RNA duplexes at "a" would serve as processing sites for excision of the same mRNA sequences as formed in (A). Either the (A) or (B) type of interaction would be mediated by the concentration of the "a" repeat transcript in the hnRNA. These models both require that specific repeats be adjacent to structural genes, as shown in the top line. For discussion, see the text and Britten and Davidson (1969), Davidson and Britten (1973) and Davidson et al. (1977).

script concentration listed in Table 2. To carry out such a calculation, we suppose arbitrarily that the RNA-driven rates of reaction in the milieu within a gastrula nucleus are those observed under standard conditions in vitro. The highly represented clone 2109B repeat provides an example. Table 2 shows that there are 640 copies of each strand of this transcript per gastrula nucleus, and the volume of each nucleus is approximately 4 μ^3. At the resulting transcript concentration, the reaction of transcripts of the 2109B families with complementary DNA or RNA targets would be 90% complete in only 30 sec. In comparison, the reaction of transcripts present at <1% of this concentration—for example, those complementary to the clone 2133B fragment—would require about twice the hnRNA half-life to reach completion.

Whatever its functional significance, cell type specificity in the representation of specific repeat families appears to have important implications for the study of genomic sequence organization. If the differences in repeat family representation in hnRNA result from differences in the extent to which they are transcribed, there must exist an organized relation between the location of transcription units and the location of the members of each repetitive sequence family in the genome.

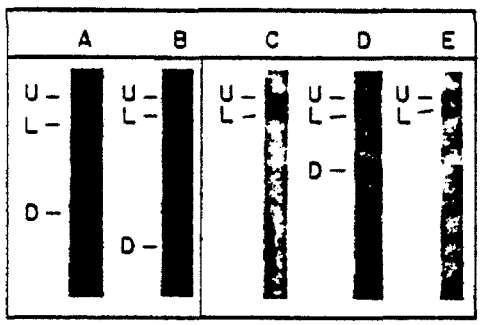


Figure 9. Strand Separation of Cloned Repetitive Sequences by Polyacrylamide Gel Electrophoresis

Supercoiled DNA was cleaved with Eco RI to release the repetitive DNA insert and then treated with bacterial alkaline phosphatase as described in the text. The 5′-hydroxyl termini of the fragments were labeled with γ-^{32}P-ATP, and the DNA was concentrated and loaded in the denatured state on polyacrylamide gels. (A and B) show ethidium bromide staining of the strand-separated clone 2090 and clone 2101 fragments, respectively. The clone 2090 repeat fragment in (A) is 190 nucleotides long (Table 1). This gel was 8% polyacrylamide. The clone 2101 fragment of (B) is 320 nucleotides long and was run on a 4% gel. The slower moving band of the doublet is referred to as upper strand and the fast moving complement is referred to as lower strand. The renatured, double-stranded DNA moves more rapidly than the separated single strands, in agreement with previous observations on fragments of this length (Maniatis, Jeffrey and van de Sande, 1975). (C, D and E) demonstrate autoradiography of other strand-separating gels. (C) contains the 2007 fragment, which is 1100 nucleotides long and was run on a 2.5% gel. (D and E) contain the repetitive sequence fragments of clones 2034 and 2090. These are 560 and 190 nucleotides long, respectively. The 560 nucleotide fragment was strand-separated on a 4.0% gel and the 190 nucleotide fragment on a 6% gel, as above.

Experimental Procedures

Preparation, Labeling and Strand Separation of Cloned Repetitive Sequence Fragments

Superhelical plasmid DNA was isolated on CsCl gradients (Scheller et al., 1977). The DNA was dialyzed into 5 mM Tris (pH 7.5) and stored frozen at −20°C. Plasmid DNA was cleaved with 1 μl Eco RI (Eco RI was a gift from P. Green and H. Boyer) per 20 μg of DNA at 37°C for 30 min in 100 mM Tris–HCl (pH 7.6), 100 mM NaCl and 5 mM $MgCl_2$. The reaction mixtures were precipitated with 1 vol of isopropanol at −20°C for at least 2 hr, followed by centrifugation. The precipitated DNA was redissolved in 10 mM Tris at a concentration <50 μg/ml, and 10 μl of bacterial alkaline phosphatase (Worthington, Code BAPF) were added per 40 μg of nucleic acid. The reaction mixture was incubated at 37°C for 30 min, then extracted twice with an equal volume of phenol-Sevag 1:1 (Sevag is a 24:1 chloroform-isoamyl alcohol solution) and once with ether. The DNA was precipitated with isopropanol as above. DNA fragments were labeled by the polynucleotide kinase reaction, essentially as described by Maxam and Gilbert (1977). The DNA pellet was dissolved in 5 mM Tris (pH 9.5), 0.01 mM EDTA, 0.1 mM spermidine at a DNA concentration of 50 μg/ml and heated at 100°C for 3 min. The solution was then brought to 50 mM Tris (pH 9.5), 10 mM $MgCl_2$ and 5 mM DTT. γ-^{32}P-ATP, synthesized by the exchange procedure (Maxam and Gilbert, 1977), was dried down, dissolved in H_2O and added to the reaction mixture at a concentration not lower than 3 μM. 3–5 μl of polynucleotide kinase solution (PL Biochemicals) were added, and the reaction was incubated for 45 min at 37°C and then extracted with phenol-Sevag 1:1 and with ether, as above. The nucleic acid was precipitated with 2 vol of ethanol for at least 6 hr at −20°C and then centrifuged. The pellet was dissolved in 0.3 M NaOH, 10% glycerol and 1 mM EDTA, and heated to 37°C for 10 min. Samples were strand-separated by electrophoresis on 1 × 20 cm cylindrical polyacrylamide gels. The polyacrylamide concentration used depended upon the size of the fragment; all gels contained 0.27% bis-acrylamide. Electrophoresis was in 50 mM Tris borate (pH 8.3), 1 mM EDTA at 15°C for 16 hr at 60 V. Gels were stained with ethidium bromide and autoradiographed, and individual bands were excised with a razor blade.

Typical examples of the strand separation gels are shown in Figure 9. Gel slices were crushed with a siliconized glass rod in a 1.5 ml Eppendorff tube. 1 ml of 0.12 M sodium phosphate buffer (PB), 0.05% SDS and 10 μg of purified calf thymus DNA carrier were added to each sample, and the mixture was incubated overnight at 37°C. The acrylamide was removed by filtering through siliconized glass wool and rinsing with an additional 1 ml of 0.12 M PB, 0.05% SDS. The eluate contained from 80–95% of the radioactivity in the gel band. The 2–2.5 ml sample was concentrated 4 fold by extraction with sec-butanol, resulting in a salt concentration of ~0.5 M PB. DNA was then incubated for 10 hr at 55°C to renature any contaminating complementary strands.

The reaction mixtures were subsequently diluted to 0.12 M PB and passed over a 1 ml column of hydroxyapatite at 50°C. The unbound fraction contained the purified strand-separated repetitive sequence. ^{32}P specific activities of these materials ranged from 3×10^6 to 3×10^7 cpm/μg, depending upon the size of the fragment and the extent of the various reactions. The final strand-separated preparations were contaminated 0–4% with their complementary strands. Reactivity of these DNA preparations with excess total sea urchin DNA ranged from 55–95%. The "nonreactive" radioactivity was shown to be of low molecular weight and is almost certainly γ-^{32}P-ATP persisting from the kinase reaction. The ^{32}P-DNA was stored in 0.12 M PB at −20°C. DNA reactivities

decreased slowly with time but were usually acceptable for at least 3 weeks.

The following experiment was carried out for each DNA preparation to determine the purity of the separated strands eluted from the polyacrylamide gels. Equal quantities of ^{32}P-DNA from the upper (U) and lower (L) strands of each clone were renatured, or the U and L strands were incubated alone. The kinetics of the reactions were measured. Since the two complementary strands had been enzymatically labeled together, their specific activities are expected to be equal and were found to be so (see Figure 4C of the accompanying paper by Costantini et al., 1978). The absolute specific activity could be calculated from the kinetics of the reactions between each pair of complementary strands, given the cloned repeat fragment length (that is, the fragment complexity). The extent of this reaction or the reaction with total sea urchin DNA determines the reactivity of the cloned tracers, and the amount of reaction obtained when each strand is incubated alone indicates its purity with respect to the complementary strand. Representative data for clones 2137 and 2108 are shown in Figure 1. The reactions between complementary strands terminate at about 80% and follow second-order kinetics. The same termination values are observed when the cloned fragments are reacted with excess sea urchin DNA (Klein et al., 1978). The incomplete (100%) reaction of the tracer is due to the persistence of a minute fraction of the γ-^{32}P-ATP originally present in the kinase labeling reactions.

RNA Preparations

Gastrula Nuclear RNA

1.3×10^8 Strongylocentrotus purpuratus eggs were suspended in three 4 liter jars of seawater containing penicillin and streptomycin (Smith et al., 1974) and fertilized. After 2.5 hr, 95.5% of the eggs were at first cleavage, 0.8% were at second cleavage and 3.7% were unfertilized. After 36 hr (early gastrula stage), the embryos were washed in cold Ca–Mg-free seawater (pH 3) and collected by centrifugation at 5000 rpm for 5 min. The embryo pellet was resuspended in 200 ml of cold 2 mM $MgCl_2$, 10 mM PIPES, 10 μg/ml polyvinyl sulfate, and sheared in a Waring blender until most of the cells were disrupted, as assayed by phase-contrast microscopy. Glucose was added to a concentration of 1 M, and the nuclei were pelleted by centrifugation at 5000 rpm for 10 min. The nuclei were lysed with 7 M urea and extracted with phenol-cresol-Sevag solution (0.15/0.85/1.0). The interface was reextracted and combined with the aqueous phase, which was reextracted twice with an equal volume of Sevag solution. The RNA was precipitated with 2 vol of 100% ethanol at −20°C and then collected by centrifugation at 10,000 rpm for 1 hr. The pellet was resuspended and treated with pancreatic DNAase (100 μg/ml) for 2 hr at room temperature with slow stirring. The solution was then brought to 50 μg/ml proteinase K and incubated at 37°C for 1 hr. The RNA was extracted and precipitated as above. The resulting material was chromatographed on Sephadex G-100, and the exclusion peak fractions were pooled and precipitated. The DNAase and proteinase K treatment followed by extraction and precipitation were then repeated. The RNA was stored in 5 mM sodium acetate (pH 6.8) at −20°C. Yields from a preparation this size ranged from 25–35 mg. The complexity and pseudo-first-order rate constants obtained in reactions of the nuclear RNA with single-copy DNA tracer were always very close to those measured by Hough et al. (1975).

Intestine Nuclear RNA

The intestine nuclear RNA preparation is described elsewhere by Wold et al. (1978). Briefly, the intestinal tissue was rinsed in EGTA containing seawater and lysed in a Triton X-100 buffer. The nuclei were pelleted through a 0.5 M sucrose cushion and the RNA was extracted as above.

Blastula Polysomal RNA

Blastula polysomal RNA was prepared from 24 hr swimming blastulae as described by Galau et al. (1976).

RNA-DNA Hybridization

All hybridization reactions were carried out in 0.12 M PB at 50°C or in 0.5 M PB at 55°C. All Cot values referred to in this work are equivalent Cots (that is, the Cot of the reaction corrected for the relative increase in rate due to salt concentrations above 0.18 M Na^+). The mixtures contained 1 mM EDTA, 0.05% SDS and calf thymus DNA carrier. RNA-DNA ratios were 5×10^5 to 1×10^7 for reactions in which RNA excess kinetics were to be measured. Titrations were carried out in ^{32}P-DNA sequence excess—that is, at lower RNA/DNA ratios. Reaction mixtures were sealed in 5–50 μl capillaries and boiled at 105°C for 2 min prior to incubation. The reactions were terminated at appropriate times by instant freezing in dry ice acetone. The possibility of DNA-DNA reaction was monitored by the low salt RNAase method of Hough and Davidson (1972) and Galau, Britten and Davidson (1974). The reaction mixtures were diluted into 2 ml of 0.02 M PB and divided into two equal aliquots. One sample was brought to 50 μg/ml RNAase A, incubated for 1 hr at 37°C, extracted with Sevag solution and adjusted to 0.12 M PB. The other was brought to 0.12 M PB without treatment with RNAase. Both samples were chromatographed on hydroxyapatite columns at concentrations not greater than 50 μg/ml nucleic acid per milliliter of hydroxyapatite. Six 2 ml washes were collected at 50°C and three 2 ml washes were collected at 99°C. The fractions were counted by liquid scintillation, and the fraction of radioactivity eluted at 99°C represented the fraction of DNA in hybrid structures. The DNA-DNA duplex content was always very small (<5%) compared with the DNA-RNA hybrid portion of the reaction.

Titration reactions and DMSO sucrose gradient centrifugation were carried out exactly as described in the accompanying paper by Costantini et al. (1978).

Acknowledgments

We wish to acknowledge the expert assistance of Mr. Clifford Beall. This research was supported by a grant from the NSF. R.H.S. and F.D.C. were supported by NIH training grants to the Division of Chemistry and to the Division of Biology. The P2 facilities used in this work were equipped with the aid of funds from a Biomedical Research Support Grant.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 13, 1978; revised June 8, 1978

References

Anderson, D. M., Galau, G. A., Britten, R. J. and Davidson, E. H. (1976). Sequence complexity of the RNA accumulated in oocytes of Arbacia punctulata. Dev. Biol. *51*, 138–145.

Bonner, T. I., Brenner, D. J., Neufeld, B. R. and Britten, R. J. (1973). Reduction in the rate of DNA reassociation by sequence divergence. J. Mol. Biol. *81*, 123–135.

Britten, R. J. and Davidson, E. H. (1969). Gene regulation for higher cells: a theory. Science *165*, 349–358.

Britten, R. J. and Davidson, E. H. (1971). Repetitive and nonrepetitive DNA sequences and a speculation on the origins of evolutionary novelty. Quart. Rev. Biol. *46*, 111–138.

Britten, R. J., Graham, D. E. and Neufeld, B. R. (1974). Analysis of repeating DNA sequences by reassociation. In Methods in Enzymology, *29E*, L. Grossman and K. Moldave, eds. (New York: Academic Press), pp. 363–406.

Campo, M. S. and Bishop, J. O. (1974). Two classes of messenger RNA in cultured rat cells: repetitive sequence transcripts and unique sequence transcripts. J. Mol. Biol. *90*, 649–663.

Chamberlin, M. E., Galau, G. A., Britten, R. J. and Davidson, E. H. (1978). Studies on nucleic acid reassociation kinetics: renaturation of DNA fragments of unequal length. Nucl. Acids Res. *5*, 2073–2084.

Costantini, F. D., Scheller, R. H., Britten, R. J. and Davidson, E. H. (1978). Repetitive sequence transcripts in the mature sea urchin oocyte. *Cell 15*, 173–187.

Darnell, J. E. and Balint, R. (1970). The distribution of rapidly hybridizing RNA sequences in heterogeneous nuclear RNA and mRNA from HeLa cells. *J. Cell Physiol. 76*, 349–356.

Davidson, E. H. Gene Activity in Early Development, second edition (New York: Academic Press).

Davidson, E. H. and Britten, R. J. (1973). Organization, transcription and regulation in the animal genome. *Quart. Rev. Biol. 48*, 565–613.

Davidson, E. H., Graham, D. E., Neufeld, B. R., Chamberlin, M. E., Amenson, C. S., Hough, B. R. and Britten, R. J. (1974). Arrangement and characterization of repetitive sequence elements in animal DNAs. *Cold Spring Harbor Symp. Quant. Biol. 38*, 295–301.

Davidson, E. H., Hough, B. R., Klein, W. H. and Britten, R. J. (1975). Structural genes adjacent to interspersed repetitive DNA sequences. *Cell 4*, 217–238.

Davidson, E. H., Klein, W. H. and Britten, R. J. (1977). Sequence organization in animal DNA and a speculation on hnRNA as a coordinate regulatory transcript. *Dev. Biol. 55*, 69–84.

Dina, D., Meza, I. and Crippa, M. (1974). Relative positions of the "repetitive," "unique" and poly(A) fragments of mRNA. *Nature 248*, 486–490.

Federoff, N. and Wall, T. R. (1978). Complementary sequences in heterogeneous nuclear RNA. In *Molecular Mechanisms in the Control of Gene Expression*. *5*, W. J. Rudder, D. P. Nierlich and C. F. Fox, eds. (New York: Academic Press), pp. 379–384.

Federoff, N., Wellauer, P. K. and Wall, R. (1977). Intermolecular duplexes in heterogeneous nuclear RNA from HeLa cells. *Cell 10*, 597–610.

Galau, G. A., Britten, R. J. and Davidson, E. H. (1974). A measurement of the sequence complexity of polysomal messenger RNA in sea urchin embryos. *Cell 2*, 9–21.

Galau, G. A., Klein, W. H., Davis, M. M., Wold, B. J., Britten, R. J. and Davidson, E. H. (1976). Structural gene sets active in embryos and adult tissues of the sea urchin. *Cell 7*, 487–505.

Galau, G. A., Britten, R. J. and Davidson, E. H. (1977). Studies on nucleic acid reassociation kinetics: rate of hybridization of excess RNA with DNA, compared to the rate of DNA renaturation. *Proc. Nat. Acad. Sci. USA 74*, 1020–1023.

Goldberg, R. B., Galau, G. A., Britten, R. J. and Davidson, E. H. (1973). Nonrepetitive DNA sequence representation in sea urchin embryo messenger RNA. *Proc. Nat. Acad. Sci. USA 70*, 3516–3520.

Goldberg, R. B., Crain, W. R., Ruderman, J. V., Moore, G. P., Barnett, T. R., Higgins, R. C., Gelfand, R. A., Galau, G. A., Britten, R. J. and Davidson, E. H. (1975). DNA sequence organization in the genomes of five marine invertebrates. *Chromosoma 51*, 225–251.

Harpold, M. M. and Craig, S. P. (1977). The evolution of repetitive DNA sequences in sea urchins. *Nucl. Acids Res. 4*, 4425–4438.

Hough, B. R. and Davidson, E. H. (1972). Studies on the repetitive sequence transcripts of Xenopus oocytes. *J. Mol. Biol. 70*, 491–509.

Hough, B. R., Smith, M. J., Britten, R. J. and Davidson, E. H. (1975). Sequence complexity of heterogeneous nuclear RNA in sea urchin embryos. *Cell 5*, 291–299.

Hough-Evans, B. R., Wold, B. J., Ernst, S. G., Britten, R. J. and Davidson, E. H. (1977). Appearance and persistence of maternal RNA sequences in sea urchin development. *Dev. Biol. 60*, 258–277.

Jelinek, W., Adesnik, M., Salditt, M., Sheiness, D., Wall, R., Molloy, G., Philipson, L. and Darnell, J. E. (1973). Further evidence on the nuclear origin and transfer to the cytoplasm of polyadenylic acid sequences in mammalian cell RNA. *J. Mol. Biol. 75*, 515–532.

Kleene, K. C. and Humphreys, T. (1977). Similarity of hnRNA sequences in blastula and pluteus stage sea urchin embryos. *Cell 12*, 143–155.

Klein, W. H., Murphy, W., Attardi, G., Britten, R. J. and Davidson, E. H. (1974). Distribution of repetitive and nonrepetitive sequence transcripts in HeLa mRNA. *Proc. Nat. Acad. Sci. USA 71*, 1785–1789.

Klein, W. H., Thomas, T. L., Lai, C., Scheller, R. H., Britten, R. J. and Davidson, E. H. (1978). Characteristics of individual repetitive sequence families in the sea urchin genome studied with cloned repeats. *Cell 14*, 889–900.

Legler, M. K. and Cohen, E. P. (1976). Estimation of the number of nucleotide sequences in mouse DNA complementary to messenger RNAs specifying a complete mouse immunoglobulin. *Biochemistry 15*, 4390–4399.

Maniatis, T., Jeffrey, A. and van de Sande, H. (1975). Chain length determination of small double and single stranded DNA molecules by polyacrylamide gel electrophoresis. *Biochemistry 14*, 3787–3794.

Maxam, A. M. and Gilbert, W. (1977). A new method for sequencing DNA. *Proc. Nat. Acad. Sci. USA 74*, 560–564.

Moore, G. P., Scheller, R. H., Davidson, E. H. and Britten, R. J. (1978). Evolutionary change in repetition frequency of sea urchin DNA sequences. *Cell*, in press.

Pearson, W. R., Davidson, E. H. and Britten, R. J. (1977). A program for least squares analysis of reassociation and hybridization data. *Nucl. Acids Res. 4*, 1727–1737.

Rabbitts, R. H. and Milstein, C. (1975). Mouse immunoglobulin genes: studies on the reiteration frequency of light-chain genes by hybridization procedures. *Eur. J. Biochem. 52*, 125–133.

Scheller, R. H., Thomas, T. L., Lee, A. S., Klein, W. H., Niles, W. D., Britten, R. J. and Davidson, E. H. (1977). Clones of individual repetitive sequences from sea urchin DNA constructed with synthetic Eco RI sites. *Science 196*, 197–200.

Schmid, C. W. and Deininger, P. L. (1975). Sequence organization of the human genome. *Cell 6*, 345–358.

Smith, M. J., Chamberlin, M. E., Hough, B. R. and Davidson, E. H. (1974). Repetitive and nonrepetitive sequence in sea urchin hnRNA. *J. Mol. Biol. 85*, 103–126.

Wallace, R. B., Dube, S. K. and Bonner, J. (1977). Localization of the globin gene in the template active fraction of chromatin of Friend leukemia cells. *Science 198*, 1166–1168.

Whiteley, A. H. (1949). The phosphorus compounds of sea urchin eggs and the uptake of radio-phosphate upon fertilization. *Am. Naturalist 83*, 249–267.

Wold, B. J., Klein, W. H., Hough-Evans, B. R., Britten, R. J. and Davidson, E. H. (1978). Sea urchin embryo mRNA sequences expressed in the nuclear RNAs of adult tissues. *Cell 14*, 941–950.

Wu, J.-R., Pearson, W., Wilkes, M. and Bonner, J. (1977). Analysis of the sequence structure of the rat genome. In *The Molecular Biology of the Mammalian Genetic Apparatus, II*, P. Ts'o, ed. (Amsterdam: Elsevier North-Holland), pp. 51–62.

CHAPTER 7

Cell, Vol. 15, 649–660, October 1978, Copyright © 1978 by MIT

Evolutionary Change in the Repetition Frequency of Sea Urchin DNA Sequences

Gordon P. Moore,* Richard H. Scheller,†
Eric H. Davidson* and Roy J. Britten*‡
*Division of Biology
†Division of Chemistry
California Institute of Technology
Pasadena, California 91125

Summary

The frequency of occurrence of particular repetitive sequence families has been estimated in the DNA of the three sea urchin species Strongylocentrotus purpuratus, Strongylocentrotus franciscanus and Lytechinus pictus using individual cloned S. purpuratus repetitive sequence elements. Cloned repetitive sequence elements as described by Scheller et al. (1977a) were prepared by reassociation of S. purpuratus DNA fragments to repetitive Cot, digestion with single-strand-specific nuclease S1 and ligation of synthetic restriction sites to their ends. The sequences were cloned by insertion at the Eco RI site of plasmid RSF2124, labeled, strand-separated and reassociated with 800–900 nucleotide long unlabeled DNA. Both kinetic (genomic DNA excess) and saturation (cloned DNA excess) estimates of frequencies were made. For nine cloned fragments, the ratio of the repetition frequency in S. purpuratus DNA to that in S. franciscanus DNA ranges from about 20 to about 1. In the four cases examined, only a few copies were detected in the DNA of L. pictus. Estimates have also been made of frequency changes in many repetitive families by measuring the reassociation of labeled repetitive DNA fractions of each species with total DNA from other species. In each reciprocal comparison, the labeled repetitive sequences reassociate more slowly with DNA of other species than with DNA of the species from which they were prepared. Thus it appears that the dominant repetitive sequence families in the DNA of each species are present at lower frequencies in the DNA of closely related species. Measurements of thermal stability have been made of S. purpuratus cloned repetitive sequences reassociated with S. franciscanus DNA or S. purpuratus DNA. Most families have changed both in frequency and sequence, although some have changed little in sequence but show great changes in frequency.

Introduction

Comparisons of repeated DNA among related eucaryotic species have shown striking similarities in the quantity of repeated DNA, the frequency distribution and the degree of divergence in sequence among copies in a genome. Nevertheless, repetitive DNA from one species reassociates incompletely with the DNA of other species. As the evolutionary distance between the species is increased, the amount of repetitive sequence held in common decreases and the intergenomic sequence divergence increases. These observations were interpreted to indicate that repetitive sequences are continually being added to the genome in the course of evolution (Britten and Waring, 1965; Britten and Kohne, 1967, 1968). Many families of repetitive sequences were supposed to undergo multiplication and/or sequence divergence. The large number of families of repetitive sequences and the wide range of frequency and sequence divergence have so far prevented a direct study of the processes of gain and loss and divergence. With the advent of recombinant DNA technology and the availability of clones of individual sea urchin repetitive DNA sequences, it is now possible to measure differences between individual families of repeated sequences in the DNA of related species. From such measurements, we expect to be able to infer the processes of evolutionary change of repetitive DNA.

The relationship of the three species of sea urchins used in this work has been reviewed by Durham (1966), and some measurements have been published on the single-copy DNA differences (Angerer, Davidson and Britten, 1976; Harpold and Craig, 1978) and on repetitive sequences (Harpold and Craig, 1977). It is probable that the lines of descent leading to S. franciscanus and S. purpuratus diverged from each other 10 to 20 million years ago, and those for the genera Lytechinus and Strongylcentrotus about 150 to 200 million years ago (Durham, 1966). In recent measurements (T. J. Hall and R. J. Britten, unpublished data), the single-copy DNA relatedness of S. franciscanus and S. purpuratus has been studied at a lower criterion of precision. With nearly complete reactions, the T_m of interspecies single-copy sequence duplexes is approximately 10°C lower than that for intraspecies duplexes. Studies with L. pictus and S. purpuratus have been carried out only at high criteria (Angerer et al., 1976; Harpold and Craig, 1978), and it appears that approximately 10% of the single-copy DNAs of the two species forms duplexes which melt within 15°C of intraspecies duplexes. This paper presents measurements which show that during recent sea urchin evolution, many repetitive sequence families have changed greatly in size and that the sequence change is less for repetitive DNA than for single-copy DNA.

‡ Also a Staff Member of the Carnegie Institution of Washington, Baltimore, Maryland 21210.

Results

Thermal Stability of Interspecies Duplexes

The measurement of the thermal stability of DNA duplexes is a standard method for the estimation of average differences in sequence between related DNAs, and is often used to study interspecies differences in either single-copy or repetitive DNA. This method has recently been used to determine the sequence divergence in S. purpuratus DNA of a number of individual families of repetitive sequence represented by cloned DNA fragments, including those used in this work (Klein et al., 1978). Sequence divergence is not only of interest in assessing evolutionary change but is technically important to the estimation of repetition frequency, since divergent sequences reassociate at reduced rate (Bonner et al., 1973). Sequences that are sufficiently similar to form stable duplexes under our conditions are termed "recognizable sequences." It is worthwhile to differentiate between the effect of sequence divergence on the rate of reassociation (which might lead to an underestimation of the number of recognizable sequences) and the possibility that many sequences related to a family may be so divergent as to be unrecognizable.

The repetitive sequence clones were constructed by Scheller et al. (1977a) from sea urchin DNA that had been renatured to Cot 40 to form repeat duplexes and then treated with S1 nuclease to remove single-stranded tails. Synthetic double-stranded decamers containing the recognition site for Eco RI endonuclease (Scheller et al., 1977b) were ligated to the S1 nuclease-resistant repeat duplexes. The repetitive DNA sequences were then ligated into the Eco RI site of plasmid vector RSF2124 (So, Gill and Falkow, 1975) and cloned. The sea urchin DNA fragments can be separated from the plasmid vector DNA by Eco RI endonuclease digestion, and for this work they were labeled at their 5' termini with polynucleotide kinase. For the purpose of the hybridization experiments and to determine whether repetitive sequence representation in the RNA is symmetrical, it was necessary to separate the two complementary strands of each cloned DNA fragment. This was accomplished preparatively by electrophoresis in neutral polyacrylamide gels loaded with alkali-denatured DNA fragments.

We have measured the thermal stability of cloned DNA fragments reassociated with total DNA from S. franciscanus, and in two cases L. pictus, compared to the products of reassociation of S. purpuratus DNA. Figure 1 shows a typical thermal stability measurement for one of the cloned repeats, CS2101, and Table 1 summarizes a number of measurements using the same procedure. The particular cloned repetitive sequence used for the measurements of Figure 1 is of interest, since the

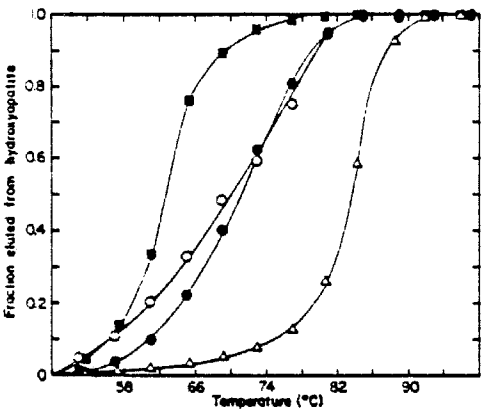


Figure 1. Hydroxyapatite Chromatography of Duplexes Formed with the Repeat Fragment of Clone CS2101 and the Genomic DNA of Three Sea Urchin Species

Sea urchin DNAs were (●) S. purpuratus, (○) S. franciscanus or (■) L. pictus. Reassociation was at 0.41 M PB, 0.05% SDS, 0.001 M EDTA at 50°C to equivalent Cot 2000. Extent of reaction was 75–85% in each case. The weight mean driver DNA fragment length was approximately 650 nucleotides. Samples were diluted to 0.12 M PB and applied to a series of 0.4 g HAP columns at 50°C. The temperature was increased in approximately 4°C increments and the DNA fragments rendered completely single-stranded were eluted with 0.12 M PB. Native ^{3}H-DNA fragments (△) of 450 nucleotides mean length were included as a standard.

thermal stability when reassociated with S. franciscanus DNA is only 0.8°C lower than when reassociated with S. purpuratus DNA. The T_m and shape of the melting curve with L. pictus driver DNA are probably affected by the criterion of incubation. Column 5 of Table 1 shows that the average divergence within this family of sequences in S. purpuratus leads to a T reduction of approximately 5°C. Very little more divergence results from the evolutionary sequence changes that have occurred in the 10 or 20 million years since the lines leading to S. purpuratus and S. franciscanus diverged. In sharp contrast, the average single-copy DNA sequence change produces a 10°C reduction in T_m. Some process has acted to stabilize the many sequences of this family. Three of the six repeats included in this table (clones CS2101, CS2109B and CS2090) show little interspecies compared to intraspecies divergence. Both repeat families represented by clones CS2109B and CS2090, however, show large intraspecies divergence and low thermal stabilities. With so great an intragenomic divergence, the criterion of precision set by the incubation conditions prevents the recognition of many copies that may be in these families. Similarly, in intraspecies comparisons, the additional divergence may lead to loss of recognition, and it is difficult to estimate the actual numbers of copies present. The two repetitive sequence fractions from S. purpuratus show an

Table 1. Thermal Stability of Cloned DNA Duplexes

Clone	Native T_m[a]	Driver DNA	T_m (50°C)[b]	S. purpuratus T_m – S. franciscanus T_m
CS2007	72.7	S. purpuratus	59.5	5.5
		S. franciscanus	54.0	
CS2034	79.4	S. purpuratus	76.0	5.5
		S. franciscanus	70.5	
		L. pictus	57.0	
CS2090	83.1	S. purpuratus	64.5	−1.0
		S. franciscanus	65.5	
CS2101	76.7	S. purpuratus	70.8	0.8
		S. franciscanus	70.0	
		L. pictus	63.4	
CS2108B	83.4	S. purpuratus	56.8	−1.5
		S. franciscanus	58.3	
CS2137		S. purpuratus	64.0	6.0
		S. franciscanus	58.0	
High Frequency Repetitive S. purpuratus DNA (Figure 4A)		S. purpuratus	74.5	4.0
		S. franciscanus	70.5	
Intermediate Repetitive S. purpuratus DNA (Figure 5)		S. purpuratus	70.0	3.5
		S. franciscanus	66.5	

[a] The native T_m is that of the nondenatured or renatured cloned DNA and is a function of the length and base composition. These values are from Klein et al. (1978).

[b] The driver lengths were approximately 650 nucleotides in each case. Reactions were run to at least 50 times their kinetically determined $Cot_{1/2}$, usually to Cot 1000. Termination points varied depending upon tracer reactability, but were usually 75–85%. The cloned repetitive sequence elements were reacted to termination with 100 μg of the indicated driver DNAs in 0.41 M PB, 0.05% SDS, 1 mM EDTA at 50°C. The reaction mixture was diluted to 0.12 M PB and applied to a 0.4 g hydroxyapatite column at 50°C. The temperature was raised in 4°C increments, and the DNA rendered single-stranded at each increment was eluted with three 2.7 ml aliquots of 0.12 M PB. An internal standard, consisting of native S. purpuratus ^{3}H–DNA of 450 nucleotides, was added to the cloned repetitive ^{32}P–DNA reaction mixture, and the thermal elution profiles of both DNAs were obtained simultaneously. The native DNA standards had a mean T_m of 83.5°C ± 0.5°C (SD). The T_m's listed have been adjusted by adding or subtracting any deviation from the mean in the internal standard T_m (Klein et al., 1978). In most cases, these adjustments were <0.5°C and the largest adjustments were 1.0°C.

interspecies divergence of 3–4°C, which is less than that of the single-copy DNA (10°C) measured under similar low criterion conditions (T. J. Hall and R. J. Britten, unpublished data). Taken together, these measurements suggest that sea urchin repetitive sequences have changed less than single-copy DNA within the genus Strongylocentrotus, as previously indicated by Harpold and Craig (1977) and by R. C. Angerer (unpublished observations).

Measurement of Repetition Frequency by Reassociation Kinetics

When a labeled DNA fraction is reassociated with an excess of total nuclear DNA, the rate of reassociation compared with the rate of reassociation of single-copy DNA sequences is a direct measure of *the number of copies present. The rate is also* somewhat affected by sequence divergence and base composition, but these corrections have a minor effect on our conclusions. Figure 2 shows the reassociation of eight cloned S. purpuratus repetitive sequences with S. purpuratus, S. franciscanus or L. pictus DNA driver present in a 10^6 fold mass excess. The reassociation and hydroxyapatite analysis were carried out at a relatively low criterion to include divergent sequences [incubation at 0.41 or 0.12 M phosphate buffer (PB), 50°C and load at 0.12 M PB, 50°C]. In the reassociation with S. franciscanus and S. purpuratus DNA, the reactions go nearly to completion, as expected with excess driver DNA. There is approximately 10–20% labeled ATP remaining in these strand-separated DNA preparations.

In six of nine cases, the reaction with S. franciscanus DNA is significantly slower than with S. purpuratus DNA. Table 2 shows the estimates of repetition frequency after correction for length based on rate constants derived from least-squares analysis of reassociation kinetics. For the three *families represented by clones* CS2108, CS2090 and CS2099, there is apparently no significant difference in the number of recognizable copies per genome between S. franciscanus and S. purpuratus.

A limited set of measurements was made with L.

Cell
652

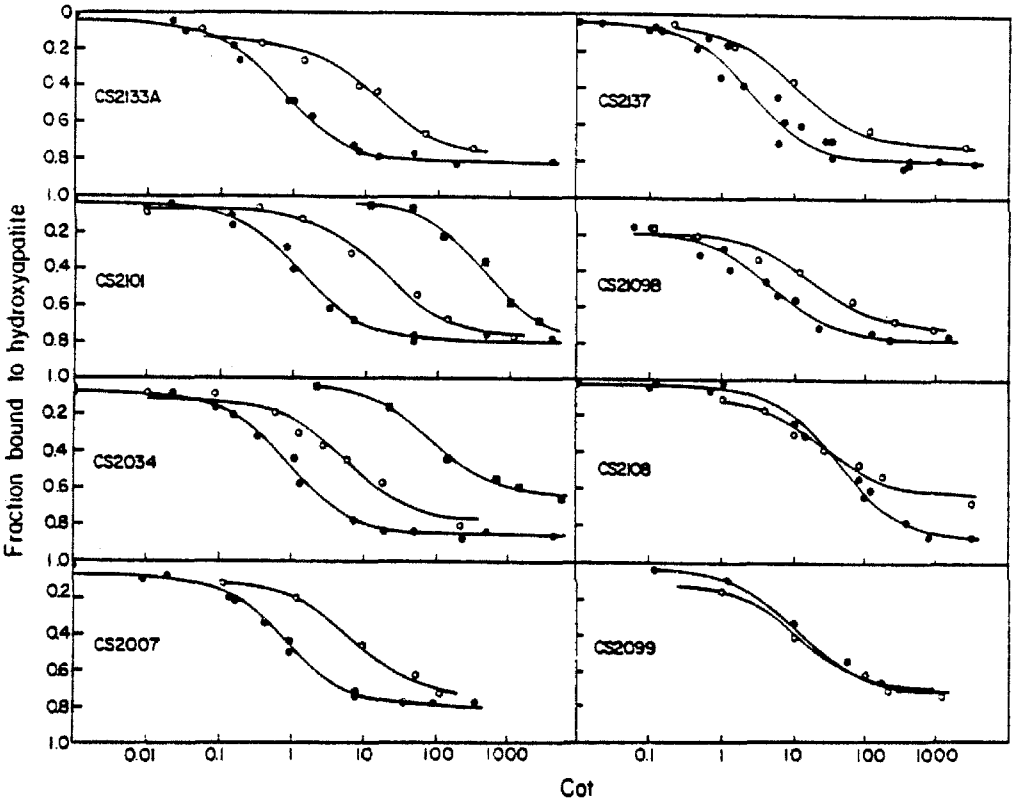


Figure 2. Reassociation of Eight Cloned Repetitive S. purpuratus Sequences with the Genomic DNA of Three Sea Urchin Species

Sea urchin DNAs were (●) S. purpuratus, (○) S. franciscanus or (■) L. pictus. Reassociation was at 0.12 M PB or 0.41 M PB with 0.05% SDS and 0.001 M EDTA at 50°C. Weight mean driver fragment lengths were approximately 650 nucleotides. After incubation, the samples (approximately 100 μg of driver, 1000 counts of tracer) were diluted to 0.12 M PB, and the fraction of molecules containing duplex regions was determined by chromatography on 0.4 g of HAP at 50°C. Unbound material was eluted with 0.12 M PB. The temperature was raised to 96°C and the bound ^{32}P-DNA was eluted with 0.12 M PB. The curves shown are the least-squares solutions assuming single second-order kinetic components (Pearson et al., 1977). With the exception of clone CS2099, the S. purpuratus data derive in part from Klein et al. (1976). The rate constants (M^{-1} sec $^{-1}$) for each cloned repeat when driven by S. purpuratus, S. franciscanus or L. pictus DNA, respectively, are: CS2133A—1.5, 0.07; CS2101—0.83, 0.07, 0.002; CS2034—1.2, 0.19, 0.01; CS2007—0.49, 0.15; CS2137—0.37, 0.13; CS2109B—0.24, 0.06; CS2108—0.03, 0.04; CS2099—0.1, 0.08.

pictus driver DNA. In two cases (clones CS2101 and CS2034), the reactions are nearly complete and rate constants can be calculated. In these cases, the thermal stability was measured and a correction has been made for the effect of divergence (Bonner et al., 1973). There is large uncertainty in the determination of the number of copies in L. pictus DNA. Many copies may not be recognized in these measurements and the frequencies listed in Table 1 for L. pictus are lower limits.

Measurement of Repetition Frequencies with an Excess of Cloned DNA Fragments

The large interspecific ratio of the frequencies of repetition for certain families seems surprising, and we have therefore confirmed the reassociation kinetic results by a totally independent method. In this approach, the extent of duplex formation of the labeled cloned DNA fragment is measured as a function of the ratio of labeled DNA to total sea urchin DNA. As the excess of cloned repeat is increased, there is a concomitant increase in the fraction of the members of its repetitive family with which it forms duplexes. Since the cloned repeat has been strand-separated and therefore does not self-reassociate, the fraction in duplexes may be assayed with hydroxyapatite and the number of members of the family it represents may be calcu-

Table 2. Frequencies Determined by Reassociation Kinetics

Clone	Length[a]	Frequency[b] (S. purpuratus)	Frequency[b] (S. franciscanus)	Ratio	Corrected Ratio[c]	Frequency[b] (L. pictus)	Ratio	Corrected Ratio[c]
CS2007	1100	400	45	9	7	low	–	–
CS2034	560	1000	160	6	4	10	100	25
CS2090	220	140	95	1.5	1.5	low	–	–
CS2099	235	80	60	1.3	–	–	–	–
CS2101	320	700	55	13	13	2	350	192
CS2108	190	20	30	0.7	–	–	–	–
CS2109B	125	200	50	4	4	–	–	–
CS2133A	310	2100	102	21	–	–	–	–
CS2137	190	530	185	3	2	–	–	–

[a] Length of cloned fragment in nucleotides. Both the nomenclature used to identify recombinant clones and the methods used to determine their lengths are described by Klein et al. (1978)

[b] Reiteration frequency is calculated as $S = K_c/K_{sc}$, where K_c is the second-order rate constant for the reaction of the cloned ^{32}P-DNA fragment with sea urchin DNA, and K_{sc} is the second-order rate constant for the reaction of the ^{3}H-DNA single-copy tracer with the same sea urchin DNA driver. Second-order rate constants were obtained by a least-squares method (Pearson et al., 1977). Since factors such as the minor differences in single-copy DNA, driver DNA and cloned DNA fragment lengths, thermal stability of the reaction products, salt concentrations and other unknown variables affect the absolute rate of the reactions, the family size estimates are probably only accurate to within a factor of 2. Generally the fragment lengths of the cloned tracers are comparable to that of the driver. In most cases the driver was approximately 650 nucleotides (weight mean), and in a few cases slightly shorter driver preparations were used. No length corrections were made on the observed rate constants, except in one case where the cloned fragment was ≥1000 nucleotides (clone CS2007). The correction in this case was obtained as follows: $K_{cL} = K_c\ (L_o/L_c)$, where K_{cL} is the length corrected second-order rate constant for the reaction of the cloned fragment with the driver DNA, K_c is the observed rate constant for the reaction, L_c is the length of the cloned fragment and L_o is the length of the sea urchin DNA driver. This correction is justified by the observations of Chamberlin et al. (1978), who showed that the observed rate constant for the reaction of a long tracer with a short driver is proportional to the tracer/driver length ratio. Measurements of the repetition frequency of these cloned sequences in the S. purpuratus genome only were made simultaneously by Klein et al. (1978) and ourselves. All the data points were combined and the best least-squares solution was determined. In seven cases, the parallel determinations agreed within a factor of two. In two cases, clones CS2090 and CS2099, agreement was within a factor of three.

[c] For the corrected ratio, the repetition frequency of the cloned sequences in the S. franciscanus and L. pictus genomes was corrected for the effect of sequence divergence on reassociation rate using the equation $F_{sp}/F = 2^{\Delta T/10}$, which is equivalent to a 2 fold rate retardation for each 10°C for sequence divergence (Bonner et al., 1973), where ΔT is the difference in T_m for S. purpuratus driver compared to the driver of the other species.

lated. Figure 3 shows the fraction of cloned DNA bound to hydroxyapatite as a function of the ratio of the cloned DNA to sea urchin DNA of three species. Since such a wide range of ratios is involved, we present the data as a log plot (which is not a log Cot plot). Care has been taken to assure that as the ratio is varied, the Cot for the actual sequences involved in the duplex formation remains constant. The ratio of the number of sites present in the DNA of two species may be estimated by inspection. The curves drawn and the values in Table 3 result from least-squares solutions as described in Experimental Procedures. Table 3 lists the ratios of urchin DNA to cloned DNA at half reaction (column 4) and the frequencies derived from them, without correction, in column 5.

In a reaction of this sort, there is a probability of breakage of the genomic DNA sequences. The resulting sequence fragments may each bind a whole cloned repetitive sequence. Excessive breakage could therefore multiply the number of apparent sites far beyond the actual number in the genome. Breakage should not affect the ratio of the frequencies for two different DNA samples if the length distributions are similar. The ratios of repetitive family sizes in S. purpuratus to those in S. franciscanus determined by saturation (Table 3) agree quite well with those determined by kinetics, as shown in Table 2.

The apparent increase in frequency caused by breakage is expected to be a function of the length of the sequence, and the data show such an effect. For a particular length distribution of the genomic DNA, there is an average length of sequence capable of binding one cloned DNA repeat which we term the "effective binding length." For an estimate, we assume that the length of each of the cloned repeats is a measure of the size of their respective family members or binding sites in the genomic DNA. Thus the effective binding length is calculated by dividing the cloned DNA length by the ratio of the uncorrected frequency to the kinetically determined frequency from Table 2. The result shows about a 2 fold variation with an average of 195 nucleotides. Column 6 shows the length-corrected frequency using 195 nucleotides as the effective binding length. Most of the corrected frequencies in column 6 are very close to the

Cell
654

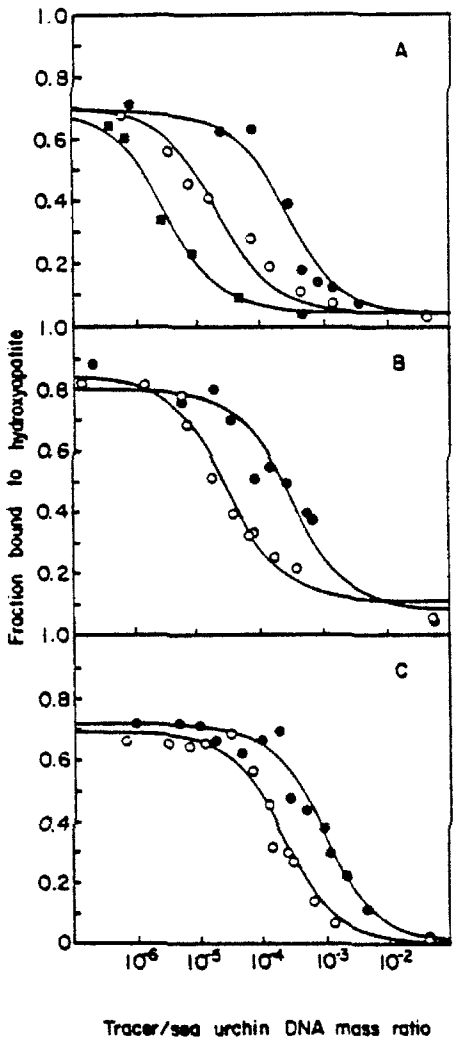


Figure 3. Genomic Repetition Frequency Estimation by the Saturation Method

Strand-separated tracer and genomic DNA from the three sea urchin species were incubated to 50 times the kinetically determined Cot_1 for each tracer sequence at 50°C in 0.41 M PB, and passed over HAP in 0.12 M PB at 50°C. The fraction of the ^{32}P-DNA bound to HAP is plotted against the log of the ratio of the ^{32}P-DNA to the genomic DNA. Tracers were the repeat fragments of (A) CS2101, (B) CS2007 and (C) CS2034. Genomic DNAs were S. purpuratus (●), S. franciscanus (○) and L. pictus (■). The curves are ideal single-component saturation curves fitted by a least-squares computer program as described in Experimental Procedures. In the case of clone CS2101, the extent of reaction at high and low ^{32}P-DNA to genomic DNA ratios was averaged for the three driver DNAs.

kinetically determined frequencies listed in Column 7. This close agreement supports the concept that the larger frequencies observed (before correction) *with the probe excess method are due to breakage* of the repetitive DNA sites in the driver DNA. Even without this length correction, the interspecific ratios of the frequencies observed by saturation agree with the kinetically determined interspecific ratios. These measurements therefore support the conclusion from the kinetic measurements that large changes in repetition frequency have occurred for some of the families represented by the cloned repetitive sequences.

Interspecies Reassociation Kinetics of Repetitive DNA Fractions

Since cloned repeats of S. Franciscanus and L. pictus are not yet available, we cannot make reciprocal cross-species measurements for individual families. Repetitive fractions were therefore prepared from each of three sea urchin species, and reassociation kinetics were determined for these tracers driven by an excess of DNA from each species, as shown in Figure 4. These tracers have an average frequency of repetition of approximately 1000–2000 copies in the species from which they were derived. The rate of reassociation of S. franciscanus repetitive tracer with S. franciscanus DNA and S. purpuratus repetitive tracer with S. purpuratus DNA is 2-3 times hgher than the corresponding interspecies reactions, as shown in Figures 4A and 4B. After correction for the effect of divergence on rate, we estimate that the frequency ratio is about a factor of *two to three. Even more striking is the comparison* of Figure 4A with Figure 4C for S. purpuratus and L. pictus tracers with S. purpuratus and L. pictus driver DNAs. These reciprocal interfamilial reactions show similar frequency contrasts. In neither case do the reactions go to completion for these distantly related sea urchins. Some of the families have changed more rapidly in frequency than others, and the average change is considerably more than a factor of 100. In the case of L. pictus repetitive tracer reacting with S. purpuratus driver DNA (right-hand curve, Figure 4C), measurements were made at a very high ratio of driver to tracer (400,000/1) so that if even one recognizable copy were present in the S. purpuratus DNA of each of *the families the reaction would go to completion.* This measurement suggests that some of the families that are now high frequency in L. pictus or S. purpuratus have changed so much in either frequency or sequence (presumably both) in the last 150–200 million years that homologous sequences are either unrecognizable or absent from the DNA of the other species.

Table 3. Frequency Estimates by Saturation with Excess Cloned Repetitive DNA

Clone and Length (Nucleotides)	Species	Q[a]	Apparent Frequency[b]	Ratio	Length Corrected Frequency[c]	Frequency by Kinetics[d]
CS2101	S. purpuratus	4,000	1250	1	761	986
L = 320	S. franciscanus	39,000	128	10	78	82
	L. pictus	444,000	11	110	7	5
CS2034	S. purpuratus	1,050	2721	1	971	1080
L = 560	S. franciscanus	4,630	617	4.4	220	126
CS2007	S. purpuratus	316	4600	1	836	465
L = 1100	S. franciscanus	3,760	393	12	71	77

[a] Ratio of sea urchin DNA to cloned DNA tracer at half-reaction using least-squares solution.
[b] Calculation using frequency = 2G/QL, where G is the genome size (8×10^8 nucleotides), L is the cloned repetitive sequence fragment length and Q is from column 3.
[c] Empirical length corrected frequency obtained by dividing column 4 by (L/195).
[d] From Table 2, corrected for the effect of divergence on rate as described in Table 2, footnote c.

Figure 5 shows measurements for a labeled repetitive fraction of S. purpuratus DNA that contains a more representative set of repetition frequencies. These data show that frequency differences occur in families with a wide range of frequency of repetition. All families do not show differences, however. The small sample of individual cloned families described in the last section included examples with differences in frequency between the genomes of S. franciscanus and S. purpuratus and several that showed no difference. Based on this small sample, we presume that Figures 4 and 5 represent an average of families with a range of different frequency contrasts between species. For the 150–200 million years (since the last common ancestor of L. pictus and S. purpuratus), few if any repeat families have survived without large frequency changes. At this great a distance, it appears that there has been substantially more change in the families that are now of high frequency in S. purpuratus than in the intermediate or low frequency families.

Discussion

The main conclusion we draw is that striking and rapid changes have occurred in the size of many individual repetitive sequence families during recent sea urchin evolution. Major support for this conclusion comes from the large contrasts in rates of reassociation observed when individual repetitive sequences and sequence fractions from one genome are reacted with the DNA of a related species. The data in Table 2 and Figure 2 show that many cloned repetitive sequences react more slowly with DNA from S. franciscanus than with DNA from S. purpuratus. Even larger contrasts in the rate of reassociation are observed in reactions with L. pictus DNA. The T_m of the duplexes with S. franciscanus DNA is lower in most cases than it is for duplexes with S. purpuratus DNA. The resulting small rate reduction, however, generally accounts for only a small fraction of the observed difference in reassociation rate. In some cases (for example, CS2109B), the thermal stability of the duplexes causes a larger uncertainty. Many copies of this sequence may exist with very low thermal stability and we cannot determine whether there is a frequency difference. The same would be true of interspecific measurements made with any of the highly divergent (class III) repeats studied by Klein et al. (1978). Such uncertainties do not apply in the case of the clone CS2101 repeat or the other relatively nondivergent (classes I and II) repetitive sequence families represented by the cloned tracers (see Table 2).

Table 4 shows no correlation between sequence divergence of the cloned repeat families and the extent of evolutionary change in family size. The lack of correlation indicates that sequence divergence is not responsible for the observed frequency differences either for technical reasons or as a result of lack of sequence recognition. The measurements made by Harpold and Craig (1977) of sea urchin interspecies repetitive DNA reassociation are consistent with the data reported here, although they interpret their data to indicate little or no frequency difference.

The number of sites in the genome that will form stable duplexes with the cloned repeat tracers was also estimated from the extent of reaction with excess of cloned DNA. The quantitative agreement between this method and the kinetic method suggests that both give reasonably accurate estimates of the number of copies of repeated DNA sequences in these genomes. The uncertainty is usually less than a factor of two. We conclude that the repetition frequencies of certain repeat families differ greatly among S. franciscanus, L. pictus and

Table 4. Lack of Correlation of Sequence and Frequency Change

		Melting Temperature Difference[a]	
		<1°C	>1°C
Frequency Ratio[b]	≤2	CS2090 (−1°C, 1.8)[c]	CS2137 (6°C, 2)
			CS2007 (5.5°C, 7)
	>2	CS2101 (0.8°C, 13)	CS2034 (5.5°C, 4)

[a] The difference in T_m from Table 1 for the cloned DNA reassociated with S. purpuratus DNA compared to that with S. franciscanus DNA. The five cases are those for which the T_m is far enough from the criterion of precision for the T_m difference to be considered reliable.
[b] The ratio of the number of members of the family with which the cloned DNA reassociates in S. purpuratus DNA compared with S. franciscanus DNA based upon kinetics of reassociation from Table 2.
[c] The measured T_m difference and the frequency ratio are listed individually in the parentheses.

pears that 95–99% of the repetitive sequences in each of these genomes either have few or no recognizable members in the other genome. Thus the dominant repetitive sequence families in one genome typically have a higher frequency than in related species genomes. It appears that copies of repeated sequences are being added to sea urchin genomes at a rate sufficient for the added species-specific copies to dominate each genome in the time since the last common ancestor of S. franciscanus and S. purpuratus.

Families of repeated sequences are apparently recognizable over long evolutionary periods although most of their members may be new. A majority of the families in S. franciscanus and S. purpuratus are related to families with very few copies in L. pictus. Thus it appears that the dominant families in modern genomes are principally the result of growth of older repetitive families rather than the result of initiation of new families. There must be termination or slowing down of the growth of many families. Otherwise those that grew in the ancestral species would simply continue to grow and the same families would dominate the genome of each of the modern species.

Frequency Changes of Interspersed Sequences

The majority of repeated sequences are probably interspersed with single-copy sequences (Graham et al., 1974) and the majority of repeated sequences appear to undergo massive changes in frequency during evolution. Thus interspersed repeated sequences must participate in large-scale changes in frequency. This leads us to consider as a major evolutionary process the wholesale removal and insertion of repetitive sequences from a milieu of single-copy sequences. We do not yet know what fraction of the members of the families represented by our set of clones are actually interspersed. There is strong evidence that most family members represented by clone CS2101 are interspersed (W. R. Pearson, unpublished data; G. P. Moore, unpublished data).

The most convincing evidence for the frequency change of interspersed sequences comes from the frequency comparisons of repeated DNA between L. pictus and S. purpuratus. Almost all repeats in S. purpuratus and L. pictus genomes are the product of events of multiplication that have occurred since the last common ancestor 150–200 million years ago. These two species have comparable repetitive frequency distribution and comparable repetitive sequence divergence. They presumably also have comparable patterns of interspersion. The implication is that repetitive sequences have been added to both genomes in many locations interspersed with single-copy DNA. An insertion or sequence rearrangement process appears to be active on an evolutionary time scale.

Generality of Frequency Changes in Eucaryotic Evolution

There is a limited amount of evidence that can be used to assess the generality of occurrence of frequency changes in the evolution of repeated DNA. Measurements have been made in two insects (P. Dunsmuir, P. Bingham and M. Meselson, personal communication), two amphibians (Galau et al., 1976) and the three sea urchins of this work. In each of these comparisons, frequency contrasts were observed even for sibling species (D. simulans and D. melanogaster). An observation that is probably due to frequency change has been made in rodents. The mouse satellite is observed in three closely related species (Sutton and McCallum, 1972; Rice and Strauss, 1973) in upwards of a million fairly precise copies which reassociate to form strand pairs whose T_m is within 5°C of that of perfect duplex. Large interspecies reassociation is observed and the interspecies duplexes melt sharply at about 25°C below perfectly complementary DNA. The best explanation is that a divergent family of repeated sequence was present in the common ancestor and different members of this family were greatly multiplied in each of the species.

Many measurements of interspecies repeated sequence reassociation have been made, and in every case at some Cot and criterion of precision (not always known), a smaller interspecific reaction is obtained. As the phylogenetic distance between the species is increased, the extent of reaction is reduced. Gillespie (1977) has measured interspe-

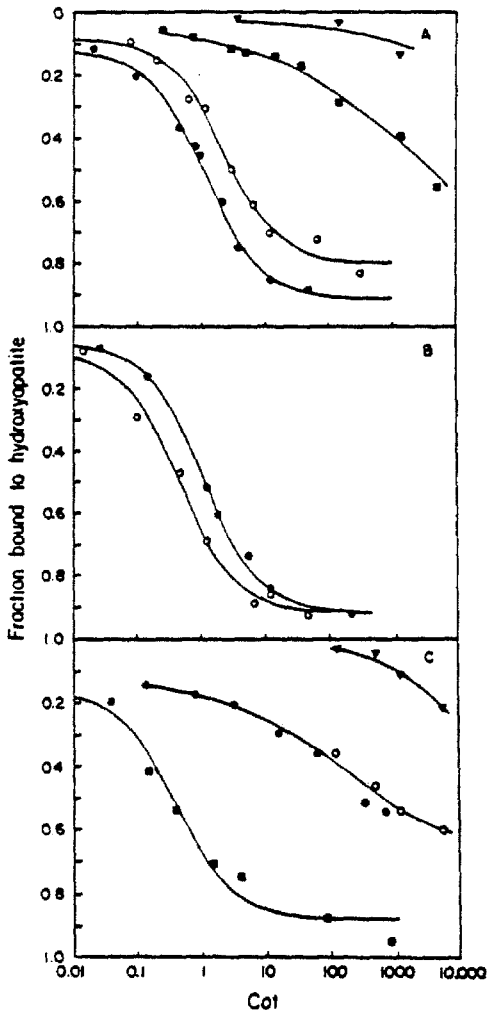


Figure 4. Reassociation of Highly Repetitive Tracer with Homologous and Heterologous Driver DNA

Weight mean driver DNA fragment lengths were approximately 850 nucleotides and the driver/tracer DNA mass ratios were approximately 2×10^4, except where otherwise indicated. Where the termination and rate constants are given, the curves are least-squares solutions assuming a single component. Preparation of repetitive tracers and reassociation and HAP conditions are described in Experimental Procedures.

(A) S. purpuratus tracer reassociated with (●) S. purpuratus, (○) S. franciscanus, (■) L. pictus or (▼) self-reacted. The S. purpuratus and S. franciscanus driven reactions terminate at 88.7 and 79.5%, with rate constants of 8.48×10^{-1} and 4.3×10^{-1} M^{-1} sec^{-1}, respectively. (B) S. franciscanus tracer reassociated with (○) S. franciscanus or (●) S. purpuratus driver DNA. The reactions terminate at 91.5 and 91.0% with rate constants of 2.2 and 0.93 M^{-1} sec^{-1}, respectively. (C) L. pictus tracer reassociated with (■) L. pictus or (●, ○) S. purpuratus DNA, or (▼) self-reacted.

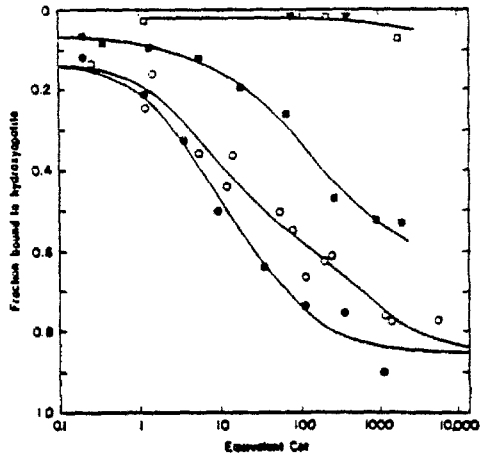


Figure 5. Reassociation of Middle Repetitive S. purpuratus Tracer with Homologous and Heterologous Driver DNAs

S. purpuratus middle repetitive tracer was reacted with (●) S. purpuratus, (○) S. franciscanus, (■) L. pictus, (□) calf DNA or (▼) self-reacted. Weight mean driver DNA fragment length was approximately 850 nucleotides and the driver/tracer DNA weight ratio was approximately 2×10^4. Preparation of the repetitive tracer and reassociation and HAP conditions are described in Experimental Procedures. The least-squares solution for the S. purpuratus driven reaction terminated at 85.0% and had a 30.1% component with a rate constant of 0.022 M^{-1} sec^{-1}, and a 41.8% component with a rate constant of 0.21 M^{-1} sec^{-1}. The S. franciscanus driven reaction fixed with the same termination point and component sizes had rate constants of 0.0017 and 0.153 M^{-1} sec^{-1}.

S. purpuratus. The measurements with genomic repetitive DNA show that many families differ in frequency among these three sea urchin species.

Evolution of Repeated Sequence Frequency in Sea Urchins

A striking feature of these observations is that five of the nine cloned repetitive sequences show considerably higher frequencies in the genome of S. purpuratus (from which they were derived) than they do in the genome of S. franciscanus. The measurements with genomic repetitive sequence fractions from all three species shown in Figures 4 and 5 are consistent with this observation. Both high and low frequency repetitive families have fewer members in the S. franciscanus genome (Figures 4A and 5). High frequency repetitive families in S. franciscanus also have fewer members in the S. purpuratus genome (Figure 4B). A similar reciprocal relationship with a very large frequency ratio is shown by the comparison of S. purpuratus and L. pictus repetitive tracers (Figure 4C). It ap-

The L. pictus driven reaction terminates at 88.3% with a rate constant of 2.6 M^{-1} sec^{-1}. For some S. purpuratus points (○), the driver/tracer DNA mass ratio was increased to 4×10^4.

cies repetitive DNA cross-reactions among primates, and was concluded that families of repeated sequences were "amplified" by a large factor. Dover (1977) has reviewed evolutionary genomic variations. In recent studies, Mizuno and MacGregor (1974) and Flavell and Smith (1976) have concluded that frequency changes or saltatory replications have occurred in amphibians and monocots. If the low extent of interspecies repetitive DNA reaction is indeed attributable to changes in repeat family size, we would conclude that it is a nearly universal process in animal evolution.

Rate of Addition and Possibility of Deletion

For quantitative comparison of rates, we define the fractional rate of addition as the fraction of the total repeated sequences added per year to the genome, not as a net rate of growth but as new copies of repetitive sequences. Sequences complementary to the dominant repetitive sequences in either S. franciscanus or S. purpuratus occur in the other with a frequency that is less by a factor of two to three. This result implies that at least one half of the repeated sequences have been added to the genome of each species in the last 15–20 million years. Thus a minimum estimate of the rate is $0.5/2 \times 10^7 = 2.5 \times 10^{-8}$ per year. There are approximately 5×10^8 interspersed repetitive sequence elements in the S. purpuratus genome, each an average of 300 nucleotides long. The fractional rate of addition implies that one such sequence is being added to the genome about every century on the average. This average includes many different families which may have quite different rates and could be affected by a few large changes in individual families termed saltatory replications (Britten and Kohne, 1967).

Alternatively, this change can be described as the addition of about 5 nucleotides of repetitive sequence per year. It is possible that the sea urchin genome is now growing at this rate, but such a rate cannot have been maintained since the lower ordovician (500 million years ago) when the sea urchin first appeared. If this were true, the sea urchin genome would have grown to more than its present size of 8×10^8 nucleotide pairs. This argument suggests a long-term balance of the processes of addition and deletion, but the question cannot yet be resolved.

Selection and Function of Repeated Sequences

Transcripts that hybridize with the nine cloned repeats examined in this paper are present in mature oocyte RNA (Costantini et al., 1978) and in nuclear RNA in early development (Scheller et al., 1978). High concentrations of transcripts from different families were shown to occur at different stages of development. Thus transcription of members of these repeated sequence families is under stage-specific control. The argument is made in these papers that the patterns of transcription show that most families of repetitive sequences carry out significant functions. This concept is supported by the observations of sequence divergence listed in Table 4. Two of the clones show melting temperatures with S. franciscanus DNA that are within 1°C of those with S. purpuratus DNA. This contrasts with the 10°C single-copy DNA sequence divergence between these two species (T. J. Hall and R. J. Britten, unpublished data), and suggests strong selective pressure dependent upon the function of many members of these families. The species which we have studied display similar repetitive sequence frequency distributions, but the corresponding frequency domains must be occupied by different repetitive sequence families. These families could have functions that depend upon their family size. How can they change so rapidly and thus appear to be independent of selection pressure on family size? Conceivably the repetitive sequence families have little to do with the functional organization of the genome and are free to change without restriction. If repeat family members do have important roles, however, it is possible that different families may carry out the same functions in related genomes. If the insertion and removal of sequences from their functional locations could occur, then whole families might replace other families in the genome. One family could be reduced in size while the other increased without interfering severely with viability. Nearly all the copies of almost all the families of repetitive sequences are different in the genomes of L. pictus and S. purpuratus. Yet these two animals are very similar in morphology, behavior and development. While many individual molecules and nucleotide sequences have changed, the principal functions of genomic organization and the regulation of gene expression in development remain essentially unaltered. We therefore propose the following hypothesis: during evolution most families of repeated sequences are replaced with others which perform slightly modified functions leading to phenotypic variation. Suppose batteries of genes were affected together—for example, as proposed by Britten and Davidson (1971) in their speculations regarding the origins of evolutionary novelty. Evolutionary changes in genomic sequence organization might exhibit coordinated characteristics and could provide a significant source of morphological change in evolution.

Experimental Procedures

DNA Extraction

Specimens of S. purpuratus, S. franciscanus and L. pictus were collected locally and dissected to obtain sperm. DNA was ex-

tracted from frozen sperm as described by Angerer et al. (1976). DNA preparations were made from combined sperm of several animals to reduce possible effects of genetic polymorphism (Britten, Cetta and Davidson, 1978). All DNA preparations had hyperchromicity of 28% or greater (A_{260} at 98°C-A_{260} at 60°C/A_{260} at 98°C × 100) in 0.12 PB [Na phosphate (pH 6.8)].

Preparation and Sizing of DNA Fragments

DNA fragments of the indicated sizes were produced by homogenization in a Virtis 60K homogenizer as previously described (Britten, Graham and Neufeld, 1974). The single-strand fragment length of DNA was determined by centrifugation through isokinetic alkaline sucrose gradients (Noll, 1967): V_{top} = 9.84 ml, C_{top} = 43% (w/v), C_{bottom} = 16% (w/v) in 0.1 N NaOH. Gradients were centrifuged at 41,000 rpm for 20–24 hr at 20°C in a Beckman SW41 rotor. The weight average fragment length was determined in duplicate measurements with respect to internal markers previously sized by electron microscopy using the equations of Studier (1965).

DNA Reassociation

DNA reassociation was carried out in 0.12 or 0.41 M PB with 0.05% SDS and 0.001 M ethylenediaminetetraacetate at 50°C in heat-sealed capillary tubes or glass ampules. All Cot values quoted in the text are equivalent Cot (Britten et al., 1974). Cot in 0.41 M PB was converted to equivalent Cot by multiplying by 5. DNA used as driver of radioactively labeled tracers was tested for driver reactability by self-reaction and optical monitoring of the resultant duplexes at 260 nm in a Unicam SP1800 spectrophotometer.

Hydroxyapatite (HAP) Chromatography

The fraction of DNA fragments in molecules containing duplexes after reassociation was determined by chromatography on HAP (BioRad, DNA grade lot No. 16399). Samples were diluted to 0.12 M PB and applied to water-jacketed HAP columns at 50°C. Unbound material was removed by washing with at least 5 bed volumes of 0.12 M PB. Duplex material was subsequently eluted by raising the temperature of the column to 98°C or by washing with at least 5 bed volumes of 0.41 M PB. DNA loads were <250 μg/cc (0.4 g) of packed HAP. When DNA loads were <20 μg (usually for kinetic fractionation of tracers), the column was presequilibrated with 20 μg of calf thymus DNA. HAP was also used to concentrate DNA solutions by dilution of the sample to 0.03 M PB, application to the column and finally elution with 0.41 M PB in a small volume.

The stability of reassociated DNA duplexes was determined by thermal chromatography from HAP. Samples of reassociated DNA were applied to columns at 50°C in 0.12 M PB. The temperature was then increased in 4°C increments, and the single-stranded DNA fragments were eluted from the column at each temperature. ^{3}H-labeled 450 nucleotide long native DNA fragments were mixed with the sample as an internal standard.

Saturation Estimate of Frequency

When the ratio of cloned repetitive DNA to nuclear DNA is varied at Cot sufficient to complete the reassociation, the maximum amount of strand-separated probe that forms duplexes is a measure of the number of copies for the homologous family. The saturation can be expressed as fraction of probe bound = 1/(1 + conc. probe/conc. sites), which has the same form as an ideal second-order reaction. The number of sites was evaluated by a least-squares solution using a computer program (Pearson, Davidson and Britten, 1977; Wallace, Dube and Bonner, 1977). The method of graphing against the log of the ratio of probe to nuclear DNA shown in Figure 3 simply gives a convenient visual indication of the ratios of frequencies and the certainty with which we can conclude that a pair of frequencies are actually distinct.

Counting Methods

Radioactive samples were counted in mixtures of 0.12 M PB and "Aquasol" (New England Nuclear) or "Handifluor" (Mallinckrodt) in the proportion 1:2.5. Alkaline samples were neutralized prior to counting.

Preparation and Labeling of Tracers

For the preparation of radioactively labeled S. purpuratus DNA, embryos were raised to the 4 cell stage as described by Hinegardner (1967). 5 mCi of ^{3}H-thymidine were then added to 1.5 × 10^7 embryos in 150 ml of seawater containing penicillin and streptomycin, and the embryos were incubated overnight to approximately hatching blastula stage. Embryos were frozen with dry ice and ground with powdered dry ice in a Waring blender, and DNA was isolated as described above. The specific activity of the preparation used was 110,000 cpm/μg and the weight mean single-strand length was 725 nucleotides, as determined by alkaline sucrose velocity sedimentation.

Radioactively labeled S. franciscanus and L. pictus DNA was prepared by the nick translation method using 50 units of E. coli polymerase I (Boehringer Mannheim, Grade 1) and 50 μg of DNA in a final reaction volume of 1 ml. The reaction mix was 50 mM Tris (pH 8.8), 5 mM MgCl, 2 × 10^{-4} M ATP, 2 × 10^{-4} M GTP, 2 × 10^{-4} M CTP, and contained 5 mCi ^{3}H-TTP (70 Ci/mmole; ICN). Incubation was at 13°C for 1 hr. The reaction was stopped by extraction with phenol-Sevag solution (1:1 ratio of phenol and Sevag which is 24:1 chloroform–isoamyl alcohol) and then by extraction with Sevag alone. The reaction mixture was adjusted to 0.12 M PB and passed over HAP, where approximately 5% of the counts bound. The specific activity of the bound tracer was 130,000 cpm/μg and 260,000 cpm/μg, and the single-strand length was 675 nucleotides and 750 nucleotides in the case of S. franciscanus and L. pictus, respectively.

The S. purpuratus tracer whose reaction is shown in Figures 4A and 5 was prepared from the in vivo labeled DNA described above as follows. For the tracer of Figure 5, 6 × 10^8 counts of starting material were incubated to Cot 192; the bound fraction on HAP (BF) was 67%. This material was incubated to Cot 25 and the BF (23% of initial input) was incubated to Cot 2.5. The unbound fraction on HAP (UBF) (9.3% or 5.7 × 10^8 cpm) constituted the tracer used. For the tracer of Figure 4A, 6 × 10^8 cpm of starting material were incubated to Cot 16 and the BF (15%) was incubated to Cot 1.2. The BF (8%) was incubated to Cot 0.01 and the UBF (6.5%) was incubated to Cot 0.028. The UBF (4.5%) was incubated to Cot 1 and the BF (which was 4% of the input counts or 2.4 × 10^8 cpm) constituted the tracer used. To prepare the S. franciscanus tracer of Figure 4B, the starting material was 4 × 10^8 cpm of the nick-translated S. franciscanus DNA described above. The DNA was incubated to Cot 0.015 and the UBF (62%) was incubated to Cot 1.6. The BF (7.5%) was incubated to Cot 1 and the BF (4.7%) was incubated to Cot 0.006. The UBF (3.5% of the input or 1.4 × 10^8 cpm) was concentrated and used as tracer in the reactions shown. To prepare the L. pictus tracer of Figure 4C, the starting material was 10^9 counts of the nick-translated L. pictus DNA described above. This tracer was incubated to Cot 0.015 and the UBF (59%) was incubated to Cot 1.95. The BF (4%) was incubated to Cot 3.2 and the BF (2.4%) was incubated to Cot 0.001. The UBF (2.3% of input) was concentrated and used as tracer.

Preparation of Cloned Repetitive Sequences

Individual S. purpuratus repetitive sequences were cloned in the bacterial plasmid RSF2124 as described by Scheller et al. (1977a). Labeling of the cloned repeats and their separation into single strands to avoid tracer self-reaction was carried out as described by Scheller et al. (1978). The characteristics of the cloned repeats used in this study are described by Klein et al. (1978).

Acknowledgments

The knowledge before publication of the measurements of P. Dunsmuir, P. Bingham and M. Meselson with Drosophila species was helpful in recognizing the generality and significance of frequency change. Unpublished measurements made by Dr. Robert C. Angerer of the conservation of repetitive DNA sequences in these sea urchins were also valuable. We acknowledge the excellent technical assistance of Linda Vock and Richard Ronan. This work has been supported by a grant from the National Institute of General Medical Sciences. G. P. M. is the recipient of a fellowship from the American Cancer Society. R. H. S. is a predoctoral fellow on an NIH training grant.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received May 1, 1978; revised June 24, 1978

References

Angerer, R. C., Davidson, E. H. and Britten, R. J. (1978). Single copy DNA and structural gene sequence relationships among four sea urchin species. Chromosoma *56*, 213–226.

Bonner, T. I., Brenner, D. J., Neufeld, B. R. and Britten, R. J. (1973). Reduction in the rate of DNA reassociation by sequence divergence. J. Mol. Biol. *81*, 123–135.

Britten, R. J. and Waring, M. J. (1965). Biophysics. Carnegie Inst. Washington Yearbook *64*, 331–333.

Britten, R. J. and Kohne, D. E. (1967). Nucleotide sequence repetition in DNA. Carnegie Inst. Washington Yearbook *65*, 78–106.

Britten, R. J. and Kohne, D. E. (1968). Repeated sequence in DNA. Science *161*, 529–540.

Britten, R. J. and Davidson, E. H. (1971). Repetitive and nonrepetitive DNA sequences and a speculation on the origins of evolutionary novelty. Quart. Rev. Biol. *46*, 111–138.

Britten, R. J., Cetta, A. and Davidson, E. H. (1978). The single-copy DNA sequence polymorphism of the sea urchin S. purpuratus DNA. Cell, in press.

Britten, R. J., Graham, D. E. and Neufeld, B. R. (1974). Analysis of repeating DNA sequences by reassociation. In Methods in Enzymology, *29E*, L. Grossman and K. Moldave, eds. (New York: Academic Press), pp. 363–406.

Chamberlin, M. E., Galau, G. A., Britten, R. J. and Davidson, E. H. (1978). Studies on nucleic acid reassociation kinetics: V. Effects of disparity in tracer and driver fragment lengths. Nucl. Acids Res. *5*, 2073–2094.

Costantini, F. D., Scheller, R. H., Britten, R. J. and Davidson, E. H. (1978). Repetitive sequence transcripts in the mature sea urchin oocyte. Cell *15*, 173–187.

Dover, G. (1977). Variation in genome organization in related species: an annotation. Chromosomes Today *6*, 105–115.

Durham, J. W. (1966). Echinoids, classification. In Treatise on Invertebrate Paleontology, Part U, Echinodermata 3, *1*, R. C. Moore, ed. (New York: The Geological Society of America and The University of Kansas Press), pp. 270–295.

Flavell, R. B. and Smith, D. B. (1976). Nucleotide sequence organization in the wheat genome. Heredity *37*, 231–252.

Galau, G. A., Chamberlin, M. E., Hough, B. R., Britten, R. J. and Davidson, E. H. (1976). Evolution of repetitive and nonrepetitive DNA. In Molecular Evolution, F. J. Ayala, ed. (Sunderland, Massachusetts: Sinauer Association), pp. 200–224.

Gillespie, D. (1977). Newly evolved repeated DNA sequences in primates. Science *196*, 889–891.

Graham, D. E., Neufeld, B. R., Davidson, E. H. and Britten, R. J. (1974). Interspersion of repetitive and nonrepetitive DNA sequences in the sea urchin genome. Cell *1*, 127–137.

Harpold, M. M. and Craig, S. P. (1977). The evolution of repetitive DNA sequences in sea urchins. Nucl. Acids Res. *4*, 4425–4438.

Harpold, M. M. and Craig, S. P. (1978). The evolution of nonrepetitive DNA in sea urchin. Differentiation *10*, 7–11.

Hinegardner, R. (1967). Echinoderms. In Methods in Developmental Biology, F. H. Wilt and N. K. Wessels, eds. (New York: Thomas Y. Crowell Co.), pp. 139–155.

Klein, W. H., Thomas, T. L., Lai, C., Scheller, R. H., Britten, R. J. and Davidson, E. H. (1978). Characteristics of individual repetitive sequence families in the sea urchin genome studied with cloned repeats. Cell *14*, 889–900.

Mizuno, S. and MacGregor, H. C. (1974). Chromosomes, DNA sequences, and evolution in salamanders of the genus Plethodon. Chromosoma *48*, 239–296.

Noll, H. (1967). Characterization of macromolecules by constant velocity sedimentation. Nature *215*, 360–363.

Pearson, W. R., Davidson, E. H. and Britten, R. J. (1977). A program for least squares analysis of reassociation and hybridization data. Nucl. Acids Res. *4*, 1727–1737.

Rice, N. R. and Strauss, N. A. (1973). Relatedness of mouse satellite deoxyribonucleic acid to deoxyribonucleic acid of various Mus species. Proc. Nat. Acad. Sci. USA *70*, 3546–3550.

Scheller, R. H., Thomas, T. L., Lee, A. S., Klein, W. J., Niles, W. D., Britten, R. J. and Davidson, E. H. (1977a). Clones of individual repetitive sequences from sea urchin DNA constructed with synthetic EcoRI sites. Science *196*, 197–200.

Scheller, R. H., Dickerson, R. E., Boyer, H. W., Riggs, A. D. and Itakura, K. (1977b). Chemical synthesis of restriction enzyme recognition sites useful for cloning. Science *196*, 177–180.

Scheller, R. H., Costantini, F. D., Kozlowski, M. R., Britten, R. J. and Davidson, E. H. (1978). Specific representation of cloned repetitive DNA sequences in sea urchin RNAs. Cell *15*, 189–203.

So, M., Gill, R. and Falkow, S. (1975). The generation of a ColE1-Apr cloning vehicle which allows detection of inserted DNA. Mol. Gen. Genet. *142*, 239–249.

Sutton, W. D. and McCallum, M. (1972). Related satellite DNA's in the genus Mus. J. Mol. Biol. *71*, 633–656.

Studier, F. W. (1965). Sedimentation studies of the size and shape of DNA. J. Mol. Biol. *11*, 373–390.

Wallace, B. R., Dube, S. K. and Bonner, J. (1977). Localization of the globin gene in the template active fraction of chromatin of Friend leukemia cells. Science *198*, 1166–1168.

CHAPTER 8

Repetitive Sequences of the Sea Urchin Genome

I. Distribution of Members of Specific Repetitive Families

D. M. ANDERSON, R. H. SCHELLER, J. W. POSAKONY, L. B. MCALLISTER,

S. W. TRABERT, R. J. BRITTEN, AND E. H. DAVIDSON

Division of Biology

California Institute of Technology

Pasadena, California 91125, U.S.A.

Running Title:

† Abbreviations used: nt, nucleotide; kb, kilobases

Three repetitive sequence families from the sea urchin genome were studied, each defined by homology with a specific cloned probe one to a few hundred nt† long. Recombinant λ-sea urchin DNA libraries were screened with these probes, and individual recombinants were selected which include genomic members of these families. Restriction mapping, gel blot, and kinetic analyses were carried out to determine the organization of each repeat family. Sequence elements belonging to the first of the three repeat families were found to be embedded in longer repeat sequences. These repeat sequences frequently occur in small clusters. Members of the second repeat family are also found in a long repetitive sequence environment, but these repeats usually occur singly in any given region of the DNA. The sequences of the third repeat are only 200 to 300 nt long, and are generally terminated by single copy DNA, though a few examples were found associated with other repeats. These three repeat sequence families constitute networks of homologous sequence elements which relate distant regions of the DNA. Such networks are probably a global feature of the organization of animal genomes.

1. Introduction

Renaturation kinetics indicate that animal DNAs contain a large variety of diverse repetitive sequence families. A repeat family may be defined experimentally as that set of homologous sequences which reacts with a given cloned repetive sequence probe. Klein et al. (1978) identified a number of such families in previous studies on sea urchin DNA, and three of these have been chosen for the present experiments. Almost nothing is known about how the individual sequences belonging to repeat families are distributed with respect to each other. To approach this issue, we isolated from λ genome libraries a number of λ-sea urchin DNA recombinants which include genomic members of the three repeat sequence families and their flanking sequences. These recombinants were used to establish the extent to which members of the same family occur in clustered arrays, as opposed to being distributed singly throughout the genome. In addition, we determined the sequence environment characteristic of each family, i.e., whether its members are characteristically embedded in single copy sequence or in other repetitive sequences, or both. The particular families we have chosen for this study include a set of short repetitive sequences interspersed in a typical way with single copy DNA, and also examples of long repetitive sequences. Though the detailed pattern of organization of each family is unique, our measurements imply that the members of all three families occur in widely separated regions of the genome.

2. Materials and Methods

(a) Preparation of unlabeled sea urchin DNA

DNA was extracted from fresh S. purpuratus sperm, as described by Britten et al. (1974). Care was taken to avoid mechanical shearing of the DNA to ensure maximum double stranded length. Driver DNA used in renaturation reactions was prepared by forcing the DNA solution through a needle valve at

50,000 psi (Britten et al., 1974). The sheared driver DNA had a weight average length of 600 nt, measured by velocity sedimentation through alkaline sucrose gradients. Unsheared DNA had a length in excess of 100,000 nt pairs as measured by electron microscopy.

(b) Partial EcoRI digestion of sea urchin DNA

To prepare DNA fragments of a length suitable for cloning (15 to 20 kb†), unsheared (>100 kb) DNA was subjected to partial EcoRI digestion in 0.06 M Tris-HCl (pH 8.0), 6 mM $MgCl_2$, 0.05 M NaCl at 37°C. We would expect an EcoRI recognition site to occur about once every 4000 (4^6) nt (uncorrected for base composition). The EcoRI digestion conditions were adjusted to cleave an average of 1 site in 5 by varying the length of digestion and the ratio of enzyme to DNA. The partially cleaved DNA (200 µg) was fractionated on preparative 10 to 40% linear sucrose gradients (0.1 M NaCl, 10 mM Tris-HCl [pH 8.0], 10 mM EDTA) in a Beckman SW27 rotor. Gradients were centrifuged at 24,000 rpm for 20 h at 15°C, and 0.5 ml fractions were collected. Aliquots of the fractions were analyzed by electrophoresis on 0.5% agarose gels using EcoRI-digested Charon 4 DNA (Blattner et al., 1977) as a molecular weight standard. The fractions containing DNA fragments 15 to 20 kb long were pooled and precipitated in isopropanol.

(c) Preparation of Charon 4 vector

Charon 4 phage were grown essentially, as described by F. R. Blattner in the protocol provided with the Charon λ phages. Phage were purified, as described by Yamamoto et al. (1970). The phage DNA was extracted as follows: The purified phage were dialyzed from CsCl against 0.01 M Tris-HCl (pH 8.0), 0.1 M NaCl, 0.01 M EDTA. SDS was added to 0.1%, and the solution was heated for 10 min

at 60°C. The DNA was then extracted twice with phenol, once with chloroform-isoamyl alcohol (24:1), dialyzed against 0.1 M Tris-HCl (pH 8.0), 0.001 M EDTA, and ethanol precipitated. The Charon 4 vector arms were isolated from the mid-pieces and prepared for use, as described by Maniatis et al. (1978).

(d) λ-Recombinant genome libraries

The Charon 4 arms were ligated to the partially digested sea urchin DNA at a 1.5 molar excess of vector to sea urchin DNA fragments. The ligation was carried out in 66 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 1 mM ATP, 15 mM DTT, 200 μg/ml gelatin for 18 h at 15°C, at a total nucleic acid concentration of 200 μg/ml. In vitro packaging and plate lysate amplification were carried out essentially as described by Maniatis et al. (1978). Cloning efficiencies of 1 to 2 x 10^4 plaque forming units (pfu) per microgram DNA were routinely obtained. Two λ libraries constructed in this manner were utilized in this work, containing 1.4 x 10^5 and 2.5 x 10^5 different phage, respectively. The first of these libraries (denoted $Sp\lambda R_1 A$) was derived from DNA of the sperm of 5 males and the second (denoted $Sp\lambda R_1 B$) from sperm of a single male. A simple Poisson calculation indicates that the number of clones included in these libraries would contain 93% and 99% respectively of the sequences in the S. purpuratus genome, i.e., if there is no selection against given sea urchin DNA sequences, and if EcoRI sites are randomly distributed. Both of these propositions are probably untrue in detail. However, a direct complexity measurement in which DNA of the smaller library was used to drive a labeled single copy tracer showed that at least 90% of the genomic complexity was in fact included in the amplified library at a sequence concentration $\geq$1/3 the average library single copy DNA sequence. A third and larger library used for some experiments was constructed by ligating EcoRI linkers onto sea urchin DNA fragments which had been partially digested with HaeIII. This library

was built and characterized by M. Chamberlin and G. Moore of this laboratory, by the methods described by Maniatis et al. (1978). This library (denoted SpλH3C) contained 7.8×10^5 individual phage. The average insert length in these libraries was 15 kb, and the range of length was 10 to 18 kb.

(e) Preparation, labeling, and strand separation of cloned repetitive sequence fragments

Recombinant plasmids containing repetitive sequence elements inserted by blunt end ligation of EcoRI linkers were described earlier (Scheller et al., 1977). The vector was RSF2124 (So et al., 1975). Briefly, sea urchin DNA was renatured to C_0t 40 and digested with S1 nuclease, and the blunt ended repeat duplexes which survived were cloned as indicated. Superhelical DNA of these plasmids was isolated on isopycnic CsCl gradients (Scheller et al., 1977). The DNA was dialyzed into 5 mM Tris-HCl (pH 8.0) and stored frozen at -20°C. Plasmid DNA was cleaved with EcoRI to release the inserted repetitive sequence fragment and ethanol precipitated. The DNA was dissolved, then treated with bacterial alkaline phosphatase, end-labeled with ^{32}P by the polynucleotide kinase reaction (Maxam & Gilbert, 1977), and strand separated after alkali denaturation on neutral acrylamide gels, as described by Scheller et al. (1978). The DNA fragments of interest were localized by autoradiography and individual bands were excised with a razor blade. Gel slices were crushed with a glass rod in sterile plastic tubes. One ml of 0.12 M sodium phosphate buffer, 0.05% SDS and 10 μg of purified calf thymus DNA carrier were added to each sample, and the mixture was incubated overnight at 37°C. The acrylamide was removed by filtering through glass wool and rinsing with an additional 1 ml of phosphate buffer. The eluate contained from 80 to 95% of the radioactivity in the gel band. The sample was then incubated overnight at 60°C and passed over a 1 ml column of hydroxyapatite at 50°C.

The unbound fraction contained the strand-separated repetitive sequence. The final preparations were contaminated only 0 to 4 % with their complementary strands. Reactivity of these DNA preparations with excess total sea urchin DNA ranged from 55 to 98%. The nonreactive component(s) were of low molecular weight and most probably consisted of unincorporated ^{32}P-ATP.

(f) Library screening

The amplified sea urchin λ libraries were screened with ^{32}P-labeled cloned repeat tracers by a modification of the Benton & Davis (1977) procedure. One to 2 x 10^4 recombinant phage were plated on 4 x 10^8 bacterial cells on 15 cm agar plates. To prevent top agar from adhering to the nitrocellulose filter when it was lifted from the plate, 0.7% agarose rather than agar was used for plating. Phage were adsorbed to nitrocellulose filters (Schleicher and Schuell, 0.45 μ pore size) for about 10 min. The DNA was denatured and bound to the filters, as described by Benton & Davis (1977).

To hybridize the filters with a labeled probe, filters were preincubated for 1 h in 4X SET (1X SET = 0.15 M NaCl, 0.03 M Tris-HCl [pH 8.0], 2 mM EDTA) plus 5X Denhardt's solution (Denhardt, 1966) and 0.1% SDS. Subsequent hybridization were carried out in the same solution in sealed plastic bags. In general, ∿2 x 10^5 cpm tracer (∿ 10^7 cpm/μg) were added per filter. Incubations were for 18 h at 50°C, or appropriately raised temperatures for higher salt concentration to produce an equivalent criterion condition, unless otherwise noted. Following hybridization, filters were washed several times in 1X SET plus 0.5% SDS at the incubation temperature. The filters were blotted dry, mounted on cardboard and exposed to preflashed Kodak X45 X-ray film with Dupont Cronex NR Xtra Life Lightning Plus intensifying screens at -70°C for 1 to 7 days. Positive plaques from the region of a plate corresponding to a spot on the autoradiogram were

picked and suspended in 1 ml SM buffer (0.05 M Tris-HCl [pH 7.4], 0.05 M NaCl, 0.01 M $MgCl_2$). The phage were titered and rescreened at a density of 200 phage per 15 cm plate. Individual positive plaques were then selected, resuspended in 1 ml of SM buffer and amplified in 2 ml liquid cultures. These were prepared by adding 10 µl of late log phase bacteria and 100 µl of the resuspended plaque to 2 ml of broth. The cultures were shaken at 37°C until lysis was evident (about 18 h). Titers of these cultures were on the order of 1×10^{10} pfu/ml. The lysate was cleared of debris by centrifugation and stored at 4°C over a drop of chloroform.

(g) DNA renaturation

DNA renaturation was carried out in 0.12 M or 0.41 M phosphate buffer with 0.05% SDS in sealed capillary tubes. All C_0t values quoted in the text are equivalent C_0t (Britten et al., 1974). For example, C_0t in 0.41 M phosphate buffer was converted to equivalent C_0t by multiplying by 5. Renaturation kinetic analyses carried out in this work included an internal single copy DNA rate standard. Single copy ^{3}H-DNA was prepared and labeled by gap translation (Galau et al., 1976). The fraction of DNA fragments in molecules containing duplexes after reassociation was determined by binding to hydroxyapatite (DNA grade, BioRad lot 17653). Samples were diluted to 0.12 M phosphate buffer and applied to water-jacketed columns at 55°C. Unbound material was removed by washing with at least 5 bed volumes of 0.12 M phosphate buffer. The duplex fraction was subsequently eluted by raising the temperature of the column to 98°C. Less than 250 µg of DNA was loaded per cc of packed hydroxyapatite.

(h) Gel blots and restriction digests

Digestion with various restriction enzymes were carried out under the conditions suggested by the manufacturers. Transfer of DNA from agarose gels to nitrocellulose filters was as described by Southern (1975). Hybridization conditions

were as described above for library screening.

(i) Recombinant DNA safety

Experiments involving the cloning or propagation of plasmids or bacteriophage λ carrying eucaryotic DNA were performed in accordance with the NIH Guidelines for recombinant DNA research. The EK1 vectors used were plasmid RSF2124 (So et al., 1975) and λ derivative Charon 4 (Blattner et al., 1977).

3. Results

(a) Three repetitive sequence families: general characteristics from reactions of cloned probes with genomic DNA

The average properties of the repetitive sequence families chosen for these experiments are summarized in Table 1. Most of the data listed are quoted from other studies, and were obtained mainly by analyses of the heteroduplexes formed by reacting genomic DNA with plasmid clones bearing the three individual repeat sequences (Klein et al., 1978; Moore et al., 1980). The three repeat families described in Table 1 are designated according to the plasmid clones which define them, viz. clones CSp2034, CSp2108 and CSp2109.

Table 1 shows that the three repeat families differ from each other in several important characteristics. Most of the individual sequence elements belonging to the 2034 family are recovered in the "long" repeat class, i.e., they occur in a context of repetitive DNA sequence extending $\geq$2 kb. The approximately 2500 members of this family appear to be closely homologous, since no distant relatives are observed in the genome, even under relaxed conditions. Sequences of the 2108 family also belong to the long repeat class. However, this family differs from the 2034 family in that there are in the genome, besides the 20 closely related members referred to in Table 1, many additional sequences which are only

distantly related to the CSp2108 probe sequence. These are discussed in detail in a later paper of this series (Scheller et al., 1980). The large 2109 family consists mainly of short repeat sequence elements, and its members display degrees of relatedness ranging down to the lowest level permitted by the reaction conditions. This is indicated by the low average thermal stability of heteroduplexes between CSp2109 probes and genomic DNA (Table 1). Family 2109 has been separated into subfractions, a set of sequences reacting with a 180 nt probe representing one end of the CSp2109 insert (2109 A), and a set of sequences reacting with a probe containing the remaining 111 nt (2109B). The genomic repeat sequence included in CSp2109 originally contained both the "A" and "B" portions (Posakony et al., 1980).

The three repeat families differ greatly in their representation in sea urchin RNAs. Family 2034 is expressed mainly in intestine nuclear RNA; family 2108 transcripts occur mainly in oocyte RNA; and that portion of the 2109 sequence represented by the "A" probe is expressed primarily in intestine nuclear RNA, while transcripts homologous to the "B" probe are most prominent in gastrula nuclear RNA. The latter observations suggest that although short, 2109 sequences have a complex structure in that some sequences homologous to the "A" portion of the repeat may occur in (transcribed) regions lacking the "B" element, and vice versa.

(b) Frequency of occurence of λ-sea urchin DNA recombinants bearing genomic members of the repetitive sequence families

The probe repeat sequences described in Table 1 were labeled at the 5' termini by the kinase procedure, and were used to screen the recombinant λ-sea urchin DNA genome libraries. The number of positive plaques obtained in these screens provides initial evidence on the distribution of the repeat family members

in the genome. Thus, if the sequences belonging to a given family were widely distributed about the genome, the positive λ-recombinants would usually each contain only a single sequence element homologous to the repeat probe. The number of positive plaques expected is directly calculated for this case from the number of plaques screened and the reiteration frequency of the probe repeat family determined by renaturation kinetics. On the other hand, if in the genome the sequences of a repeat family occur in clusters, positive λ recombinants should often contain several copies, given the 10 to 18 kb length of the inserts, which is large compared to the length of most repeat sequences. In this case, the number of positive plaques expected would be correspondingly smaller.

Radioautographs of 2109A and 2108 plaque screens are shown in Figure 1(A) and 1(B). Both plates contained about 2×10^4 plaque-forming units (pfu). While 393 plaques scored as positive with the 2109A tracer, only 9 plaques reacted with the 2108 tracer. If all or most members of these families occurred singly in the λ recombinants, the reiteration frequencies listed in Table 1 predict that about 370 clones per plate should have reacted with the 2109A probe, and about 8 clones per plate should have reacted with the 2108 probe. Since these expectations are close to the observed values, it follows that the majority of the members of each of these two repetitive sequence families are scattered in diverse regions of the genome, rather than clustered.

Data for screens with all four probe sequences are given in Table 2. Here it can be seen that a screen carried out with the 2109B probe yields a result consistent with the 2109A experiment shown in Figure 1(a), and thus also indicates a nonclustered organization for the 2109 family. In contrast, the number of positive 2034 plaques is only about one-third that expected if members of this family were generally to occur singly. The implication is that there are an average of three

2034 sequences per positive λ-recombinant. Additional evidence suggesting multiple local clusters of 2034 family members in the λ recombinants is that many plaques react with the labeled probe sequence much more intensely than others. Since, as Table 1 shows, the members of this family are all closely related, this is likely to be due to differences in the number of copies per λ insert, rather than to variability in the degree of homology of the sequences.

When library screens were carried out with the 2108 probe at a more permissive criterion than used for the experiment shown in Figure 1(b), many more reactive plaques could be detected (not shown). These recombinants contained distant genomic relatives of the CSp2108 probe sequence. The set of 2108 plaques selected for rescreening and clonal purification included examples of both closely and distantly related sequences. The number of λ recombinants isolated for more detailed analysis of the sequence organization of all three repeat families is listed in the last column of Table 2.

(c) Occurrence of the specific repeat sequence in λ genome library recombinants

The number of sequence elements reacting with the cloned repeat probes was estimated for many of the λ recombinants by the gel blot method (Southern, 1975). The DNA was digested with various restriction enzymes, and the fragments were separated by agarose gel electrophoresis and blotted onto nitrocellulose filters for hybridization with the respective repeat tracers. Results from several representative experiments are shown in Figure 2. It can be seen in Figure 2(a), for example, that only one out of many restriction fragments reacts with the CSp2109B probe in each of two λ2109 recombinants (lanes A and B), while in contrast, the two λ2034 recombinants shown (lanes C and D) apparently contain multiple copies of this repeat, or elements thereof. The two λ2108 digests (lanes

E and F) differ in that one contains a single band which reacts with the CSp2108 probe, while the other displays three reactive subfragments. In Figure 2(b) are reproduced several subdigests and blot hybridizations which were carried out on individual restriction subfragments initially found to contain reactive sequence(s). The examples shown include 5 cases in which the homologous repeat element is confined to only one relatively short subfragment (lanes A, B, C, E and G) and two cases in which there were clearly multiple copies of the relevant sequence clustered within the original reactive subfragment (i.e., λ2034-4, lane F; and λ2109B-16, lane D).

Since the repeat sequence elements are at least several hundred ntp in length, restriction sites will not infrequently occur within them. Indeed, primary sequence data (Posakony et al., 1980) indicate that sites for some of the enzymes used in these experiments occur in the cloned repeat probes themselves, and it is probable that some members of the repeat family will contain the same sites. In a gel blot experiment, this would result in two (or more) bands for a single repeat sequence element. On the other hand, multiple homologous sequence elements could be present in any single reactive fragment large enough to contain them, as in the cases illustrated in Figure 2(b). To decrease the probability of misinterpretation due to either of these causes, we subjected each of the λ recombinants analyzed to digestion with a number of different restriction enzymes. Table 3 displays the results of many such experiments, and includes our best estimates for the number of relevant repeat family members per whole sea urchin DNA insert.

Table 3 shows that sequences of the 2109 repeat family, probed with either A or B tracers, generally occur only once in a given λ recombinant. However, there are several exceptions. λ2109B-8 clearly contains at least two copies of

the repeat, as does λ2109B-16, among others. In these particular cases, the multiple copies are known not to be contiguous to each other from mapping data, some of which are presented below. Though only a minority of the λ-2109 recombinants include multiple copies of the repeat sequence, we note that the frequency with which they do occur is higher than would be expected if the elements of a 1000 member family were distributed completely at random throughout the genome.

The 2034 clones examined except one (λ2034-14) contain multiple copies of this sequence. In the sample included in Table 3, the average number of these sequences per insert is 2.6, which can be compared to the estimate of 3 copies per insert derived from the data in Table 2. The agreement indicates that the presence of multiple copies of this repeat in the various 2034 λ recombinants does not induce frequent deletions, at least through the several rounds of replication required for each liquid culture amplification.

The 2108 family is a complex assemblage of variably related sequences, as noted above. With one exception, those λ recombinants selected at high criterion each contain only a single copy of the 2108 sequence. Though the particular restriction fragments which include this sequence are all the same in size (Table 3) the inserts of the various high criterion λ-2108 recombinants are diverse with respect to their overall restriction digest pattern. It follows that the "high criterion" 2108 sequence element is part of a longer repeat unit which is itself interspersed in many different regions of the genome. The λ-2108 recombinants selected at a more permissive criterion occur in 1, 2, or 3 copies per insert. The degree of relatedness and the actual size of this "superfamily" of diverse sequences is not well known (see Scheller et al., 1980), and the statistical significance of the multiple local occurrences of these sequences remains uncertain.

(d) Sequence environment around repeat family members

In the following experiments, we determined the genomic repetition frequency of sequences flanking several members of each repeat family. While most repeats in the sea urchin genome are short and are surrounded by single copy regions, data from other sources (summarized in Table 1) indicated that the 2108 and 2034 families belong to the long repeat class. Thus, it is expected that on the average, the sequences immediately flanking the elements reacting with these two probes in the λ recombinants would be repetitive as well. There was no previous information regarding either the repeat length or the sequence environment characteristic of the 2109 family.

Restriction subfragments bearing members of the three repeat families were isolated from a number of the selected λ recombinants. Figure 3 displays the location of various enzyme sites within these subfragments, and indicates the positions of the specific repeat sequences (Star). In most cases, there is only one such sequence element within the several thousand ntp of the mapped fragment, as expected from the data of Table 3, for at least the majority of 2109 and 2108 examples. However, the restriction maps provide interesting additional information regarding the arrangement of the 2034 repeat sequences. At least two different subfragments bearing these repeats were isolated and mapped from λ2034-4, λ2034-9, and λ2034-10. Figure 3 shows that all of the mapped subfragments from these recombinants contain multiple copies of the 2034 repeat. In the case of λ2034-4 the two mapped subfragments are identical in the pattern of restriction sites except for one HaeIII site. All the other 2034 restriction maps are unique. These data suggest that the genome includes both tandem 2034 sequences, some of which are near exact replicas, and tightly clustered but non-exact replicas of regions carrying 2034 sequences, as well as the somewhat more widely spaced

2034 sequence clusters indicated in Table 3. Figure 3 also reveals a 2109 sequence cluster (λ2109B-16). This shows that a repeat family in which most members occur singly throughout the genome may also include occasional blocks of several locally contiguous sequence elements.

Genomic reiteration frequencies were determined within the mapped regions by isolating and labeling the appropriate DNA subfragments with ^{32}P, and reacting them with excess sea urchin DNA. Single copy ^{3}H-DNA kinetic standards were included in each reaction. Figure 4 shows representative examples of renaturation kinetics for the relevant regions of three λ recombinants. Reactions carried out with a set of subfragments from λ2109B-9 are illustrated in Figure 4(a). The 2109 family member on subfragment A reacts 680 times faster than the internal single copy standard, which is consistent with the overall repetition frequency estimated previously for this family (Table 1). All of the flanking fragments clearly react as single copy sequences. In contrast, the measurements shown in Figure 4(b) and 4(c) demonstrate the repetitive nature of the sequence environments of a 2108 repeat in λ2108-16, and of a 2034 family member in λ2034-18.

Results of a number of experiments of this nature are summarized in Table 4. It is clear from these data that the 2109 family consists primarily of short (i.e., several hundred nt) repeat sequences interspersed in regions of single copy DNA. Two individual exceptions are found in λ2109B-17 and λ2109B-18, in which all of the fragments tested were moderately repetitive. In several cases, the rate of reaction of the flanking sequences was slightly faster than that of the single copy standard, suggesting that one or two related sequences exist somewhere within the sea urchin genome. Table 4 shows that the repeat elements of both the 2108 and 2034 families are usually, though not invariably, flanked by sequences reiterated in the genome to about the same extent.

These determinations of reiteration frequency are probably subject to at least a twofold error (see Table legend). The significance of even some of the larger rate differences observed within the same region (e.g., in λ2108-16) is therefore not clear. If real, these differences could indicate that some of the contiguous repeat sequence elements occur separately, elsewhere in the genome. In one case, the mapped subfragment appears to include a terminal junction of the 2108 long repeat. Thus, all the tested sequences on the left end of the λ2108-15 subfragment are repetitive, while in the orientation shown, the right end is single copy.

(e) Heteroduplex analysis of λ2109 sequences

In Figure 5 are shown electron micrographs of heteroduplexes formed between DNA molecules from different λ2109 recombinants. The structures observed are consistent with the conclusions drawn above regarding the organization of this family. In Figure 5(a), the complete inserts of λ2109A-22 and λ2109A-24 can be observed extending between the forks that represent the beginning of the right and left areas of the Charon 4 vector. The single duplex structure within the sea urchin DNA must be the 2109 sequence, though there are almost certainly many other repeat sequences in both of these long inserts. No other complementary regions can be seen. The duplex structure is 210 ± 25 nt long (Fig. 5(b)). Figure 5(c) displays a heteroduplex between restriction subfragments from λ2109B-9 and λ2109B-14. The only repetitive sequence in both these subfragments is located where the 2109B probe reacts, since the remainder of both subfragments is single copy or near single copy (Table 4). Thus, the identification of the heteroduplex as a 2109 repeat is in this case unequivocal. The length of the homologous region between the four nonhomologous single copy tails is again about 200 nt. Figure 5(d) shows the HhaI subfragment of λ2109B-16, reacting

with itself at a homologous but out of register site. This is almost certainly the 2109 sequence, since the map of this fragment in Figure 3 indicated that it contains at least two 2109 repeat sequence elements. Figure 5(3) shows that the same length of duplex occurs in about the same position relative to the ends of the λ2109B-16 subfragment when this subfragment is reacted with the λ2109-14 subfragment used in Figure 5(c). The experiment thus confirms the unusual double occurrence of the 2109 repeat in the local region of the genome included in λ2109B-16. It is interesting that, as also can be deduced from the map data in Figure 3, the two copies of this repeat are separated by more than 1000 nt of other DNA sequences.

Discussion

(a) Organization of the three repeat families

We now present a brief summary and discussion of the organization of the three repeat families, based on conclusions drawn in Results as well as data from other sources.

The 2109 repeat family includes about 1000 reasonably well matched sequences, though additional more distant relatives exist in the genome. The repetitive sequences of this family are organized for the most part as 200 to 300 nt elements interspersed in single copy DNA. Most individual family members occur distantly from each other, since few of the 10 to 18 kb inserts in the 2109 λ recombinants include more than a single 2109 sequence, and since the number of recombinants reacting with the 2109 probe in genome screens correctly predicts the size of this family. However, there is a minor fraction of 2109 family members which are located only a short distance from other 2109 repeats. Thus, restriction subfragments from several recombinants contained more than one 2109 sequence element. Three such subfragments were mapped (Fig. 3) and one, from λ2109B-16,

includes members of this family which may be separated by no more than 1000 nt of the other sequence. The fraction of 2109 family members in the genome which occurs in multiple local arrays may be overestimated in the set of 2109 λ recombinants, since their stronger screening signals could have resulted in a bias toward their selection. We cannot exclude the alternative that there is actually more clustered 2109 sequence organization in the genome than we observed if such arrangements are unstable in the λ recombinants. However, we regard this as highly unlikely because of the quantitative agreement between the reiteration frequency of the 2109 family measured in whole DNA and the screening results shown in Table 2. Furthermore, many recombinants containing clustered copies of 2034 repeat sequences have been recovered and studied without difficulty.

The 2108 sequence is an element of a longer repeat which extends for several thousand nt, as shown for several examples in Table 4. Some members of this family studied by Scheller et al. (1980) are about 4.5 kb in length. This result is consistent with the observation (Moore et al., 1980) that the 2108 probe preferentially reacts with a long repeat fraction of the genome. The 2108 family is organized in a complex and interesting way. When reactions are carried out under fairly stringent criterion conditions, only about 20 sequences closely related to the CSp2108 probe are observed. However, many more distant relatives, constituting a large "superfamily", exist in the genome. Differences in the apparent reiteration frequencies between several of the starred 2108 sequences in Table 4 are due to the fact that these sequences belong to different subsets of the 2108 superfamily. Despite this complication, it is clear from the experiments shown in Table 3 that most (or all) of the closely related, 20 member 2108 sequence subfamily occur singly; i.e., that they reside in regions of the genome which are at least 10 to 20 kb apart. Table 3 shows that the large number of less closely related members of the 2108 superfamily must also occur in many different locations.

The 2034 sequence is also an element of a long repeat, but members of this family are all closely related. The amount of sequence divergence between various 2034 family members (about 4%) is no greater than between average single copy sequences of different individual S. purpuratus genomes (Britten et al., 1978). This large family has about 2500 members and these sequences are partially clustered and partially distributed in the genome. At one extreme are regions such as those in Figure 3 which include tandem or tightly clustered 2034 sequences. However, most clusters cannot contain many copies of the 2034 sequence, since the number of reactive restriction subfragments per λ recombinant is not very large (Table 3), and since the 2034 genome screens revealed as many as 1/3 of the number of positive plaques expected on the basis of single occurrences per insert (Table 2). Thus, there are probably several hundred small clusters of 2034 sequences in the genome, as well as some single occurrences (e.g., in λ2034-14). In addition, the genome contains one or more very large deposits of tandemly arranged 2034 sequences. The evidence for this is presented elsewhere by Moore et al. (1980), who noticed a prominent band in 2034 genome blots which includes a significant though minor fraction of the 2034 sequence. The restriction pattern and tandem sequence organization expected for an element derived from this genomic repository of 2034 sequences is displayed by the subfragment of λ2034-4 shown in Figure 3.

(b) Global distribution of members of individual repeat families

The major finding reported here is that the members of all three repeat families populate many separate regions of the genome. Suppose, for example, that the organization of the 1000 member 2109 family in the genome is reasonably represented in the set of λ recombinants listed in Table 3. There are 40 single occurrences of 2109 sequences per average 15 kb insert, plus seven double occurrences. It follows that even on the extreme assumption that all 2109 sequences

are as close as allowed by these statistics, (i.e., about one 2109 sequence every 20 kb), this short repeat family would be dispersed over no less than 2×10^7 nt of DNA. Thus, unless the λ recombinants are not representative, the 2109 family could not be confined to one or a few chromosomal "domains" of the size observed in Drosophila cells (Benyajati & Worcel, 1976), or of the size of polytene chromosome bands. Most likely, the sequences of the 2109 family are dispersed over a large portion of the genome, although this is not shown. It is also clear from the foregoing that long repeats, such as the 2108 and 2034 sequences, are widely distributed in the genome. In situ hybridization has demonstrated extensive genomic dispersion of several cloned Drosophila long repeats as well (Finnegan et al., 1978; Potter et al., 1979; Wensink et al., 1979).

Interspersed short repeats and their flanking sequences account for a major portion of the genome in organisms such as Xenopus (Davidson et al., 1973; Chamberlain et al., 1975), human (Schmid & Deininger, 1975), and sea urchin (Graham et al., 1974). Our present findings illuminate an important consequence of repetitive sequence interspersion. Each interspersed repeat family organized like the 2109 family endows the genome with a network of sequence relationships which links physically distant single copy regions of the DNA. In the sea urchin genome there are perhaps half a million short repeat sequence elements, belonging to at least 10^3 nonhomologous families. The matrix of relationships constructed by these sequences extends throughout most of the genome, since only a minor fraction of the single copy DNA is devoid of them.

The concept of genomic sequence organization advanced here is abstract, and is based solely on structural information. It may or may not have relevance to the coordination of genome function. However, it is clearly important in considering the processes by which the genome has evolved.

(c) The evolutionary process of repeat sequence dispersion

In previous articles, we have drawn attention to the possibility that evolutionary mechanisms exist for the insertion (and disappearance) of interspersed repetitive sequences. This is directly implied by a recent study (Moore et al., 1978) which showed that some repeat sequence families differ markedly in size among related species of sea urchin. One way of envisioning such a process is as follows. At an early stage in the growth of a repeat sequence family, one or a few potentially large blocks of tandem repeats are copied from a preexisting sequence. A mechanism then begins to function by which the length of these blocks is progressively decreased, as fragments of them are excised and inserted elsewhere in the genome. For example, out of register paring might occur, followed by excision of nonaligned sequences, some of which might then reinsert at random nonhomologous locations by the same kinds of reactions as are responsible for insertion of foreign DNA elements in DNA transformation experiments (Wigler et al., 1979). This process would ultimately result in a family of singly occurring repeats interspersed in many regions of the genome. Eventually, decay of recognition among the separated family members would be likely to occur, by means of sequence divergence, deletion, or internal reorganization.

Viewed in light of these speculations, the 2034 family seems to be in an early stage of expansion and dispersion in the genome of S. purpuratus. The large repository of 2034 sequences cited above (Moore et al., 1980) is evidently of recent origin, since this set of tandem repeats is absent from the genome of S. franciscanus, even though the repeat family itself is well represented in the latter species. Most other 2034 sequences now in the S. purpuratus genome are scattered about in clusters, and only a few are as yet singly interspersed. These clustered sequences are probably also of recent origin, given the low overall degree

of sequence divergence in this family (Klein et al., 1978). In contrast, the 2109 family appears to be at an advanced stage of its evolutionary dispersion, though a few nearly contiguous 2109 sequences still exist in the genome, since one example was encountered in our scan of 2109 family recombinants. The relatively large differences between the sequences of many 2109 family members (Klein et al., 1978; Posakony et al., 1980) suggest that the replication event(s) giving rise to the present family occurred longer ago than in the case of the 2034 family.

(d) Repeat family organization and repeat sequence transcription

Interspersed repetitive sequence families are extensively represented in the long heterogeneous nuclear RNA of animal cells, as has been shown by many workers (Darnell & Balint, 1970; Smith et al., 1974; Holmes & Bonner, 1974; Federoff et al., 1977; Scheller et al., 1978). Certain repeat families are also transcribed by Pol III into short nuclear RNAs (Steitz et al., 1980). The representation of specific repeat families in long nuclear RNA is strikingly tissue-specific, at least in sea urchins (Scheller et al., 1978). An interesting aspect is that both strands of each repeat sequence family are represented in different interspersed RNA molecules. That is, multiple members of each repeat sequence are transcribed (asymmetrically) in the diverse transcription units of the nucleus. Interspersed repeats are also represented in a specific pattern in the stored poly(A) RNA of the sea urchin egg, covalently linked to single copy sequence transcripts (Costantini et al., 1980). Among many others transcripts of the 2108 and 2109 sequence families are concentrated in the egg poly(A) RNA. Costantini et al. (1980) concluded that the highly expressed repetitive sequences in egg RNA represent a nonrandom subset of the repeat families in the genome. In other words, members of a minority of families are located in the vicinity of those single copy sequences represented in egg poly(A) RNA.

Repeat family size and organization are remarkably plastic. This raises the question of how evidence for specificity of repeat family expression and for nonrandom genomic location of repeat sequences is to be interpreted. A possible answer is that some repeat family members are transcribed and are in some sense functional, while particularly in cases of large, rapidly expanding families, many of the "newer" copies are not. The networks of homology within the genome resulting from evolutionary repeat family dispersion may thus include both presently expressed sets of repeat sequences, and sets which may potentially be included in useful patterns of expression in future evolutionary time.

This research was supported by NIH Grant GM-20927. D. M. A. was supported by an NIH Postdoctoral Fellowship, and R. H. S. and J. W. P. were supported by an NIH Predoctoral Training Grant.

TABLE 1

Characteristics of three repetitive sequence families

Family[a]	Length of probe sequence (nt)[b]	Repeat length class[c]	Family size: Genomic reiteration frequency[d]	Intrafamilial thermal stability:[e] T_m (°C)	Representation of family in RNA:[f] Oocyte RNA	Gastrula nRNA	Intestine nRNA
2034	498	long	2500	3.7	0.2	0.9	48
2108	204	long	20	5.7	590	95	55
2109 A	180		900	19.7	0.3	1	22
2109 B	110	short	1000	25.3	5	64	1.8

[a]Repetitive sequences were isolated and cloned in the EcoRI site of the plasmid RSF2124 as described in "Materials and Methods" and by Scheller et al. (1977). The 2109A and 2109B probes were isolated from the same plasmid, CSp2109, where they are separated by EcoRI site (Posakony et al., 1980).

[b]Lengths of the cloned repeat fragments have been obtained from their primary sequences (Posakony et al., 1980). The values shown represent the total length of sea urchin DNA present in the plasmids CSp2034, CSp2108 and CSp2109 (i.e., the sum of 2109A and 2109B fragments).

[c]Both long (≥2000 ntp) and short (∿300 ntp) repeat classes have been identified in the sea urchin genome (Eden et al., 1977; Moore et al., 1980). Long repeats are defined as those DNA sequences which are excluded from Sepharose CL2B after low C_0t incubation followed by S1 nuclease digestion, while the short repeats are retarded in gel filtration columns to an extent indicating a mean length of about 300 nt. This is also the typical length of short repeat duplexes observed in electron micrographs of partially renatured DNA (e.g., Chamberlin et al., 1975). The observed repeat length may underestimate the actual length of the repetitive sequence elements, since the renatured duplexes will terminate at the beginning of any nonhomologous sequences, whether these are single copy sequences or other repetitive sequences. While most sea urchin short repeats are in fact terminated by single copy sequence (Graham et al., 1974), this cannot be taken for granted in any specific case. The 2108 and 2034 data are from Moore et al. (1980). 2109A and 2109B data are from these studies and from Scheller et al. (1980).

[d]Measured by the renaturation kinetics of reactions between the labeled cloned probes and excess genomic DNA, assayed by hydroxyapatite binding. Data for 2109A and B are from Moore et al. (1980). The conditions for which these determinations are accurate are 55°C, 0.12 M phosphate buffer. Klein et al. (1978) gave 1000 copies as the size of the 2034 family. However, the average of the four independent kinetic determinations shown later in this paper for 2034 sequences indicate that the proper value is 2500 copies per haploid genome. The measurement shown for 2108 sequences is that of Klein et al. (1978).

[e]Intrafamilial thermal stability (ΔT_m) is the mean thermal stability of the cloned probe fragment minus the mean thermal stability of the population of heteroduplexes formed by reacting the cloned probe with genomic DNA (Klein

et al., 1978). The ΔT_m values shown could be due to scattered mismatch or to short length of homologous sequence elements, or a combination of both factors (cf. Posakony et al., 1980). Mean terminal stabilities were determined (Klein et al., 1978) as the temperature at which 50% of the homologous duplex or heteroduplex populations eluted from hydroxyapatite columns.

[f]Percent representation in oocyte total RNA, gastrula mRNA and intestine nRNA is calculated according to the equation: % representation = $\frac{T_c}{F_c \times T_{sc}} \times 100$, where T_c is the number of transcripts complementary to the cloned tracer per cell, T_{sc} is the number of copies of a typical single copy transcript, and F_c is the genomic reiteration frequency of the repeat family to which a given clone belongs (Scheller et al., 1978).

TABLE 2

Recovery of positive plaques in λ-genome library screens with cloned repeat sequence probes[a]

Repeat family	Positive plaques per plate	Approximate expectation if repeat sequences occur singly[c]	Total No. λ recombinants plaque-purified[d]
2034	304	920	7
2108[b]	9	7	15
2109 A probe[b]	343	370	13
2109 B probe	375	330	27

[a]Probes were obtained from the repeat plasmids CSp2108, CSp2109, and CSp2034 as described in text. Length of the probe sequences is shown in Table 1. All plates contained 2×10^4 pfu per 15 cm plate. Only representative data are shown; a large number of different screening experiments was actually carried out with the several λ-genome libraries described in Materials and Methods, under a variety of conditions. Except for the 2108 experiment, which is described in the Legend to Figure 1, all screens for which data are presented here were carried out under the criterion conditions specified in Materials and Methods.

[b]These screens are shown in Figure 1.

[c]The expected number of positive plaques, N, is calculated for the assumption that the DNA of each recombinant

λ which reacts with the probe sequence contains only 1 copy of this sequence. N= PLR/G, where P is the number of pfu per plate; L is the mean length of the sea urchin DNA insert in the λ recombinants, for which we take 1.5 x 10^4 ntp; R is the repetition frequency of the repeat family, from Table 1; and G is the haploid genome size for S. purpuratus, or 8.1 x 10^8 ntp.

[d]This column lists the total number of positive clones ultimately selected from each family for further analysis.

TABLE 3

Number of restriction enzyme fragments bearing specific repeat family members in selected λ recombinant

Family	λ isolate[a]	Enzyme[b] HaeIII	HhaI	HpaII	HinfI	Estimated number of family sequences per λ recombinant[c]
2109	λ2109A-6	1	1	1	1	1
	8	1	1	1	1	1
	11	1	1	1	1	2[d]
	12	1	1	1		1
	21	1	1	1		1
	22	1	2	2		2
	23	1	1	2		1
	24	1	1	2		1
	25	1	2	1		1
	26	1	1	1		1
	27	1	1	1		1
	28	1	2	1		1
	29	1	1	1		1
	λ2109B-6	1	1	2	2	2
	8	2	2	2	2	2
	9	1	1	1	1	1
	10	1	1	1	1	1
	11	1	1	1	1	1
	14	1	1	1	1	1
	15	2	2	2	3	2

TABLE 3 Continued

Family	λ isolate[a]	Enzyme[b] HaeIII	HhaI	HpaII	HinfI	Estimated number of family sequences per λ recombinant[c]
2109	λ2109B-16	2	2	3	1	2
	17	1	1	1	1	1
	18	1	2			1
	30		1	1		1
	31	1	1	1		1
	32		1	1		1
	38		1	1		1
	39		1	1		1
	40		1	1		1
	42	1	1	1		1
	44	1	1	1		1
	45	1	1	1		1
	46	1	1	1		1
	48	1	1	1		1
	49	1		1		1
	50	1	1	1		1
	52	1	2	1		1
	53	1	1	1		1
	55	1	1	1		1
	56	1	1	1		1
2034	λ2034- 3	2	2			2
	4	2	5	2		4
	9	3	3	3		3
	10	5		4		4

TABLE 3 Continued

Family	λ isolate[a]	Enzyme[b] HaeIII	HhaI	HpaII	HinfI	Estimated number of family sequences per λ recombinant[c]
2034	λ2034-13	2	2	2		2
	14	1		1	1	1
	18	2	2	2		2
2108	λ2108-16	3	2	1	2	2
	21	1	1	1	1	1
high	32	1	1			1
criterion	33	1	1			1
selection	34	1	1			1
	35	1	1			1
	36	1	1			1
	37	1	1			1
	λ2108- 8	3	1	1	1	2
	11	1		1	1	1
low	12	2	1	2	2	2
criterion	15	1	1	2	3	2
selection	17	2		3	3	3
	18		1	2	2	2
	20	1		1	2	1

NOTES TO TABLE 3

[a]All 2034 λ recombinants, 2108 recombinants 1-30, 2109A recombinants 1-20, and 2109 recombinants 1-40 were isolated from genome library SpλR_1B. The other λ recombinants listed were isolated from library SpλH_3C.

[b]Indicated are the lengths of the restriction enzyme fragments which react with the respective probes in gel blot hybridizations such as those shown in Figure 2.

[c]The presence of two or more positive bands in two or more digests was taken to indicate the existence of multiple copies of the repeated sequence. Where two bands occur in only two digests, the conclusion is less certain, as indicated by parentheses.

[d]In this case, the listed fragments all occur twice within the insert.

TABLE 4

Reiteration frequencies for regions surrounding specific repeat sequence elements

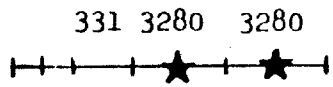
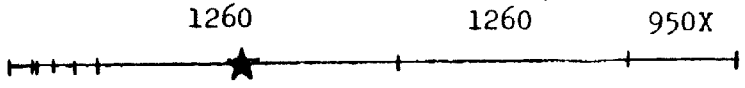
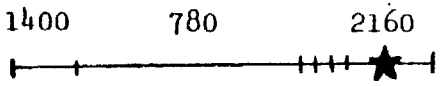
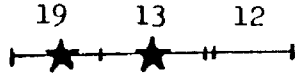
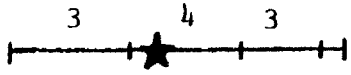
λ recombinant	Digest[a]	Fragment length	K[b]	K corrected[c]	Genomic reiteration frequency[d]	Restriction map[e]
λ2034- 9	HaeIII	640	4.4	4.1	3280	
		640	4.4	4.1	3280	
		560	4.0	4.1	3310	
λ2034-17	HhaI+ HinfI	1400	3.7	1.6	1260	
		1300	3.5	1.6	1260	
		750	1.5	1.2	950	
λ2034-18	HhaI	1380	1.5	0.60	780	
		500	1.6	1.8	2160	
		380	0.90	1.13	1400	
λ2108-11	HinfI	640	0.015	0.014	13	
		560	0.02	0.021	19	
		480	0.012	0.013	12	
λ2108-12	HaeIII	720	0.004	0.0033	3	
		680	0.004	0.0035	4	
		520	0.003	0.0032	3	

TABLE 4 Continued

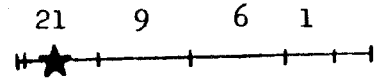
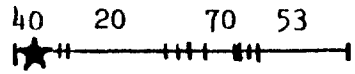
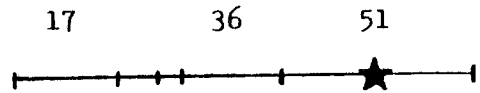
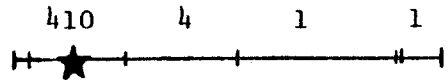
λ recombinant	Digest[a]	Fragment length	K[b]	K corrected[c]	Genomic reiteration frequency[d]	Restriction map[e]
λ2108-15	HinfI	570	0.008	0.009	9	
		510	0.02	0.021	21	
		400	0.005	0.006	6	
		315	0.001	0.0013	1	
λ2108-16	HinfI+ HpaII	600	0.02	0.02	20	
		580	0.052	0.053	53	
		240	0.025	0.04	40	
		160	0.036	0.07	70	
λ2108-21	HpaII	1170	0.10	0.051	51	
		780	0.047	0.036	36	
		650	0.018	0.017	17	
λ2109A-6	HpaII+ HinfI	995	0.0027	0.0016	1	
		680	0.0064	0.0056	4	
		600	0.53	0.53	410	
		260	0.0001	0.00015	1	
λ2109A-8	HpaII	950	0.0025	0.0013	1	
		700	1.4	1.2	920	

TABLE 4 Continued

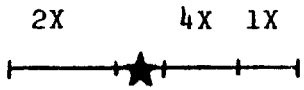
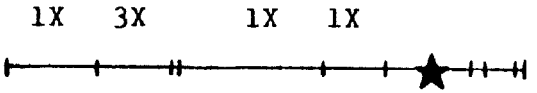
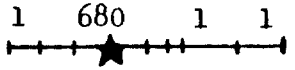
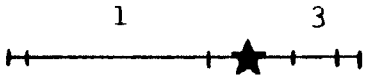
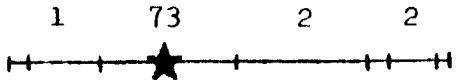
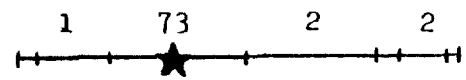
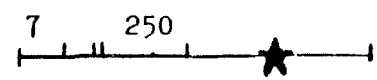
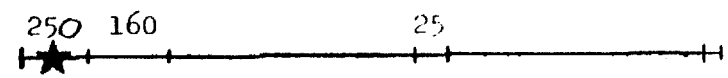
λ recombinant	Digest[a]	Fragment length	K[b]	K corrected[c]	Genomic reiteration frequency[d]	Restriction map[e]
λ2109A-11	HpaII	670	0.0024	0.0021	2	
		450	0.0045	0.0052	4	
		360	0.0012	0.0015	1	
λ2109A-12	HaeIII+ HpaII	900	0.0011	0.00073	1	
		560	0.0011	0.0011	1	
		440	0.0033	0.0039	3	
		400	0.00071	0.0009	1	
λ2109B- 9	HaeIII+ HinfI	440	1.4	1.63	680	
		360	0.0016	0.0021	1	
		340	0.0013	0.0017	1	
		210	0.002	0.0033	1	
λ2109B-11	HhaI+ HincII	1120	0.0038	0.0020	1	
		305	0.0044	0.0061	3	
λ2109B-14	HpaII+ HaeIII	850	0.16	0.11	70	
		800	0.005	0.0037	2	

TABLE 4 Continued

λ recombinant	Digest[a]	Fragment length	K[b]	K corrected[c]	Genomic reiteration frequency[d]	Restriction map[e]
λ2109B-14		300	0.0035	0.0049	3	
		150	0.001	0.002	1	
λ2109B-17[e]	HpaII+ HinfI	1070				
		380			250	
		300			7	
		100				
λ2109B-18	HaeIII	500	0.158	0.173	160	
		440	0.24	0.280	250	

NOTES TO TABLE 4

[a]The restriction enzyme used to digest the subfragment whose map is shown in the last column. For other enzyme sites and further identification of the subfragments, see Figure 3.

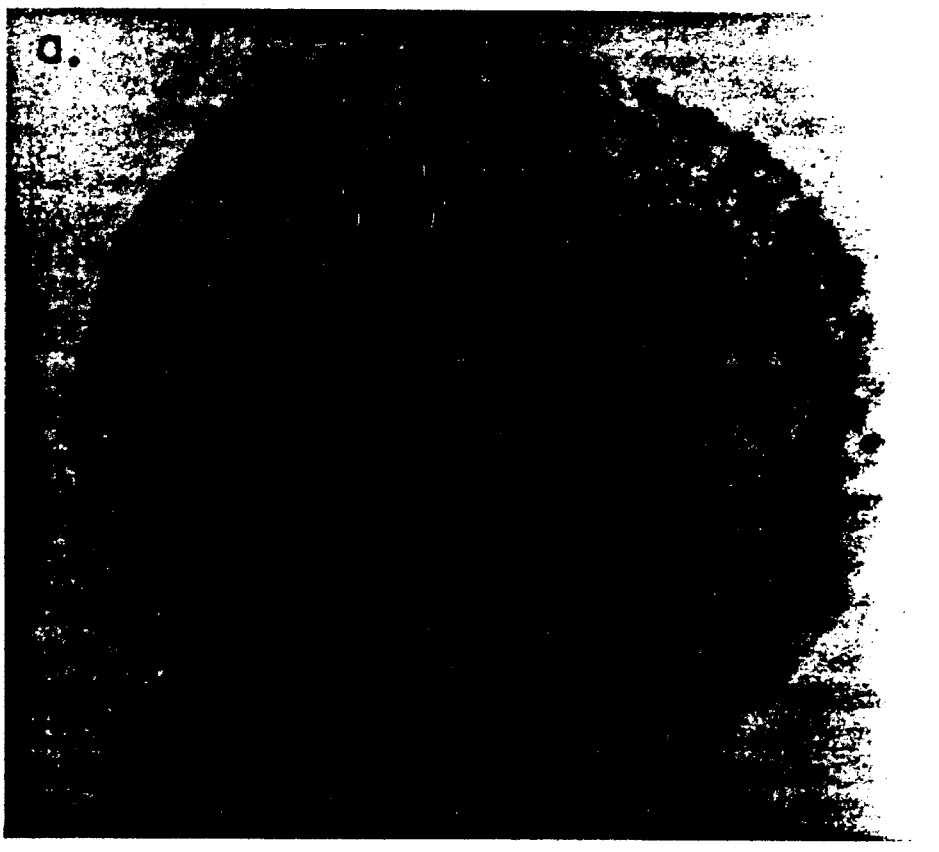
[b]K is the experimentally determined second order rate constant and is given in units of M^{-1} sec^{-1}.

[c]The corrected rate is the rate of reassociation after correcting for the disparity in length between the driver DNA (600 nt) and the tracer fragment (Chamberlin et al., 1978).

[d]The reiteration frequency is obtained by dividing the corrected fragment rate by the rate of reassociation of the internal single copy standard, which varied slightly from experiment to experiment, and averaged about 0.0014 M^{-1} sec^{-1}. The accuracy of these estimates is limited by (1) the small number of data points necessitated by the large number of samples run; (2) the effects of lack of complete homology where repeat sequences are reacted. This error cannot be specifically evaluated for the flanking repeats, since in most cases, thermal stability estimates for the reaction products was not determined. Despite the presence of the internal single copy standard, these factors could induce a two- to threefold error in the estimates given, except for a single copy sequences, and except for 2034 sequences which are in general highly homologous.

[e]Reiteration frequencies measured by the titration method (Scheller et al., 1980).

FIG. 1. Recombinant λ-sea urchin genome libraries screened with cloned repeat tracers. (a) Radioautograph of a plate containing 2×10^4 pfu from library $Sp\lambda R_1 A$, and screened with the 2109A repeat probe described in Table 1. Hybridization was in 0.6 M Na^+ at 55°C, and the screen was carried out by the procedure of Benton and Davis (1977), as described in Materials and Methods. (b) Radioautograph of a plate containing 2×10^4 pfu from library $Sp\lambda H_3 C$, and screened with the 2108 repeat probe. Procedures were as in (a) except that hybridization and subsequent washing of the filter were carried out in 0.1 M Na^+ at 70°C.


a.

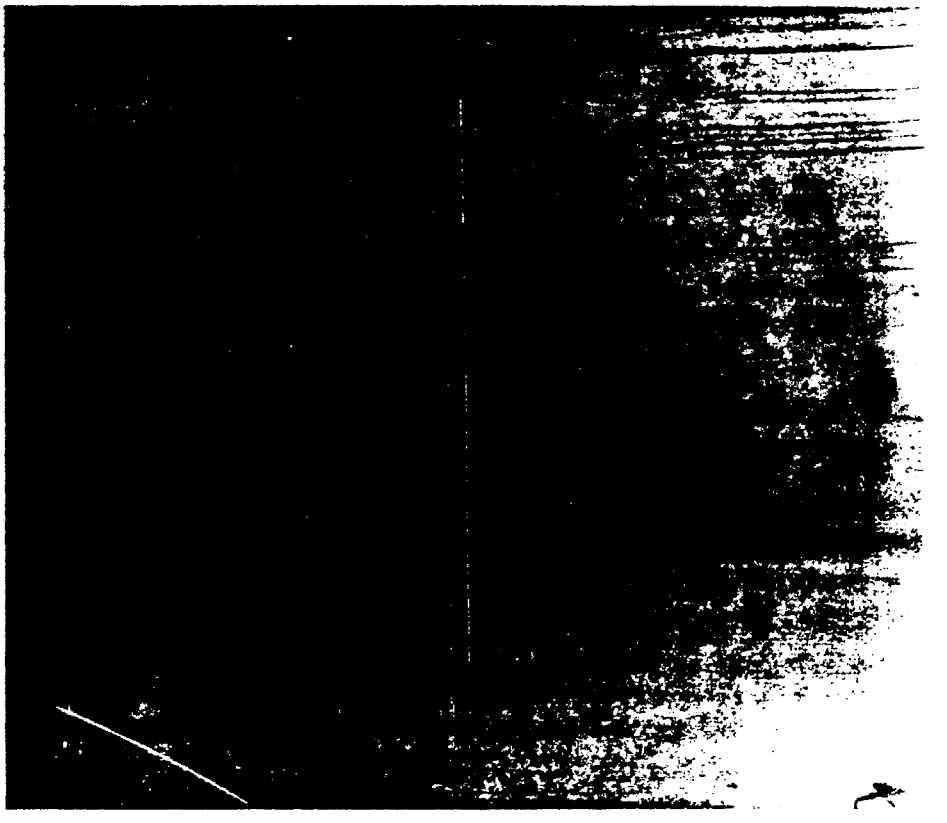
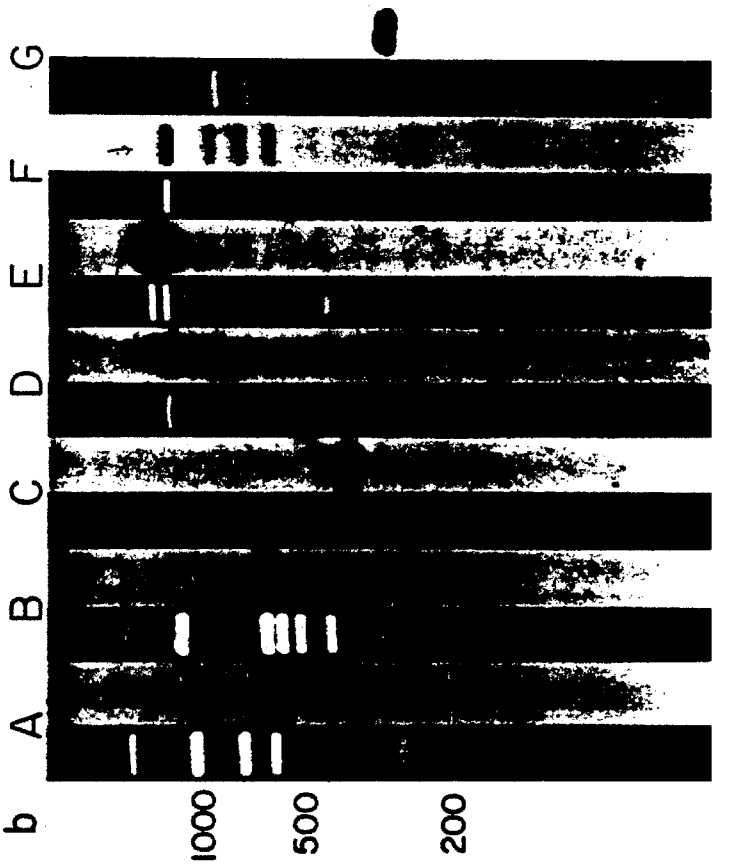


FIG. 2. Gel blot analyses of specific repeat sequences in selected λ recombinants. Isolates from the λ genome libraries were selected by screening with the cloned repeat probes as described in text. (a) The DNAs of the recombinant phage were digested with the indicated restriction enzymes and blotted for hybridization with the same probe fragment initially used to select the respective library isolates (see Materials and Methods for details). Each lane shows on the left the ethidium bromide fluorescence pattern of the digest, and on the right the radioautograph of the corresponding gel blot. Arrows indicate the position of length standards. Lanes A and B: Family 2109 isolates, digests reacted with 5'-^{32}P-2109B probe from plasmid CSp2109. A, λ2109B-9 DNA digested with HhaI; B, λ-2109B-14 DNA digested with HinfI. Lanes C and D: Family 2034 isolates, digests reacted with 5'-^{32}P-CSp2034 probe. C, λ2034-10 DNA digested with HpaII; D, λ2034-13 DNA digested with HpaII. Lanes E and F: Family 2108 isolates, digests reacted with 5'-^{32}P-CSp2108 probe. E, λ-2108-15 DNA digested with HinfI; F, λ-2108-17 DNA digested with HinfI. (b) Subdigests and gel blots of single restriction enzyme fragments which reacted with the repeat probes in experiments such as that shown in (a). The described fragments were eluted from agarose gels, redigested with the indicated restriction enzymes, and the digest again displayed on a gel. Restriction maps of many of the subfragments are shown below, in Figure 3. As in (a), the left hand track in each lane represents the ethidium bromide fluorescence pattern of the subdigest, while the right hand track is the radioautograph of the hybridized subdigest gel blot. A-E: Subfragments from λ2109 isolates, reacted with 5'-^{32}P-2109A probe (A-C) or with 5'-^{32}P-2109B probe (D and E). A, HhaI subfragment of λ2109-6, digested with HinfI + HpaII; B, HhaI subfragment of λ2109A-8, digested with HpaII + HaeIII. C, HhaI subfragment of λ2109A-11 digested with HpaII + HaeIII; D, HhaI subfragment of λ2109B-16, digested with

FIG. 2 (continued)

HpaII; E, HhaI subfragment of λ2109B-18, digested with HindIII. F, HhaI subfragment of λ2034-4, digested with HpaII and hybridized with 5'-^{32}P-2034 probe. G, HhaI subfragment of λ2108-16, digested with HpaII, and hybridized with 5'-^{32}P-2108 probe.


b
A
B
C
D
E
F
G
1000
500
200

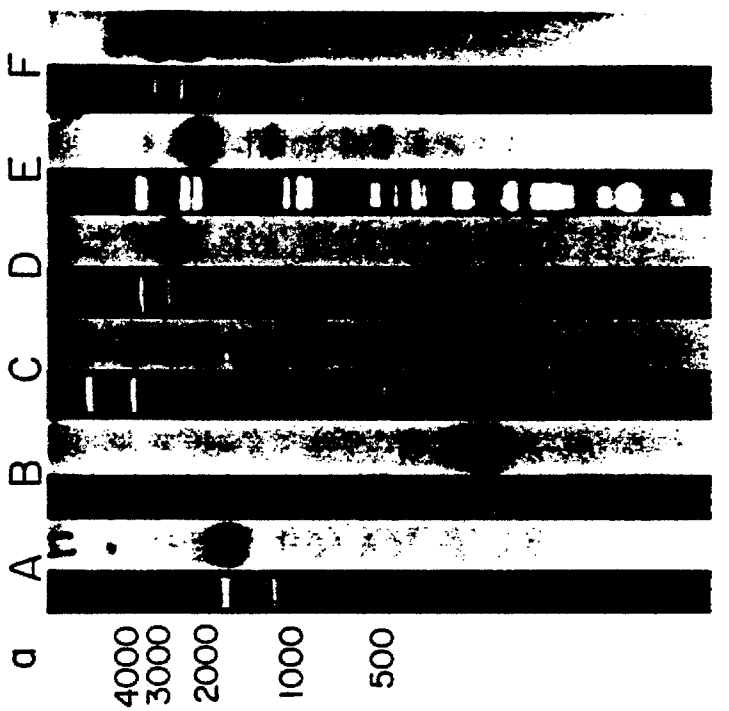

a
A
B
C
D
E
F
4000
3000
2000
1000
500

FIG. 3. Restriction maps of subfragments from λ recombinants bearing repeat family members. Subfragments were obtained from the various λ recombinant DNAs after digestion with the restriction enzymes indicated at the termini. The subfragments shown were mapped by partial and double digests with other enzymes: 1, HaeIII; 2, HhaI; 3, HpaII; 4, HinfI; 5, HincII; 6, TaqI; 7, AvaI; 8, AvaII; 9, HindIII; 0, EcoRI; X, XbaI. Where two subfragments bearing members of the repeat family in a given λ recombinant are released by the same enzyme, these are denoted "A" and "B" in order of decreasing size. The position of the relevant repeat family member(s) (star) was determined by gel blot hybridization as described in the previous section of this paper. The exact location of the repeat sequence within the region defined by the nearest restriction sites is unknown.

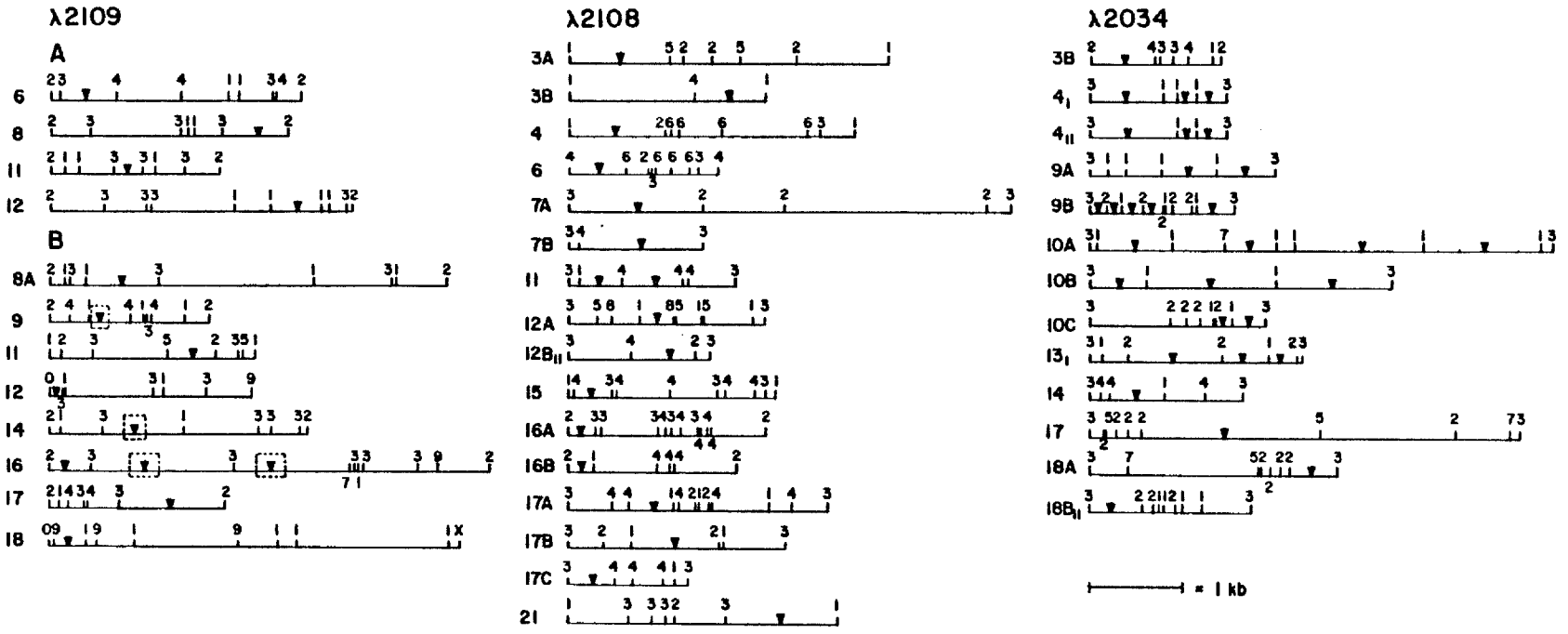

λ2109
A
6
8
11
12
B
8A
9
11
12
14
16
17
18
λ2108
3A
3B
4
6
7A
7B
11
12A
12B II
15
16A
16B
17A
17B
17C
21
λ2034
3B
4 I
4 II
9A
9B
10A
10B
10C
13 I
14
17
18A
18B II
= 1 kb

FIG. 4. Renaturation of subfragments from λ recombinants with excess sea urchin DNA. The subfragments indicated in the restriction map in each panel (see Figure 3 for enzyme sites) were labeled with ^{32}P by the kinase reaction, and reacted with sheared sea urchin DNA in the presence of a ^{3}H-DNA single copy tracer (dashed lines). Details are given in Materials and Methods. The starred subfragment includes the relevant repeat sequence family member. Individual kinetic measurements are indicated by the letters which denote the restriction fragments tested. The rate constants for these reactions are shown in Table 4. (a) HhaI subfragment from λ2109B-9; (b) HhaI "A" subfragment from λ2108-16; (c) HpaII "A" subfragment from λ2034-18.

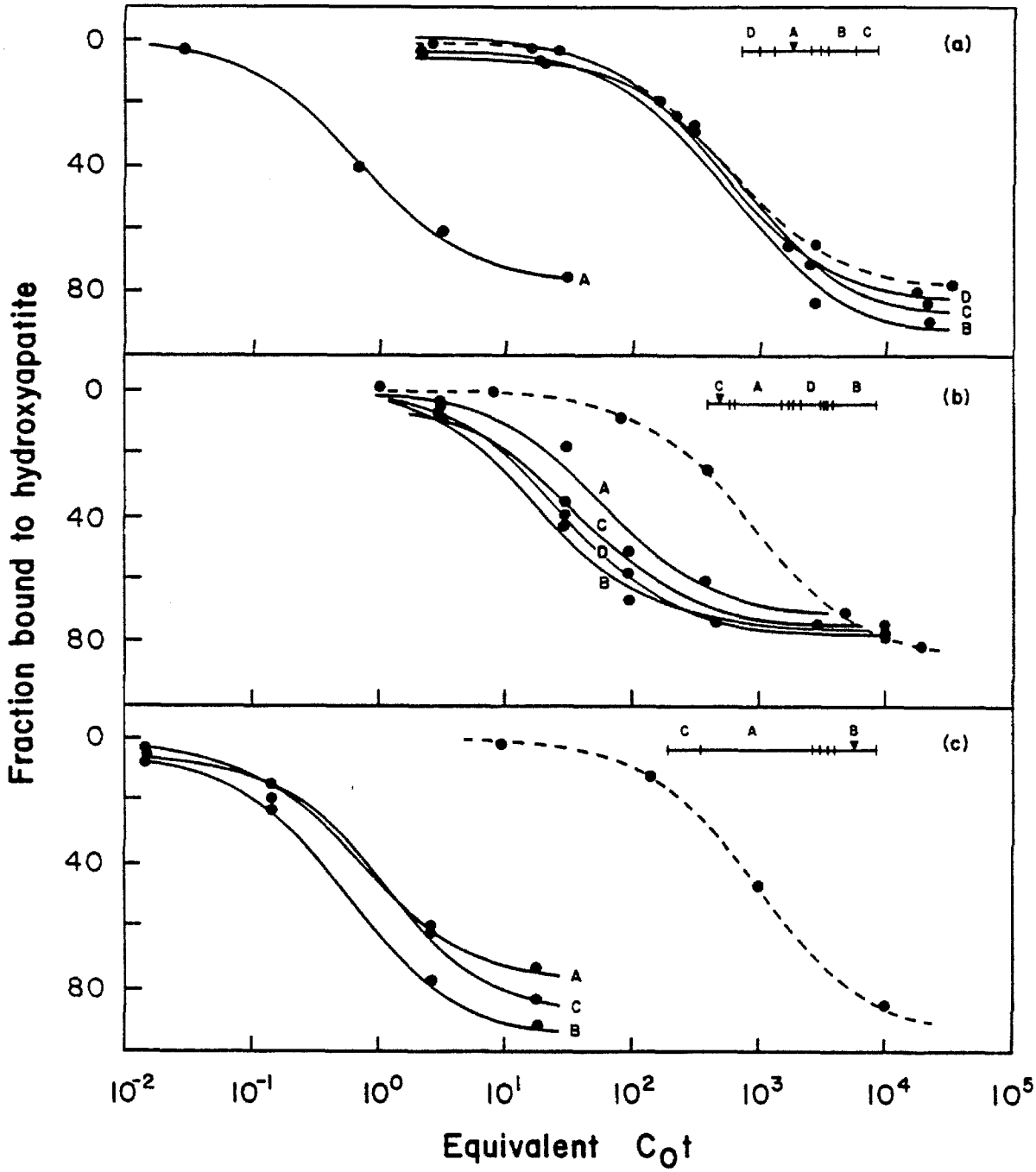
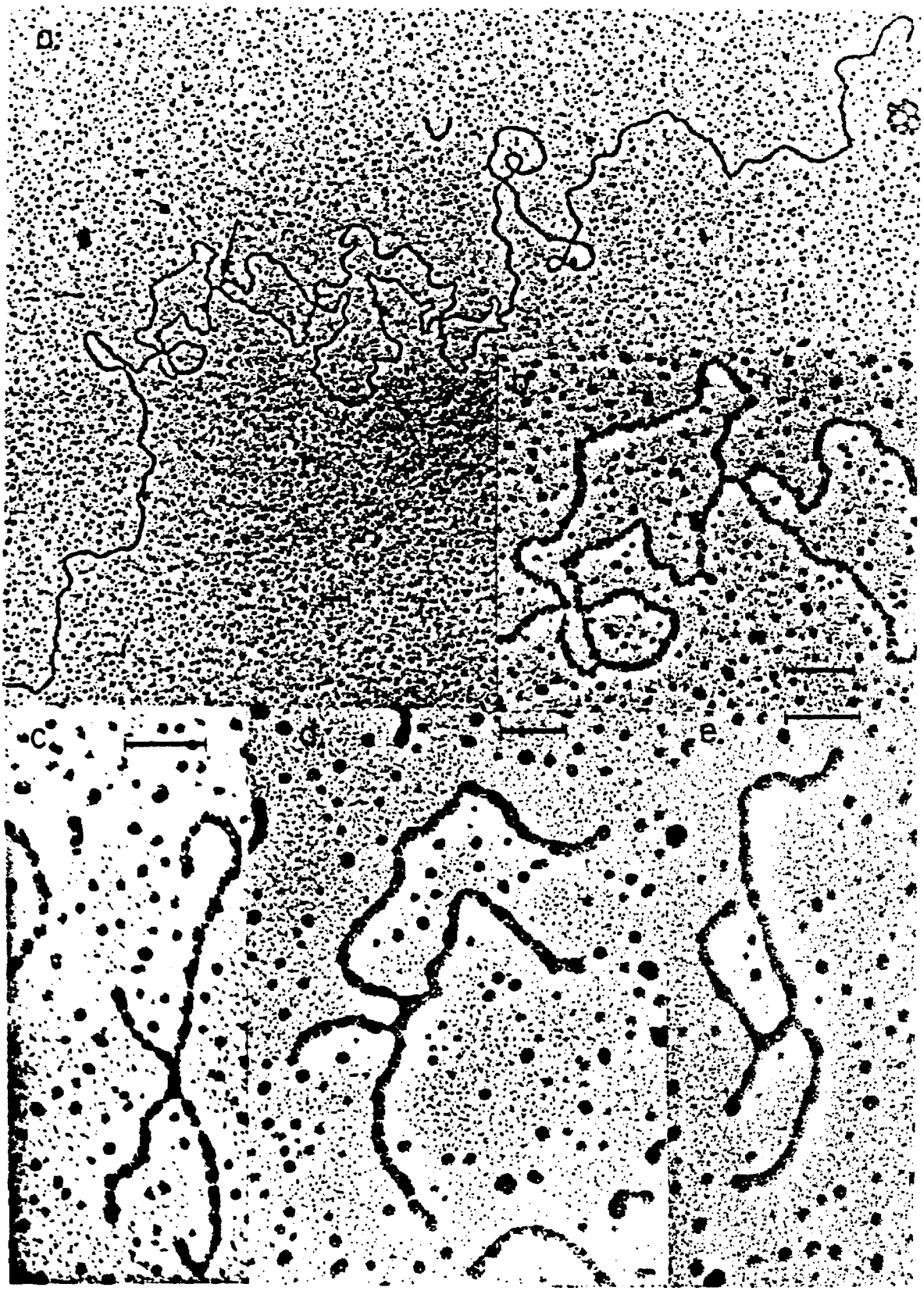

Fraction bound to hydroxyapatite
0
40
80
(a)
D A B C
A
D
C
B
(b)
C A D B
A
C
D
B
(c)
C A B
A
C
B
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}
10^{5}
Equivalent C_0t

FIG. 5. Electron micrographs of heteroduplexes formed between homologous regions of 2109 λ recombinants. (a) Heteroduplex between complete genomes of λ2109A-22 and λ2109A-24. The arrow indicates the homologous region between the sea urchin DNA inserts. The phage were mixed, treated with alkali, and after neutralization hybridized as described by Davis et al. (1971). The DNA was spread from a hypophase containing 55% formamide and a hyperphase of containing 25% formamide (Davis et al., 1971; see Scheller et al., 1980 for details of procedure for this and following panels). (b) Further magnification of the heteroduplex in (a). (c) Heteroduplex between 1.7 kb HhaI subfragment of λ2109B-9 and 2.85 kb HhaI subfragment of λ2109B-14. Restriction maps and reiteration frequencies for these subfragments may be found in Figure 3 and Table 4. (d) Out of register duplex formed in self reaction of 4.7 kb (HhaI subfragment from λ2109B-16). This subfragment contains at least two 2109B sequence elements (Fig. 3). (e) Heteroduplex between 4.7 kb HhaI subfragment of λ2109B-16 and 2.85 kb HhaI subfragment of λ2109B-14. The length of the bar in (b)-(e) is equal to 0.5 kb.


a
c
d
e

References

Benton, W. D. and Davis, R. W. (1977). Science **196**, 180-182.

Benyajati, C. and Worcel, A. (1976). Cell **9**, 393-407.

Blattner, F. R., Williams, B. G., Blechl, A. E., Denniston-Thompson, K., Faber, H. E., Furlong, L.-A., Grunwald, D. J., Kiefer, D. O., Moore, D. D., Schumm, J. W., Sheldon, E. L. and Smithies, O. (1977). Science **196**, 161-169.

Britten, R. J., Graham, D. E. and Neufeld, B. R. (1974). In Methods in Enzymology (Grossman, L. and Moldave, K., eds.), pp. 363-406, Academic Press, New York.

Britten, R. J., Graham, D. E., Eden, F. C., Painchaud, D. M. and Davidson, E. H. (1976). J. Mol. Evol. **9**, 1-23.

Britten, R. J., Cetta, A. and Davidson, E. H. (1978). Cell **15**, 1175-1186.

Chamberlin, M. E., Britten, R. J. and Davidson, E. H. (1975). J. Mol. Biol. **96**, 317-333.

Chamberlin, M. E., Galau, G. A., Britten, R. J. and Davidson, E. H. (1978). Nucleic Acids Res. **5**, 2073-2094.

Costantini, F. D., Britten, R. J. and Davidson, E. H. (1980). Message sequences and short repetitive sequences are interspersed in sea urchin egg poly(A)+ RNAs. Submitted for publication.

Darnell, J. E. and Balint, R. (1970). J. Cell Physiol. **76**, 349-356.

Davidson, E. H., Hough, B. R., Amenson, C. S. and Britten, R. J. (1973). Chromosoma Biol. **77**, 1-23.

Davidson, E. H., Galau, G. A., Angerer, R. C. and Britten, R. J. (1975). Chromosoma **51**, 253-259.

Davis, R. W., Simon, M. and Davidson, N. (1971). In Methods in Enzymology (Grossman, L. and Moldave, K., eds.), pp. 413-428, Academic Press, New York.

Denhardt, D. T. (1966). Biochem. Biophys. Res. Commun. **23**, 641-646.

Eden, F. C., Graham, D. E., Davidson, E. H. and Britten, R. J. (1977). Nucleic Acids Res. **4**, 1553-1567.

Federoff, N., Wellauer, P. K. and Wall, R. (1977). Cell **10**, 597-610.

Finnegan, D. J., Rubin, G. M., Young, M. W. and Hogness, D. S. (1978). Cold Spring Harbor Symp. Quant. Biol. **42**, 1053-1063.

Galau, G. A., Klein, W. H., Davis, M. M., Wold, B. J., Britten, R. J. and Davidson, E. H. (1976). Cell **7**, 487-505.

Graham, D. E., Neufeld, B. R., Davidson, E. H. and Britten, R. J. (1974). Cell **1**, 127-137.

Holmes, D. S. and Bonner, J. (1974). Proc. Nat. Acad. Sci. USA **71**, 1108-1112.

Klein, W. H., Thomas, T. L., Lai, C., Scheller, R. H., Britten, R. J. and Davidson, E. H. (1978). Cell **14**, 889-900.

Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C. Quon, D., Sim, G. K. and Efstratiadis, A. (1978). Cell **15**, 687-701.

Maxam, A. M. and Gilbert, W. (1977). Proc. Nat. Acad. Sci. USA **74**, 560-564.

Moore, G. P., Scheller, R. H., Davidson, E. H. and Britten, R. J. (1978). Cell **15**, 649-660.

Moore, G. P., Costantini, F. D., Posakony, J. W., Britten, R. J. and Davidson, E. H. (1980). Science, in press.

Posakony, J. W., Scheller, R. H., Anderson, D. M., Britten, R. J. and Davidson, E. H. In preparation.

Potter, S. S., Brorein, W. J., Dunsmuir, P. and Rubin, G. M. (1979). Cell **17**, 415-427.

Scheller, R. H., Costantini, F. D., Kozlowski, M. R., Britten, R. J. and Davidson, E. H. (1978). Cell **15**, 189-203.

Scheller, R. H., Anderson, D. M., Posakony, J. W., McAllister, L., Britten, R. J. and Davidson, E. H. In preparation.

So, M., Gill, R. and Falkow, S. (1975). Mol. Gen. Genet. **142**, 239-249.

Southern, E. M. (1975). J. Mol. Biol. **98**, 503-517.

Wensink, P. C., Tabata, S. and Pachl, C. (1979). Cell **18**, 1231-1246.

Wigler, M., Sweet, R., Sim, G. K., Wold, B. J., Pellicer, A., Lacy, E., Maniatis, T., Silverstein, S. and Axel, R. (1979). Cell **16**, 777-785.

Yamamoto, K. R., Alberts, B. M., Benzinger, R., Lawhorne, L. and Treiber, C. (1970). Virology **40**, 734-744.