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 reassocciated 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>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Chemical Synthesis of Restriction Enzyme Recognition Sites Useful for Cloning</td>
<td>19</td>
</tr>
<tr>
<td>3</td>
<td>Clones of Individual Repetitive Sequences from Sea Urchin DNA Constructed with Synthetic Eco RI Sites</td>
<td>24</td>
</tr>
<tr>
<td>4</td>
<td>Characteristics of Individual Repetitive Sequence Families in the Sea Urchin Genome Studied with Cloned Repeats</td>
<td>29</td>
</tr>
<tr>
<td>5</td>
<td>Repetitive Sequence Transcripts in the Mature Sea Urchin Oocyte</td>
<td>42</td>
</tr>
<tr>
<td>6</td>
<td>Specific Representation of Cloned Repetitive DNA Sequences in Sea Urchin RNAs</td>
<td>58</td>
</tr>
<tr>
<td>7</td>
<td>Evolutionary Change in Repetition Frequency of Sea Urchin DNA Sequences</td>
<td>74</td>
</tr>
<tr>
<td>8</td>
<td>Repetitive Sequences of the Sea Urchin Genome</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I. Distribution of Members of Specific Repetitive Families</td>
<td>87</td>
</tr>
</tbody>
</table>
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 $8.3 \times 10^8$ 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 $3.7 \times 10^7$ 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
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 $2.6 \times 10^6$ nucleotide pairs is observed. Adult tissues are characterized by relatively small messenger RNA complexities of about $6 \times 10^6$ 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 $1.7 \times 10^8$ 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, $3.5 \times 10^7$ 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 occurred 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 Tm of 42°C in 0.1 M NaCl. Sea urchin DNA greater than 2000 nucleotides in length was denatured and re-natured, 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\text{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 prevalence 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 many individual member 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 \( \lambda \) clones were determined. The 2034 hybrids all melted \(< 6\degree C\) below the self \( T_m \). Clone \( \lambda 2034-4 \) which contains elements from the 1.7 kb prominent genomic land (Moore et al., 1980b), and melts only 2\degree C below the self \( T_m \) indicating polymorphism less than the 4\degree C observed for the average single copy sequence (Britten et al., 1978). The 2108 \( \lambda \) clones, however, show a different organization. There are two types of hybrids formed, those with stability \(< 6\degree C\) below the self \( T_m \), and those with \( \Delta T_m \) \( > 20\degree C \). No intermediate stability fragments were found.

The clones that showed (high) \( \Delta T_m \) s 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 \( > 5 \) 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 $T_m$'s. 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 recent 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 to 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


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.


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


CHAPTER 2
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 (5-7). 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 octadecanucleotide 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-triisoprop-
pybenezesulfonyl 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-cyanomethyl group on the 3’ ends. These conditions subsequently were changed for the Hind III and Bam I decamer syntheses (see legend of Fig. 3). The pyridine-ethanol mixture provided a faster, more efficient removal of β-cyanomethyl protecting groups than tetrathydrofuran did. 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 NH₄OH 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-TM urea gel electrophoresis, with the same Rₚ 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 Tm of about 48°C in 0.3M NaCl, 0.1 mM EDTA, 10 mM Tris-HCl, pH 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-32P]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 Hind III restriction endonuclease site (GGCC). Polymers of the three linkers are cleaved back to decamers by Hae III enzyme (lane 9 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 polydeoxycadenylate [poly(dA)] or polydeoxycytidylate [poly(dT)] (9) or, more recently, with polydeoxyguanylate and polydeoxycytidylate [7]; and A. Otsuka, personal communication; (iii) blunt-end ligation by T4 ligase (12). This latter method is based on the ability of T4 ligase to readily 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 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'-phosphateto3'-hydroxyl flux 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 RSP2124 has an Eco RI site within its colicin genes (10), and therefore one can screen for insertion by looking for disruption of colicin production after recombination. The plasmids pBR316 and pBR322 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 reagents were described in the 1983 edition of this text. The purification of the protected polynucleotide fragments was aided by the use of medium-pressure liquid chromatography. Chromatography glass columns with Fluid Membr KPG-50 solvent pumps were used to purify the intermediate and decameric products. The 2-inch columns were run at 25 pounds per square inch. The 1-inch columns were run at 100 pounds per square inch. The 15-inch columns were run at 100 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 0.1M 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 (i). Final deblocking of the decameric products was done in a 1:2 mixture of pyridine and 30 percent NH₄OH 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 were then run on cellulose thin-layer chromatography plates in a mixture of 100 ml H₂O and 60 ml of isobutaric acid for 16 hours. The plates were dried, and the slowest moving band was recovered from the cellulose in 3 percent NH₄OH. Evaporation of solvent gave a white residue, which was dissolved in 70 mM phosphate buffer at pH 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 0.1M ammonium bicarbonate buffer, pH 7.5. Abbreviations: DMT, 4,4-dimethoxytrityl protecting group for the 3' end; CCCN, β-cyanoethyl protecting group for the 5' end; Ac, acetic ester protecting group for the hydroxyl end. For details about TPST, see (i). 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 labeled with T₄ polynucleotide kinase (5), then treated as indicated. Conditions for T₄ DNA ligase (blunt-end ligation) were 66 mM tris, pH 7.6, 6.6 mM MgCl₂, 10 mM dithiothreitol, 0.4 mM adenosine triphosphate (ATP), 1 μM [γ-⁵³⁷P]ATP, 0.05 units of T₄ ligase, 27°C, 16 hours, 15-μl volume. Conditions for Eco RI were 100 mM tris, pH 7.5, 5 mM MgCl₂, 0.02 percent Triton X-100, 1 μg of [γ-⁵³⁷P]DNA, 20°C, 5 hours, 10-μl volume. Conditions for Bam I were 100 mM tris, pH 7.5, 5 mM MgCl₂, 1 μg of [γ-⁵³⁷P]DNA, 27°C, 5 hours, 10-μl volume. The enzymes Eco RI, Bam I, Hpa I, Alu I, Hae III, T4 DNA ligase, and T4 polynucleotide kinase were the gifts of D. 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, 7M urea, at 15 volts/cm for various times in 0.05M tris-borate, pH 8.3.
(R. Rodriguez, F. Bolivar, H. Goodman, H. Boyer, and M. Bestach, 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 (13), 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 R1 (H. Heyneker, personal communication), and Eco R1 to Bam I (A. Riggs, unpublished data). Therefore, the chemically synthesized decamer linkers reported here are generally applicable tools for molecular cloning experiments.

The three decamer linkers themselves were constructed with CC...GG at the ends for efficiency 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, "polyniker" clones have not yet been obtained, perhaps because of immobilization 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 Technology, Pasedena 91125
Institute of Technology, Pasedena 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

2. The Eco R1 decamer (CAAGCTTGGG) contains the specific sequences for Bam I, Bst I, Bsu N1, Eam I, Mbo I, Mse I, Hpa II, Hae II, and Mnl I. The Hind III decamer (CCCGAATTCCTGG) contains sequences for Bgl I, Bcl I, Bsa I, Bst I, Hind II, and Hpa II. After polymerization, each decamer will contain the sequence GGGG recognized by Hae III, Bsu I, Bst I, Hpa I, Pst I, and Sca I. For a review of restriction enzymes and their recognition sequences, see R. Roberts, "Control of DNA Synthesis," in press.
17. We thank J. Rosenom 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 described in this report, and L. Stoessel for technical help. Supported by NIH grants GM-12121 and HD-04420, NSF grants PC-3671-03864 and GB-3631, and an NIH predoctoral fellowship to R. H. S. This is contribution No. 143 from the Norman W. Church Laboratory of Chemical Technology, California Institute of Technology.

4 February 1977

Copyright © 1977 by the American Association for the Advancement of Science
CHAPTER 3
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\text{m} 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 isolated 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 $C_{\text{m}}$ 40 ($C_{\text{m}}$ 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 $C_{\text{m}}$ 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 200 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 5'-hydroxyl and 3'-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-
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) 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 $10^{-4}$ Amp 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 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 colonies 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 result in unit length linear plasmid DNA molecules plus repetitive DNA fragments. We examined a number of amp 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 considered 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-
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<sup>®</sup> 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 RSFI24 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 ³₂P by the T4 kinase method (18) (Fig. 2). The DNA was renatured after labeling and placed on 3 percent agarose gels along with an Hae III digest of RSFI24. 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 hydroxypatite chromatography and is plotted as a function of the driver DNA C<sub>D</sub> 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 × 10<sup>-3</sup> M<sup>-1</sup> sec<sup>-1</sup> 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 × 10<sup>-4</sup> M<sup>-1</sup> sec<sup>-1</sup>. Thus clone CS2108 contains a DNA sequence which is reassociated 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

![Plasmid DNA](image)

Fig. 2. Analysis of a recombinant plasmid clone constructed by blunt-end ligation. (A) Detection and end labeling of inserted sea urchin fragments. Amp<sup>®</sup> clone CS2108 was grown in 40 ml of L broth plus ampicillin (15 µg/ml) to 4 × 10<sup>8</sup> to 6 × 10<sup>8</sup> cells per milliliter, and the plasmid content was amplified approximately 25 times by growth in L broth plus ampicillin (200 µg/ml). The cells were harvested by centrifugation and washed with a solution of 10 mM tris- HCl (pH 7.4) and 1 mEDTA. They were then centrifuged and resuspended in 3 ml of 25 percent sucrose, 50 mM tris- HCl (pH 8.0), and 40 mM EDTA at 0°C. Lysozyme (1.5 mg/ml Sigma) was added, and the solution was maintained at 0°C for 5 minutes. Then 0.3 ml of 0.5M EDTA and 0.375M EGTA (pH 8.0) was added for 5 minutes at 0°C. Four ml of Brij lysozym buffer (1 percent Brij 58, 0.4 percent sodium deoxycholate, 0.063M EDTA, 0.05M tris- HCl (pH 8.0)) was added and mixed vigorously. Chromosomal DNA was removed from the lysate by centrifugation at 50,000 rpm 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; ρ<sub>e</sub> = 1.665 g/cm<sup>3</sup> in a 50 Ti rotor at 39,000 rpm). Superhelical DNA was removed from the gradient by speed puncture. Ethidium bromide was removed by extraction with isopropanol, and the DNA was dialyzed against 100 mM tris- HCl (pH 7.5), 100 mM NaCl, 5 mM MgCl<sub>2</sub> 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 0.5-ml BioRad column. The DNA was precipitated at −70°C for 2 hours with 1 volume of isopropanol and mixed with 3 ml of sodium acetate (10.8 M) and 30 percent chloroform:isoamyl alcohol (24:1). The mixture was incubated for 2.5 hours and 1 ml of 0.3M sodium acetate at (pH 6.8) was added. The solution was extracted twice with a mixture of 80 percent phenol (pH 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 (19) with [γ<sup>³₂P</sup>]-ATP. Specific activity 1200 cpm/mole (Winston-Sauser'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 redissolved in a volume of 200 µl of 0.1 M NaCl, 0.05M tris- HCl (pH 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<sup>10</sup> to 2 × 10<sup>10</sup> cpm/micron. 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 RSFI24 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 RSFI24 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 5M NaCIO4 and binding the DNA to hydroxyapatite in 0.125M phosphate buffer (PB), followed by thermal elution of the ³₂P-labeled DNA. This DNA was reassociated with excess 500 nucleotide long (5 × 10<sup>5</sup> mass ratio) sheared sea urchin DNA in 0.12M PB at 55°C. The fraction ³₂P-labeled DNA (θ) and driver DNA (Λ) reassociated was determined by hydroxyapatite chromatography in 0.125M PB at 55°C. The solid line is the least-squares solution for the reaction of ³₂P-labeled DNA driven by total sea urchin DNA. The rate constant determined for this reaction is 3.16 × 10<sup>-4</sup> M<sup>-1</sup> sec<sup>-1</sup>. 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 × 10<sup>-4</sup> M<sup>-1</sup> sec<sup>-1</sup>. The dotted line shows the renaturation of total sea urchin DNA (Ω).
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 \( \Phi \), 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.
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-05755 and GM-29027 and by NSF grant BMS 75-03789. R.H.S. is supported by an NIH postdoctoral training grant; T.L.T. by an NIH postdoctoral fellowship; A.S.L. by fellowship J-289; and W.H.K. by Lehn fellow-ship J-160, both from the American Cancer Society-\ California Division.

February 1977
CHAPTER 4
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. The DNA was 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 are present in 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 Strongylocentrotus 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, 0.1 M Na+ are indistinguishable 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., 1975; 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 are cloned with a given 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.
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 hybridization 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 repetitive duplexes included within the original sequence, due to strand inversions, 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 event. 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 RSP2124 (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 rearranged 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 long. 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 32P 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 three-fourths 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

<table>
<thead>
<tr>
<th>Clone</th>
<th>Length (Nucleotide Pairs)</th>
<th>Rate Constant* (M⁻¹ sec⁻¹)</th>
<th>Approximate Reiteration Frequency* (Copies per Genome)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>240</td>
<td>1.5 ± 0.11</td>
<td>2,100</td>
</tr>
<tr>
<td>2007</td>
<td>1100</td>
<td>0.49 ± 0.006</td>
<td>400</td>
</tr>
<tr>
<td>2004</td>
<td>560</td>
<td>1.2 ± 0.12</td>
<td>1,000</td>
</tr>
<tr>
<td>2007A</td>
<td>325</td>
<td>1.1 ± 0.13</td>
<td>900</td>
</tr>
<tr>
<td>2007B</td>
<td>290</td>
<td>0.11 ± 0.013</td>
<td>90</td>
</tr>
<tr>
<td>2008</td>
<td>165</td>
<td>0.19 ± 0.028</td>
<td>130*</td>
</tr>
<tr>
<td>2008A</td>
<td>420</td>
<td>0.0041 ± 0.00034</td>
<td>3</td>
</tr>
<tr>
<td>2009</td>
<td>220</td>
<td>0.17 ± 0.027</td>
<td>140*</td>
</tr>
<tr>
<td>2008B</td>
<td>260</td>
<td>0.0073 ± 0.00060</td>
<td>6</td>
</tr>
<tr>
<td>2007</td>
<td>900</td>
<td>0.044 ± 0.0047</td>
<td>40</td>
</tr>
<tr>
<td>2009</td>
<td>235</td>
<td>0.10 ± 0.021</td>
<td>80</td>
</tr>
<tr>
<td>2100</td>
<td>1000</td>
<td>3.7 ± 0.54</td>
<td>3,100</td>
</tr>
<tr>
<td>2101</td>
<td>220</td>
<td>0.93 ± 0.12</td>
<td>700</td>
</tr>
<tr>
<td>2102</td>
<td>220</td>
<td>0.83 ± 0.19</td>
<td>530</td>
</tr>
<tr>
<td>2104</td>
<td>190</td>
<td>2.1 ± 0.18</td>
<td>1,800</td>
</tr>
<tr>
<td>2108</td>
<td>190</td>
<td>0.025 ± 0.0034</td>
<td>20</td>
</tr>
<tr>
<td>2108A</td>
<td>200</td>
<td>1.1 ± 0.21</td>
<td>900*</td>
</tr>
<tr>
<td>2109</td>
<td>125</td>
<td>0.24 ± 0.026</td>
<td>200*</td>
</tr>
<tr>
<td>2110</td>
<td>260</td>
<td>1.9 ± 0.23</td>
<td>1,800</td>
</tr>
<tr>
<td>2111</td>
<td>153</td>
<td>15 ± 2.1</td>
<td>12,500</td>
</tr>
<tr>
<td>2112</td>
<td>220</td>
<td>0.54 ± 0.087</td>
<td>450</td>
</tr>
<tr>
<td>2125</td>
<td>180</td>
<td>0.082 ± 0.0088</td>
<td>90*</td>
</tr>
<tr>
<td>2131</td>
<td>210</td>
<td>0.97 ± 0.11</td>
<td>1,400</td>
</tr>
<tr>
<td>2133A</td>
<td>310</td>
<td>1.5 ± 0.21</td>
<td>2,100</td>
</tr>
<tr>
<td>2133B</td>
<td>310</td>
<td>0.042 ± 0.013</td>
<td>60</td>
</tr>
<tr>
<td>2137</td>
<td>190</td>
<td>0.37 ± 0.038</td>
<td>530</td>
</tr>
</tbody>
</table>

* 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: Kc = Kd(Lc/Ld), where Kd is the length-corrected rate constant, Kc is the observed rate constant, Lc is the length of the cloned fragment and Ld is the length of the sea urchin DNA driver (Chamberlin et al., 1978).

The reiteration frequency, S, is calculated as S = Kd/Km, where Km is the second-order rate constant for the reaction of the δ42-DNA single-copy tracer with the same sea urchin DNA driver. The mean length of the single-copy tracer, 320 nucleotides, was too small to that of the driver to warrant a length correction. For all samples except clones 2006, 2137, 2133A, 2133B, 2131 and 2125, the single-copy rate constant measured was ~1.3 × 10⁻¹¹ M⁻¹ sec⁻¹. For the latter clones, the rate constant was 0.7 × 10⁻¹¹ M⁻¹ sec⁻¹, probably due to a minor difference in salt concentration.

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 Tm's are dependent upon the reaction conditions. Thus the reaction temperature is lowered to 45°C, the family appears to expand by including more divergent members (see text). For example, when clone 2108B 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.

Measurements (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. Retention Frequencies for Sea Urchin Repetitive DNA Fractions and for the 28 Family Cloned Sample

<table>
<thead>
<tr>
<th>Repetitive DNA*</th>
<th>Component Name</th>
<th>Rate Constant (M⁻¹ sec⁻¹)</th>
<th>Component Size</th>
<th>Calculated Retention Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total* Repeats</td>
<td>1</td>
<td>3.3</td>
<td>0.36</td>
<td>2800</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.097</td>
<td>0.54</td>
<td>80</td>
</tr>
<tr>
<td>Short* Repeats</td>
<td>1</td>
<td>2.8</td>
<td>0.48</td>
<td>2000</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.048</td>
<td>0.54</td>
<td>40</td>
</tr>
<tr>
<td>28 Cloned Repeats</td>
<td>1</td>
<td>–</td>
<td>0.4</td>
<td>1600*</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>–</td>
<td>0.8</td>
<td>50*</td>
</tr>
</tbody>
</table>

* Reactions with the “total” and “short” repeat RH-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 retention times were derived from a least-squares fit of these reactions. Ratios were derived from a least-squares fit of these reactions. Ratios were derived from a least-squares fit of these reactions.

The preparation of the DNA fraction described by Davison et al. (1975), 1600 nucleotides long RH-DNA labeled in vivo was reacted to Cot 8 x 10⁻¹, 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 30. The duplex fraction was collected on hydroxyapatite. The DNA was sheared to 300 nucleotides and reacted to Cot 100, and the duplex fraction ("total repeats") was collected on hydroxyapatite.

The preparation of the DNA fraction described by Costantini et al. (1978), 2700 nucleotides long RH-DNA labeled in vivo was reacted to Cot 40 and treated with S1 nuclease. 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 50, and the duplex fraction ("short repeats") was collected on hydroxyapatite.

* Calculated as log f = (3.5 log I) / 3 L, where f is the retention frequency of an individual repetitive sequence family, and I is the calculated retention frequency of the component.

renaturated 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 tracer 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 clone repeat 2133A. These T_m 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_m for the cloned tracers are listed in column 2 of Table 3. The major sources of difference in the T_m 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 21098, 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_m, 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 $10^4-10^5$ copies per family. No evidence for discrete frequency components is to be found, at least in this 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.

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 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 $\gamma$-HATP from the labeling reaction. Closed circles (a) and (b) show the reactions of the cloned $\gamma$H-DNA with the driver DNA (the reassociation of which is not shown); closed triangles (a) show the reactions of the single-copy $\gamma$H-DNA tracer with the driver DNA. Points taken at Cot 23,500 and 49,200 indicate terminal reactions of the $\gamma$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 2055; (c) clone 2057B; (d) clone 2100.

---

**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 footnote 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 smooched is probably small on the scale shown here.
### Table 3. Thermal Stability Data for the Cloned Repetitive Sequence Families

<table>
<thead>
<tr>
<th>Clone</th>
<th>Native $T_m$ (°C)</th>
<th>$T_m$ (45°C)</th>
<th>$T_m$ (55°C)</th>
<th>$JT_m/T_m$ Native − $T_m$ 45°C</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>62.3</td>
<td>73.2</td>
<td>74.8</td>
<td>9.1 (9)</td>
</tr>
<tr>
<td>2007</td>
<td>72.7</td>
<td>83.1</td>
<td>70.1</td>
<td>3.6 (9)</td>
</tr>
<tr>
<td>2024</td>
<td>79.4</td>
<td>73.7</td>
<td>70.0</td>
<td>3.7 (9)</td>
</tr>
<tr>
<td>2007A</td>
<td>−</td>
<td>−</td>
<td>73.3</td>
<td>− (or II)*</td>
</tr>
<tr>
<td>2007B</td>
<td>−</td>
<td>−</td>
<td>75.8</td>
<td>− (or II)*</td>
</tr>
<tr>
<td>2008</td>
<td>79.8</td>
<td>56.7</td>
<td>64.5</td>
<td>&gt;19.9 (9)</td>
</tr>
<tr>
<td>2009</td>
<td>56.7</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>2000</td>
<td>83.1</td>
<td>82.8</td>
<td>64.3</td>
<td>&gt;20.3 (9)</td>
</tr>
<tr>
<td>2100</td>
<td>−</td>
<td>81.8</td>
<td>52.4</td>
<td>− (or II)*</td>
</tr>
<tr>
<td>2101</td>
<td>78.8</td>
<td>69.1</td>
<td>71.4</td>
<td>7.3 (9)</td>
</tr>
<tr>
<td>2104</td>
<td>83.6</td>
<td>74.3</td>
<td>75.2</td>
<td>9.3 (9)</td>
</tr>
<tr>
<td>2108</td>
<td>88.7</td>
<td>81.0</td>
<td>79.9</td>
<td>3.7 (9)</td>
</tr>
<tr>
<td>2105A</td>
<td>80.5</td>
<td>80.8</td>
<td>58.3</td>
<td>&gt;18.7 (9)</td>
</tr>
<tr>
<td>2105B</td>
<td>83.4</td>
<td>56.1</td>
<td>−</td>
<td>&gt;25.3 (9)</td>
</tr>
<tr>
<td>2110</td>
<td>78.8</td>
<td>72.5*</td>
<td>73.3*</td>
<td>8.1 (9)</td>
</tr>
<tr>
<td>2112</td>
<td>83.3</td>
<td>77.8</td>
<td>81.6</td>
<td>5.7 (9)</td>
</tr>
<tr>
<td>2125</td>
<td>81.7</td>
<td>80.2</td>
<td>64.8</td>
<td>&gt;21.5 (9)</td>
</tr>
<tr>
<td>2131</td>
<td>88.2</td>
<td>75.4</td>
<td>77.0</td>
<td>10.8 (9)</td>
</tr>
<tr>
<td>2133A</td>
<td>88.5</td>
<td>83.0</td>
<td>83.5</td>
<td>3.6 (9)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Short Repetitive DNA Fraction</th>
<th>$T_m$ (45°C)</th>
<th>$T_m$ (55°C)</th>
<th>$JT_m/T_m$ 45°C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>71.8</td>
<td>72.7</td>
<td>11.0</td>
</tr>
</tbody>
</table>

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

*Calculated as $JT_m/(JT_m + JT_m')$. All cloned repetitive sequence families for which $JT_m$ was determined, except clone 2007, should be represented in the short repetitive DNA preparation obtained in the inclusion peak of the CL-Sephacryl column after 51 nuclease digestion. For purposes of comparison to $JT_m$, clone 2007 was therefore excised from this calculation. Roman numerals indicate the degree of intraspecies sequence divergence according to the arbitrary classification scheme described in the text.*

* $JT_m$ 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 16s-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 °C + 4 °C = 80/300 = 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° + 60/300 = 73.8 °C. Thus $JT_m$ = 84.8° − 73.8° = 11.0 °C.*

*Clones 2057A and B were not completely analyzed. From the absence of any low thermal stability fractions in the thermal elution profiles (data not shown), however, it is clear that the $T_m$s 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 $JT_m$ could not be calculated. The high $T_m$s 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
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 Tm 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 3l (clone 2101). The values of Tm, the difference between the Tm of the heteroduplex population and the Tm of the double-stranded cloned fragment, lies between 5.7 and 10.6°C for the seven class II families (Table 3). The weight mean value of Tm for these families is ~4°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 non-reacted, 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 Tm. 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, $\hat{a}$, 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, $\bar{a}$, 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 $2\bar{a}$. For a total repeat fraction such as that shown in Figure 3a, $\overline{\Delta T_m} =$ 50 gives the weight average of the individual values of $2\bar{a}$ 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 $\Delta T = C + \bar{a}$. When many different clones are studied, the average value of $\Delta T$ should, of course, approach the weight average value of $2\bar{a}$ 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 all different 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, $2\Delta T$, 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 $\Delta T$, listed in the last column of Table 3 should approximate the mean intramolecular divergence in the genome if the clones studied are a representative set.

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 present 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 \( \Delta T_c \) values for the cloned repeat families is 10.2°C. As shown above, the weight mean for \( \Delta T_c \) should approximate the measured value of \( \Delta T_{meas} \), 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 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 experiments of Figure 3a were prepared using S1 nucleases. We rely on our earlier studies which show that under appropriate conditions, S1 nucleases 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 reproductive 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 replicates. Repeat families distributed throughout the genome in 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.

Restrain 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 Procedure**

**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-EX1 conditions, as specified in the NIH Guidelines.

**Preparation of Cloned Repetitive rDNA.**

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 CaCl2-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 µl Tris-HCl (pH 7.8), 100 µM NaCl and 5 mM MgCl2. 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-Savag solution (1:1) (Savag solution is chloroform-isoamyl alcohol 24:1 and isopropyl or with ether. Dephosphorylated Eco RI-cleaved DNA was labeled at the 5' termini with γ-32P-ATP essentially as described by Maxam and Gilbert (1977). DNA was dissolved in 5 mM Tris (pH 9.5), 0.61 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 MgCl2 and 5 mM dithiothreitol. γ-32P-ATP was dried down, dissolved in H2O 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-Savag (1:1) and once with ether and ethanol-precipitated. Specific activities ranged from 5 x 106 to 2 x 107 cpm µg-1, depending upon the length of the cloned fragment and the efficiency of the individual kinase reactions.

The 5'-32P-labeled sea urchin DNA fragments were isolated from the plasmid DNA in one of two ways. For clones 2057A, 2057B, 2090 (preparation a), 2098, 2099, 2096, 2100, 2102, 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 5 min and then equilibrated to >10 x C0141 for each fragment. The renatured DNA was then added at or less than a 3% agarose gel and subjected to electrophoresis at 50 V for 3-3.5 hr in 30 mM Tris, 20 mM sodium acetate, 10 mM EDTA (pH 7.6) at room temperature. The 5'-32P-labeled DNA fragments were excised from the gel and eluted by dissolving the agarose in 5 M NaClO4, 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 2055, 2057, 2053, 2055, 2058, 2090 (preparation b), 2102, 2108 (preparation b), 2109A, 2106B, 2112, 2125, 2131, 2133A, 2132B and 2137, the labeled, dried precipitates were dissolved in 0.3 M NaOH, 1% 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 6-9% gels at 80 V in 50 mM Tris-borate buffer (pH 8.3), 1 mM EDTA at 37°C for 1 hr. After electrophoresis, gels were stained with ethidium bromide and autoradiographed. The electrophoretic conditions used were essentially the same 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 Eppendorf 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 2050 (preparation b) and 2108 (preparation b), strand-separated DNA was
used for the thermal stability measurements. The strand-sep-
arated DNA was concentrated 4-fold by extension with 2-butanol
to approximately 0.5 M phosphate buffer. The HP-DNA was then
incubated for 10 hr at 55°C to allow any contaminating comple-
mentary strands to renature. The mixtures were then diluted to
0.12 M phosphate buffer and passed over 1 ml of hydroxyapatite
at 30°C in 0.12 M phosphate buffer. The nonbound fraction
contained single-stranded fragments. These fragments were con-
taminated from 0-5% with their complementary strands and up to
30% unincorporated γ32P-ATP. The labeled DNA was stored at
−20°C after dialysis versus 0.12 M phosphate buffer. 0.1% SDS.

Length Measurements of Cloaked Repetitive Sequence Elements

Lengths of cloned DNA fragments were determined by electo-
phoresis on 3% agarose gels, by comparison to DNA length
standards. An Hae ill digest of RS2124 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 spherized 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
two preparations had weight average lengths of 450
nucleotides and one had a weight average length of 650 nucleo-
tides as measured on alkaline sucrose gradients (Davison et al.,
1975).

Preparation of Single-Copy H-DNA

H-DNA was extracted from sea urchin embryos grown in H-
thymidine and was used to prepare single-copy H-DNA as de-
scribed by Gaitus, Britten and Davidson (1974). The specific
activity of the single-copy H-DNA was 2 × 108 cpm μg−1 under
our culturing conditions.

Reactions with Cloaked Repetitive HP-DNA

Reaction mixtures varied considerably with regard to amounts of
cloaked repetitive HP-DNA and sea urchin DNA. In general, 100–
500 cpm of nonstrand-separated HP-DNA and 2–30 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 1.0–10.0 μg of single-copy H-DNA. After denatur-
ing at 97°C, the reactions were carried out in sealed glass ampoules
or in microcapillaries at 55°C. Samples were reacted to Cu values
ranging from 10–100 μM Cu; reaction times were from 30 sec
to 24 hr. In a few cases for high Cu reactions, 0.41 μM phosphate
buffer, 0.05% SDS were used. These Cu values and all other Cu
values referred to in this work are equivalent Cu (i.e., the Cu
of the reaction corrected for the relative increase in rate due to
cu concentrations above 0.18 μM Cu). Reactions were diluted to
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 95°C. In all
but four cases, the sequence ratios of drider to cloned HP-DNA
exceeded 20, so reaction of the tracer with itself was negligible.
Clones 2086, 2096, 2097 and 2108, diet crude to their low renaturation
frequencies, were driven by 2–4 fold sequence excess of sea urchin
NA. When corrections are applied for reaction of the HP-
DNA with itself, however, no appreciable effect on the renaturation
frequency is observed in these cases.

Thermal Chromatography

Reactions with Sea Urchin DNA

Reaction mixtures generally contained 200–2000 cpm of non-
strand-separated cloned repetitive HP-DNA and 200–800 μg of
sea urchin DNA. In two cases (2108 and 2096), only one strand
of the cloned HP-DNA was used. The reactions were all carried out
to kinetic termination (>10 × the Ct). The mixtures were then
diluted into 3 ml of 0.12 M phosphate buffer, 0.05% SDS. 1000–
2000 cpm of renatured single-copy 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 53°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
column and the relative elution rates, the rate of temperature increase.
The single-copy tracer melting profiles were relatively reproducible. The mean Tm for all of the
single-copy standards was 77°C, with a standard deviation of
at least 0.3°C. The TmS measured for the single-copy standards
were used to correct the experimental TmS to a uniform standard
melting condition. The magnitude of these corrections was usu-
ally only a few tenths of a degree C and never more than 1.3°C.

Thermal Chromatography of Native Cloaked HP-DNA

For thermal chromatography of native HP-DNA fragments, the
eution procedure followed was as described above. Where the
cloaked repetitive HP-DNA was renatured, 10× 107 cpm of the HP-
DNA were reacted in 0.5 M phosphate buffer, 0.05% SDS at 67°C
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 er-
ors if very poorly matched duplexes (for example, 25–30% mis-
matched) fail to bind even though they are not single-stranded.
To test whether poorly matched duplexes bind to hydroxyapatite,
we reacted S. franciscanum DNA with S. purpuratus single-copy
HP-DNA in 0.5 M phosphate buffer, 0.05% SDS at 50°C to Cot
20,000. This heterologous reaction is known to produce mis-
matched duplexes; the TmS between the homologous S. purpur-
tatus reaction and the heterologous S. franciscanum-S. purpur-
tatus reaction is approximately 10°C (Anger et al., 1976; Harpold
and Craig, 1977; T. J. Holz, E. H. Davidson and R. J. Britten, unpub-
lished 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 H-DNA passed through the column, approximately as
expected for the heterologous reaction at this creation and Cot.
The bound HP-DNA displayed a broad thermal stability profile,
clearly limited by the reaction criterion. The apparent Tm was
~64°C, and 30% of the HP-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 HP-DNA eluted 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 H-DNA sample consisting of the same single-copy
HP-DNA plus S. franciscanum driver DNA. The digestion kinetics
(Britten et al., 1978) 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 HP-DNA and 0.075 of the H-DNA not bound to hydrox-
apatite were resistant to digestion. Thus the maximum fraction of
the duplexes present which was not bound on hydroxyapatite at
45°C was:

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

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


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 calculates third-order polynomials through the data points and interpolates 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 3a-3f. Tm from the polynonal 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 Leukemia Fellowship from the American Cancer Society, California Division, and T.L.T. is the recipient of an NIH postdoctoral fellowship. H.S. is a postdoctoral fellow on an NIH training grant. The PE facilities used in this work were equipped with the aid of funds from a Biomedical Research Support Grant.

The cost of publication of this article was 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

CHAPTER 5
Repetitive Sequence Transcripts in the Mature Sea Urchin Oocyte

Franklin O. 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. 4H-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 (>10^10 copies per oocyte), while others are rare (~10^6 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 ~2000 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 repetitive 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 Gilson, Gilson 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
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., 1978), 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 3H-DNA preparations. The reactions were performed under relatively low criterion conditions (0.4 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 31 Nucleases-Resistant Repetitive DNA](image)

Total sea urchin 3H-DNA was sheared to an average single-strand length of 3300 nucleotides, renatured to Cot 40 and digested with 31 nucleases to remove single-stranded regions. The 31 nucleases-resistant doublet fraction (21%) was recovered on hydroxyapatite and chromatographed on a Sepharose CL-2B gel filtration column in 0.12 M PB at 60°C (b). Fractions were collected and an aliquot of each fraction was counted. DNA in the excluded peak (fractions 8-12) was pooled and chromatographed (C), and again the excluded fractions (7-12) were pooled. This material constituted the long repetitive DNA fraction. Similarly, the included peak material (fractions 18-23) was pooled and chromatographed (D), 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 3H-orthophosphate. The x<sub>0,1</sub> of 300 nucleotides long native DNA on this column is 0.74.
The short repeat \(^{3}H\)-DNA (2) and long repeat \(^{3}H\)-DNA (9) tracers were prepared as described in Figure 1 and the text, and hybridized with excess oocyte RNA at 55°C in 0.1 M PB, 0.2% SDS, 0.005 M EDTA. The fraction of \(^{3}H\)-DNA in RNA-\(^{3}H\)-DNA hybrids (ordinate) is shown as a function of RNA:DNA mass ratio at 55°C. The fraction was measured by binding to hydroxyapatite at 50°C in 0.12 M PB, 0.05% SDS. Since the single amount of tracer binding due to RNA-DNA duplex formation (<2%) at the highest RNA:DNA mass ratio 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 10:1 in all reactions; no reactions to high RNA:DNA mass ratio used was used to prevent tracer self-react. 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 by dimethyl sulfoxide at 55°C prior to hybridization (8). These data are indistinguishable from the data obtained after the usual procedure of 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 (ki) of 1.06 \times 10^{-10} M^{-2} sec^{-1}, 21% hybridizes with a ki of 2.48 \times 10^{-10} M^{-2} sec^{-1}, and 33% hybridizes with a ki of 9.98 \times 10^{-12} M^{-2} 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 ki of 2.48 \times 10^{-10} M^{-2} sec^{-1}, and 19% with a ki of 7.68 \times 10^{-10} M^{-2} 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 (ki) 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 (ki) of 2.3 \times 10^{-10} M^{-2} sec^{-1} (Hough-Evans et al., 1977). The scale for this reaction is shown on the right-hand ordinate.

Figure 2 shows that the long repeat tracer (closed circles) hybridizes significantly less than the short repeat tracer. The RNA/\(^{3}H\)-DNA mass ratio used in the high RNA:DNA 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:DNA mass ratio 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 $3 \times 10^4$ 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 $6 \times 10^6$ nucleotide pairs. Highly prevalent oocyte RNA transcripts such as histone mRNA and rRNA clearly cannot account for the order 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 (Schellier 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 RSP2124 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 Schellier 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 Cot1, 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 Schellier 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. Schellier 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 2108B (open circles), since its reaction occurs most rapidly. On the basis of the rate of the reaction of the clone 2108B upper strand repeat fragment, we estimate that there are $-8.3 \times 10^4$ 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**

<table>
<thead>
<tr>
<th>Clone</th>
<th>Strand*</th>
<th>Length* (nt)</th>
<th>Genomic Repetition Frequency*</th>
<th>Second-Order Rate Constant* (k₂)</th>
<th>Transcripts per Oocyte*</th>
<th>RNA Excess Kinetics</th>
<th>Titrations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>k₂</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td>U</td>
<td>1100</td>
<td>400</td>
<td>1.5 x 10⁻⁸</td>
<td>7000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2034</td>
<td>L</td>
<td>560</td>
<td>1000</td>
<td>3.6 x 10⁻⁷</td>
<td>3000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.6 x 10⁻⁷</td>
<td>5300</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2090</td>
<td>U</td>
<td>220</td>
<td>140</td>
<td>1.7 x 10⁻⁸</td>
<td>35000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.7 x 10⁻⁸</td>
<td>57000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2101</td>
<td>L</td>
<td>320</td>
<td>700</td>
<td>2.5 x 10⁻⁷</td>
<td>5400</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.9 x 10⁻⁷</td>
<td>3000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2106</td>
<td>U</td>
<td>190</td>
<td>20</td>
<td>3.6 x 10⁻³</td>
<td>79000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.3 x 10⁻³</td>
<td>27000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.6 x 10⁻⁴</td>
<td>160000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2106A</td>
<td>U</td>
<td>200</td>
<td>900</td>
<td>1.6 x 10⁻⁴</td>
<td>2500</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.0 x 10⁻⁴</td>
<td>2800</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2109B</td>
<td>L</td>
<td>125</td>
<td>1000</td>
<td>6.0 x 10⁻³</td>
<td>65000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.0 x 10⁻⁴</td>
<td>82000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2123B</td>
<td>U</td>
<td>310</td>
<td>60</td>
<td>4.1 x 10⁻⁷</td>
<td>8600</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.4 x 10⁻⁷</td>
<td>5600</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2137</td>
<td>L</td>
<td>190</td>
<td>530</td>
<td>9.0 x 10⁻⁴</td>
<td>19000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.6 x 10⁻⁷</td>
<td>18000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.3 x 10⁻⁴</td>
<td>16000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

1 The two complementary strands of each cloned repetitive sequence were designated "upper" and "lower," according to their electrophoretic mobility on polyacrylamide gels after denaturation (Schierer et al., 1979). The lower strand is the strand which migrates more rapidly.

2 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).

3 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.

4 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).

5 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 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 Schierer et al. (1978). Insertion of the appropriate numerical values yields:

\[
\text{transcripts per oocyte} = \frac{k_2}{k_1} \times \frac{L}{1800}(L)
\]

where \(k_1\) is the observed rate constant for the hybridization of the cloned DNA sequence with excess oocyte RNA from column 5 of the table, \(2.3 \times 10^{-11} \text{ M}^{-1} \text{ sec}^{-1}\) 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 1800 is the average number of RNA transcripts of each single-copy sequence per oocyte, determined by Hough-Evans et al. (1977). \(k_2\) is a factor which corrects for the rate retardation observed when the length of the oocyte (the RNA in this case) exceeds the length of the transcript. \(k_2\) is approximately equal to \((L_{\text{transcript}}/L_{\text{DNA}})^{1/2}\) (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 \(k_2\) of 3.0 has been applied. For sequences whose complementary transcript length has not been measured, we use a conservative value of \(k_2 = 2\). An additional minor kinetic uncertainty derives from the sequence mismatch with characteristics 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 Schierer et al. (1978). The intrastrialal sequence divergence of these cloned repetitive sequence families was measured by Klein et al. (1978), and is listed in Table 2 of Schierer 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

\[
\text{transcripts per oocyte} = \frac{k_{\text{transcribed}}(3 \times 10^{-4})(0.03 \times 10^{10})}{k_{\text{transcribed}}(350)(L)}
\]
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:32P-DNA ratio used was insufficient to permit complete reaction.

Titrations 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 1.6 × 10^6 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 apparent incomplete 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 reactivity 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 2106, 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 clustered 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 sequence 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 repeat 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 RNA.
Oocyte Repetitive Sequence Transcripts

175

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

The separated strands of cloned repetitive sequences were hybridized with a 1-5 x 10^6 molar excess of oocyte RNA at 55°C in 0.4 M PB, 0.2% SDS, 0.005 M EDTA. The cloned tracers were terminally labeled with ^32P. Hybridization was assayed by binding to hydroxylapatite at the indicated values of RNA:DNA. The observed hydroxylapatite binding was 90% 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) is 2.3 x 10^-5 sec^-1 (Houch-Evans et al., 1977). Clone 21096 upper strand (C) hybridizes with a second-order rate constant (k) of 8 x 10^-14 M^-1 sec^-1; for clone 2090, upper strand (C) and lower strand (A), k = 1.7 x 10^-14 M^-1 sec^-1; Clone 2103 upper strand (C) and lower strand (B) hybridize incompletely at the oocyte RNA:DNA mass ratios attainable in these experiments because the tracer is in sequence excess (see text). The observed fraction of ^32P-DNA bound to hydroxylapatite was normalized by the fraction of ^32P in each cloned DNA preparation which is capable of reassociating with excess sea urchin DNA (that is, the tracer reactivity). This fraction ranged from 0.8 to 0.8 for the tracers shown. The nonreactive ^32P was mainly γ-ATP persisting from the poly-Nucleotide 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 fraction 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 unfractonated 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 x 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 unfractonated 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 the 96% of the selected repeat tracer which is reactive, 19% consists of repeats occurring only 20 times per haploid genome, and 39% 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 x 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
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 38S; RNA in size class II sediments between 11S and 19.5S; RNA in size class III sediments 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 210B repeat, as with prevalent transcripts, such as those complementary to the clone 210B 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 x 10^{-7}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 67% bound. This second bound fraction was eluted, and the \( ^{3}H \)-DNA was purified of RNA by pronase 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 hydroxyapatite (C) 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 79 x 10^{-3} 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 8. 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 (8), or that fraction of the short repeat \( ^{3}H \)-DNA tracer selected by low Cot hybridization to oocyte RNA (C) (see Figure 5), was reacted with a 1-5 x 10^{-8} fold mass excess of total sea urchin DNA that was sheared to a weight mean single-strand fragment length of 850 nucleotides. The ordinates show the fraction of \( ^{3}H \)-DNA binding to hydroxyapatite at 50°C in 0.12 M PB, 0.08% SDS. The Cot is calculated with respect to the concentration of the genomic DNA. Included as an internal kinetic standard in each reaction mixture was a \( ^{3}P \)-labeled, cloned repetitive sequence (clone 2034; data not shown). The rate of reassociation of this \( ^{3}P \)-DNA sequence with the driver DNA was for all reaction mixtures consistent with the 1000-fold renaturation frequency measured previously for this sequence (Klein et al., 1978). The solid line shows a least-squares solution for the reaction of the unfractonated short repeat tracer, assuming three second-order kinetic components. This solution uses the three rate constants measured by Graham et al. (1974) for the repeated components of whole 3. purpurea DNA: 8.2 M^{-1} sec^{-1} for the fastest component (8500 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, 0.32, and 0.001% of the \( ^{3}H \)-cpm are nonreactive and 2% bind at Cot <10^{-4}. 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 \( ^{3}P \)-cpm are nonreactive).

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 x 10^{-4}. After this procedure, its weight mean single-strand fragment length was 2600 nucleotides. The tracer was
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 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.

hydroxyapatite in the absence of added RNA is due to minor contamination of the lower strand of the clone 2050 repeat fragment with the complementary upper strand. The increase in $^{3}P$-DNA binding to hydroxyapatite with added RNA is due to the formation of oocyte RNA-$^{3}P$-DNA hybrids and is labile to low salt RNAase treatment. The graph shows the relation between the amount of $^{3}P$-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 2000 repeat in each size class is calculated from the mass of $^{3}P$-DNA hybridized by a given volume of RNA solution in the linear region of the graph (that is, when the $^{3}P$-DNA is in large excess). The chart 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 (18S μg), the fraction of each size class used in the experiment and the amount of RNA per oocyte ($3 \times 10^{14}$ g). The number obtained is 37,000 complementary transcripts per oocyte.

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 $^{3}P$-DNA had a specific activity of $6.4 \times 10^{6}$ 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.

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 $^{3}P$-DNA (spec. act. $8 \times 10^{5}$ 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 300,000 per oocyte.

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, 1.3 or 3.5 μl of each RNA size class was used. Each hybridization contained 900 cpm of $^{3}P$-DNA (spec. act. 3.2 $\times 10^{5}$ cpm/μg). The total number per oocyte of transcripts complementary to the clone 2101 upper strand tracer was 9000.
Oocyte Repetitive Sequence Transcripts

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

A 3H-DNA tracer of weight average single-strand fragment length 2600 nucleotides, from which the fraction binding to hydroxyapatite at Cot $5 \times 10^{-6}$ (20%) had been removed, was hybridized with a $10^6$-fold excess of oocyte RNA. The fraction of 3H-DNA bound to hydroxyapatite has been corrected, as described in Experimental Procedures, for any self-reaction of the DNA (35%). 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, $F_3$, was calculated at various values of RNA Cot from the data in Figure 3, using the relation $F_3 = (F_0 - F_0^*)/(F_0 - F_0^*)$, where $F_0$ and $F_0^*$ are the fractions of short and long repetitive DNA hybridizing with oocyte RNA at a given RNA Cot. and $F_0$ and $F_0^*$ are the fractions of whole sea urchin DNA consisting of short and long repetitive sequences (0.13 and 0.86, respectively).

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-
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. (1979) demonstrate that complementary repeat transcripts also exist in nuclear RNAs. They also serve 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 matRNAs molecules. From the data of Figures 2 and 8, 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 maturing RNA 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 maturing RNA 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 isolated in vivo by a modification of the method of Kleene and Humphreys (1977). Strongly-sedimented purpuratus embryos were cultured at 13°C in Millipore-filtered seawater containing penicillin and streptomycin (Memminger, 1967; Smith et al., 1974). At the 8-cell stage (~4.5 hr after fertilization), embryos were detached and resuspended at 7 × 10⁴ embryos per ml with 50 μg/ml hydrazine (90 Cl/meromela), 50 μg/ml 2′-3′-adenosine (125 Cl/meromela) and 26 μg/ml 5,6-uridine (35 Cl/meromela). 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 (12 hr after fertilization) then harvested by centrifugation. They were washed once in seawater and sonicated in 0.1 M NaCl, 0.05 M EDTA (pH 8.0) and then resuspended in a few milliliters of EDTA. The resulting slurry was dropped from a Pasteur pipette onto a bed of powdered dry ice, so that the drops froze immediately. The embryo-dry ice mixture was blended into a Waring blender to a fine homogeneous powder. The powder was then suspended in sonicated 10 ml of 10% normal embryo. 0.5% SDS was added with stirring to complete lysis, and DNA was isolated by standard procedures including phenol-chloroform-acetic acid extraction, washing, and Rnase A and pronase digestion (Graham et al., 1974; Angeier, Davidson, and Britten, 1975). The DNA obtained had a specific radioactivity of 1.15 × 10⁸ cpm/μg.

Preparation of Long and Short Repetitive DNA Fractions

Sea urchin H4-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 a Chelsea 100 (Bio-Rad) to remove any contaminating heavy metal ions. Following denaturation with 0.1 M NaOH for 10 min at 25°C, the DNA was resuspended in 0.14 M sodium phosphate buffer, 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 50°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, 1972) (10 μg/mg of DNA) in 0.15 M NaCl, 0.025 M Na acetate, 0.005 M PIPES, 0.1 mM ZnSO₄, 0.002 M β-mercaptoethanol (pH 4.4). The S1 nuclease conditions correspond to a G to I of 0.8, as defined by Britten et al., 1975). This amount of digestion is sufficient to remove single-stranded regions without destroying most repetitive sequence duplicities (Britten et al., 1976; Eden et al., 1977). Phospho-buffer was added to 0.12 M and the DNA was extracted once with IAC (24:1 chloroform:scalenal 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.

The 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 n-butanol (Stafford and Bieger, 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 iodinated with a Branson Model 5125 sonifier and microspin for 6 × 20 sec at 2 K rpm in 0.12 M PB. The resulting fragments averaged 570 nucleotides in length, according to measurements made by position sedimentation in alkaline sucrose gradients. To remove any contaminating single-copy DNA and unreproducible DNA fragments, the DNA was deproteinized and reextracted 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 (80%) was eluted with 0.5 M PB and constitutes the long repetitive H4-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 repeat comprises about 13% of the DNA, this is comparable to the whole DNA Cot of about 150). Long and short repeat fractions were isolated from unseeded sea urchin DNA in the same fashion.

DNA-DNA Resequencing

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. Resequencing mixtures also contained 2.5 mM EDTA and 0.1-10% SDS, and some included cast rhodium DNA as carrier. The mixtures were sealed in siliconized glass capillaries, boiled at 95°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 columns on hydroxyapatite at 0.12 M PB, 0.05% SDS at 50°C. Bound fragments were eluted from the column at 98°C, and fractions were counted by liquid scintillation in Hankfioril (Masukocho) or assayed for absorbance at 260 nm. Bio-Gel HTP DNA-grade hydroxyapatite, lot 415535 (Bio-Rad), was used throughout this study. All values of DNA or RNA Cot (DNA or total RNA concentration x time, in units of moles of nucleotides per liter x 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).

DNA-RNA Hybridizations

Incubation conditions and assays for DNA-RNA 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 Gasco 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 the 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 32P-DNA binding to hydroxyapatite, Fₚ, is the sum of two, the fraction of 32P-DNA containing an RNA-DNA hybrid, and Fₚ, the fraction of 32P-DNA containing a DNA-DNA duplex. To determine Fₚ, the second aliquot in 0.02 M PB was treated with 50 μg/ml RNase A at 37°C for 1 hr, a condition sufficient to digest the RNA in RNA-DNA hybrids, leaving only the DNA-RNA 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 H4-DNA binding to hydroxyapatite after the RNAase digestion s fH4, which was usually between 0.01 and 0.1. The fraction of H4-DNA containing a RNA-DNA hybrid, s fH4, was calculated as fH4 = (fH4 - fH4) / (fH4 - fH4), where fH4 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 streaked 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 32P-Labeled Repetitive DNA Sequences
The purification, labeling and strand separation of cloned repetitive sea urchin DNA sequences are described in the accompanying paper by Schiefer et al. (1978).

RNA Fractionation of DMSO-Sucrose Gradients
Total oocyte RNA (175 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 8.5). This sample was heated at 55°C for 5 min, then diluted to 0.4 ml with 0.1 M LiCl, 0.005 M EDTA, 0.2% SDS and 0.01 M Tris-HCl (ph 8.5). One third of this sample was layered onto each of three sucrose gradients (40-20%) containing 90% (v/v) DMSO, 0.1 M LiCl, 0.02 M Tris-HCl (ph 8.5), 0.005 M EDTA and 0.05% SDS. After centrifugation for 65 hr 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 8.5) 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 7°C.

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


Specific Representation of Cloned Repetitive DNA Sequences in Sea Urchin RNAs

Richard H. Scheider
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 unfertilized sea urchin eggs, 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 Deragoner, 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, 1975). 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 postdoctoral fellow 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 (Schellek 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 concentration for each clone, 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 2004, belong to families displaying very little intramolecular 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 2007 and 2004 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).

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 \(~2 \times 10^6\) 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

<table>
<thead>
<tr>
<th>Clones*</th>
<th>Genomic Retention Frequency*</th>
<th>Length* (NTP)</th>
<th>Approximate Base Composition*</th>
<th>Estimated Intramiatal Divergence* (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>400</td>
<td>1100</td>
<td>7</td>
<td>3.6</td>
</tr>
<tr>
<td>2034</td>
<td>1000</td>
<td>560</td>
<td>25</td>
<td>3.7</td>
</tr>
<tr>
<td>2090</td>
<td>140</td>
<td>220</td>
<td>38</td>
<td>&gt;20.5</td>
</tr>
<tr>
<td>2101</td>
<td>700</td>
<td>320</td>
<td>20</td>
<td>7.5</td>
</tr>
<tr>
<td>2108</td>
<td>20</td>
<td>190</td>
<td>48</td>
<td>5.7</td>
</tr>
<tr>
<td>2109A</td>
<td>900</td>
<td>200</td>
<td>32</td>
<td>&gt;19.7</td>
</tr>
<tr>
<td>2109B</td>
<td>1000</td>
<td>125</td>
<td>44</td>
<td>&gt;25.3</td>
</tr>
<tr>
<td>2132</td>
<td>60</td>
<td>310</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2137</td>
<td>530</td>
<td>190</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Clones of this series constructed in this laboratory by blunt end ligation of restriction enzyme fragments. The Cs prefix is omitted in this paper.
* Estimated by radiometric labeling of repetitive DNA and hybridization to the labeled cloned tracer. (Klein et al., 1978).
* Values derived from a computer program, the program assumes that the base composition of a repeat family is equal to the base content of the sequence.

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. Gaal et al. (1978) and Wold et al. (1978) showed that the complexity of intestine polyosomal mRNA is only 50 x 10^6 nucleotides as compared with 17 x 10^6 nucleotides for gastrula. About half of the intestine mRNA sequence set is also represented in gastrula or blastula polyosomal 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 of 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, Tc, is given by:

\[ T_c = \frac{K_r}{K_w} \frac{T_{tr}}{I_c} \]  

In this equation, Ttr is the number of transcripts of an average single-copy sequence per nucleus, as cited above for each nuclear RNA; \( K_w \) is a first-order rate constant for the effect of the kinetics of disparity in tracer and driver length (Chamberlin et al., 1978); \( K_r \) is the pseudo-first-order single-copy hybridization rate constant; and \( I_c \) 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 Tc are listed in the legend to Table 2. The intramiatal 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, 37°C) provide a reaction criterion significantly below the Tm of the RNA-DNA DNA hybrids formed in these experiments (data not shown).

We know from the small intramiatal 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 lower 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 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>
<thead>
<tr>
<th>Clone</th>
<th>Strand</th>
<th>Transcripts per Nucleus by Kinetics</th>
<th>RNA Fraction</th>
<th>Represen-</th>
<th>Rate Constantb</th>
<th>Transcripts per Nucleus by Kinetics</th>
<th>RNA Fraction</th>
<th>Represen-</th>
<th>Rate Constantb</th>
<th>Transcripts per Nucleus by Kinetics</th>
<th>RNA Fraction</th>
<th>Represen-</th>
<th>Rate Constantb</th>
<th>Transcripts per Nucleus by Kinetics</th>
<th>RNA Fraction</th>
<th>Represen-</th>
<th>Rate Constantb</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>U</td>
<td>$1.4 \times 10^{4}$</td>
<td>3.2</td>
<td>3</td>
<td>$3.9 \times 10^{-4}$</td>
<td>31</td>
<td>39</td>
<td>7.000</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>$1.2 \times 10^{4}$</td>
<td>2.7</td>
<td>2</td>
<td>$4.2 \times 10^{-4}$</td>
<td>31</td>
<td>43</td>
<td>7.000</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2034</td>
<td>U</td>
<td>$1.1 \times 10^{4}$</td>
<td>0.9</td>
<td>1</td>
<td>$6.1 \times 10^{-4}$</td>
<td>48</td>
<td>48</td>
<td>3.6 x 10</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>$5.0 \times 10^{3}$</td>
<td>1.0</td>
<td>1</td>
<td>$3.9 \times 10^{-3}$</td>
<td>31</td>
<td>31</td>
<td>5.8 x 10</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2090</td>
<td>U</td>
<td>$3.3 \times 10^{4}$</td>
<td>3.0</td>
<td>1</td>
<td>$3.0 \times 10^{-4}$</td>
<td>47</td>
<td>53</td>
<td>1.7 x 10</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>$2.1 \times 10^{4}$</td>
<td>4.0</td>
<td>1</td>
<td>$1.5 \times 10^{-4}$</td>
<td>45</td>
<td>66</td>
<td>1.7 x 10</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2101</td>
<td>U</td>
<td>$5.4 \times 10^{3}$</td>
<td>1.1</td>
<td>1</td>
<td>$6.4 \times 10^{-3}$</td>
<td>51</td>
<td>73</td>
<td>5.8 x 10</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>$4.8 \times 10^{3}$</td>
<td>1.1</td>
<td>1</td>
<td>$3.8 \times 10^{-3}$</td>
<td>30</td>
<td>43</td>
<td>1.9 x 10</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2108</td>
<td>U</td>
<td>$1.2 \times 10^{4}$</td>
<td>1.1</td>
<td>1</td>
<td>$2.1 \times 10^{-4}$</td>
<td>55</td>
<td>5.8 x 10</td>
<td>79.000</td>
<td>58000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>$5.3 \times 10^{3}$</td>
<td>0.8</td>
<td>1</td>
<td>$5.9 \times 10^{-3}$</td>
<td>27</td>
<td>1.3 x 10</td>
<td>27.000</td>
<td>32000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2109A</td>
<td>U</td>
<td>$5.8 \times 10^{3}$</td>
<td>1.0</td>
<td>2</td>
<td>$2.0 \times 10^{-2}$</td>
<td>18</td>
<td>2.3 x 10</td>
<td>2.500</td>
<td>0.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>$5.8 \times 10^{3}$</td>
<td>1.0</td>
<td>2</td>
<td>$2.0 \times 10^{-2}$</td>
<td>2</td>
<td>2.0 x 10</td>
<td>2.600</td>
<td>0.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2109B</td>
<td>U</td>
<td>$3.5 \times 10^{4}$</td>
<td>0.4</td>
<td>1</td>
<td>$2.4 \times 10^{-4}$</td>
<td>18</td>
<td>1.8</td>
<td>6.6 x 10</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>$3.5 \times 10^{4}$</td>
<td>0.4</td>
<td>1</td>
<td>$1.1 \times 10^{-4}$</td>
<td>0.8</td>
<td>0.9</td>
<td>6.6 x 10</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2133B</td>
<td>U</td>
<td>$1.4 \times 10^{4}$</td>
<td>2.7</td>
<td>2</td>
<td>$8.7 \times 10^{-2}$</td>
<td>48</td>
<td>41 x 10</td>
<td>6.600</td>
<td>7.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>$5.5 \times 10^{3}$</td>
<td>1.0</td>
<td>1</td>
<td>$5.3 \times 10^{-3}$</td>
<td>0.29</td>
<td>3.4 x 10</td>
<td>5.600</td>
<td>5.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2137</td>
<td>U</td>
<td>$2.7 \times 10^{4}$</td>
<td>6.0</td>
<td>1</td>
<td>$7.4 \times 10^{-1}$</td>
<td>59</td>
<td>11</td>
<td>9 x 10</td>
<td>2.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>$2.9 \times 10^{3}$</td>
<td>1.2</td>
<td>2</td>
<td>$5.3 \times 10^{-2}$</td>
<td>42</td>
<td>7.9</td>
<td>7.8 x 10</td>
<td>2.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The second-order rate constants - that is, $k_1$ 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). 

* 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 $k_2 = 1.1 \times 10^{-4}$, $T_d = 1$; intestine $k_2 = 2.3 \times 10^{-4}$, $T_d = 0.1$; oocyte $k_2 = 3.8 \times 10^{-4}$, $T_d = 1$. The value of $k_2$ calculated from the data of Figure 6 for the clone 2137 fragment reacting with gastrula nuclear RNA was 4, for the clone 2137 fragment reacting with gastrula nuclear RNA the value of $k_2$ was 2, for all other reactions with gastrula and intestine nuclear RNA except those with the clone 2007 fragment, a minimum of 2 was applied - that is, assuming that the hybridizing RNA molecules are at least 4 times larger than the cloned fragment. Chamberlain et al. (1978) found that the rate of long driver short tracer reactions is established by a factor approximately equal to $(k_2/k_1)^n$, where $k_2$ and $k_1$ are tracer and driver fragment lengths. For the long clone 2007 fragment, $k_2$ was taken as 5. 

* Values for $k_2$ for oocyte reactions are those found in Table 1 of the accompanying paper by Costantini et al. (1978). 

* 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 average of the least-squares solution for (a) in these expressions. 

* 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: gastrula nuclear RNA, $Q = 2 \times 10^{-10}$ g; for intestine nuclear RNA, $Q = 1 \times 10^{-10}$ g; for oocyte RNA, $Q = 3 \times 10^{-10}$ g (Wilson, 1949; Howes et al., 1975; Wold et al. 1979, also our unpublished data). 

* Percentage of representation is calculated according to equation (3), using titration data where possible. 

* 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 hydroxypatite 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 YH-DNA tracer with excess intestine nuclear RNA under the same conditions, and the right-hand dashed line shows the reaction of the single-copy YH-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 hydroxypatite 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 (A) and gastrula (A) nuclear RNAs. The reactions were carried out in 0.5 M phosphate buffer at 55°C and assayed by hydroxypatite binding. The possible presence of RNA-DNA duplex was routinely monitored by measuring the binding to hydroxypatite 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 YH-DNA tracer with gastrula nuclear RNA (data from Hough et al., 1975; also our unpublished data) and with intestine nuclear RNA (data from Wod et al., 1978) are indicated in each panel. Pseudo-first-order rate constants for these reactions are: for intestine nuclear RNA, 2.5 \times 10^{-8} M^{-1} sec^{-1} (solid line), and for gastrula nuclear RNA, 1.1 \times 10^{-9} 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" 32P radiactivity in the annealing mixtures is mainly 32P-ATP persisting from the kinase reaction used for labeling the fragments. The intestine nuclear RNA was present at a 7 \times 10^6 fold mass excess, and the gastrula nuclear RNA at a 4 \times 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 \times 10^6 fold mass excess, and the gastrula nuclear RNA in a 5 \times 10^5 fold mass excess with respect to the 32P-DNA tracer. The lower termination with the intestine RNA is due to the fact that the sequence excess for the 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) Reactions 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 \times 10^6 fold mass excess, and the gastrula nuclear RNA in a 2.5 \times 10^5 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 (gastrula) 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, 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-separated 32P-DNA tracer with increasing amounts of RNA. As more RNA is added, a greater amount of the 32P-DNA is hybridized and can be bound to hydroxyapatite. The reactions were run to kinetic termination (that is, \( t_{\text{max}} \)) with respect to the excess partner—in this case, 32P-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 higher, 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 sequences 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):

\[
\frac{T}{T_n} = \frac{1}{1 - a/R}
\]  

(2)

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

\[
\frac{T}{T_n} = \left(\frac{1}{a}\right) R.
\]  

(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/Tc, 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 a1 are shown in Table 2. Given an estimate of (a), the number of complementary transcripts for each cloned strand per nucleus, Tc, is calculated in the following manner:

\[ T_c = \frac{Q}{a} \frac{1}{350L} N \]  

(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 Tc calculated by application of equation (4) to titration data are listed in Table 2, where they can be compared to the kinetic estimates of Tc.

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 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](image_url)

(A) Renaturation of the upper and lower strands of the clone 2109A fragment, as in Figure 1 (B). Failure of separated strands to react with themselves is also shown: 2109A U strand (C) and 2109A L strand (B). Data have been normalized to 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 traces was calculated to be \( 9 \times 10^5 \) 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 (C) and L (A) strands 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 \( 9 \times 10^6 \) in this experiment. The solid line indicates the kinetics of the reaction of gastrula nuclear RNA with a single-copy H-DNA tracer (data from Hough et al., 1975).

(C) Reaction of the same 2109A U (C) and L (A) strands with intestine nuclear RNA. Transcripts complementary to the two strands differ in concentration by at least 3 fold. The failure of renaturation 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 \( 5 \times 10^6 \). 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 H-DNA tracer (data from Wold et al., 1975).
intestine nuclear RNA, however, the upper and lower strand of the clone 210SA 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 210SA upper strand). The upper strand reacts completely. We estimate that there are 16-20 transcripts complementary to the upper strand of the 210SA fragment per intestine nucleus, but only about 2-6 transcripts complementary to the 210SA 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 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. (1976) 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 \( \% \text{representation} \):

\[
\% \text{representation} = \frac{T_r}{F_r} \times 100
\]

where, as above, \( T_r \) is the number of transcripts complementary to the cloned tracer per oocyte or per gastrula or intestine nucleus. \( F_r \) is the number of copies of a typical single-copy transcript in the oocyte or the nuclear RNAs and \( F_r \) 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 level 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 \( T_r \) by \( F_r \) normalizes for the different repeat families according to their size and dividing by \( F_r \) 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 (\( T_r \)) is approximately 1600 (Galau et al., 1976; Hough-Evans et al., 1977), while as stated above, the gastrula nuclear RNA, \( T_r \) is 1 or less, and in intestine nuclear RNA, \( T_r \) 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 \( \sim 20,000 \)), from about 1-600 for the gastrula nucleus (average \( \sim 100 \)) and from about 1-50 for an intestine nucleus (average \( \sim 10 \)). 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 the 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 10^4 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 \( T_r \) 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 cloned tracer in each RNA, while Figure 7A shows the representation observed for the nine clones 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

---

*Repetitive Sequence Representation in RNAs* 197
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 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 subtraction 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 impression 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 strands in a given repeat is that 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 of each member are represented 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 asymmetric at most of 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., 1975; 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 oocyte 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 mRNAs 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 sequences in polysomal mRNAs of sea urchin embryos (Goldberg et al., 1973), HeLa cells (Klein et al., 1974) or rodent cells (Campos and Bishop, 1974; Rabbits 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 observations of Costantini et al. (1978) that both strands of the cloned repeat transcripts 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 X 10^4 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 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 nonpolysonal 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 gastrular 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 polysonal RNA compared with its representation in oocyte RNA or gastrular 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 polysonal message, they are clearly associated with nuclear RNA, as shown by this and earlier work (for example, Darnell and Baint, 1970; Jelinek et al., 1973; Smith et al., 1974). A hypothesis which may be relevant to our present results is that hnrRNA 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 hnrRNA 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 hnrRNA molecules rather than with the genomic DNA. Fedoroff and Wall (1976) and Fedoroff, Wallauer and Wall (1977) showed that hnrRNA 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 hnrRNA 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, Fedoroff et al., 1977). In terms of the kinetics of the putative repetitive sequence reactions, the genomic and hnrRNA 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 diagrammatically 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](image)

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 hnrRNA 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 hatched 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 RNA 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 hnrRNA. 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 μl. At the resulting transcript concentration, the fraction of transcrits 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 2109B fragment—would require about twice the 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 mRNA 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.

Experimental Procedures

Preparation, Labeling and Strand Separation of Cloned Repetitive Sequence Fragments

Supercoiled plasmid DNA was isolated on CsCl gradients (Scheffler et al., 1977). The DNA was dissolved 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 F. Green and H. Boysen per 20 μg of DNA at 37°C for 30 min in 100 mM NaCl and 5 mM MgCl₂. The reaction mixtures were precipitated with 1 vol of acetone at −20°C for at least 2 hr followed by centrifugation. The precipitated DNA was redisolved in 10 mM Tris at a concentration <50 μg/ml, and 10 μl of bacterial alkaline phosphatase ( Worthington, Code SAP1) 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-chloroform-isoamyl alcohol solution and once with ether. The DNA was precipitated with isopropanol as above. DNA fragments were isolated by the phenol-chloroform-isoamyl alcohol 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 MgCl₂, and 5 mM DTT. γ²⁵⁰-ATP, synthesized by the exchange procedure (Maxam and Gilbert, 1977), was dried down, dissolved in H₂O, and added to the reaction mixture at a concentration not lower than 3 μM. 3-5 μl of polynucleotide kinase solution (Pharmacia) were added, and the reaction was incubated for 45 min at 37°C and then extracted with phenol-Saevag 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 at 37°C for 10 min. Samples were strand-separated by electrophoresis on 1x 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 at 50 mA using 0.5 M Tris borate (pH 8.3), 1 mM EDTA at 15°C for 16 hr at 50 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 1.5 ml Eppendorf tube. 1 ml of 0.12 M sodium phosphate buffer (PB), 0.05% SDS and 10 μl 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 0.2 μl PB and passed over a 1 ml column of hydroxyapatite at 50°C. The unbound fraction contained the purified strand-separated repetitive sequence. γ²⁵⁰ specific activities of these materials ranged from 3 x 10⁷ to 3 x 10⁸ cpm/ml, 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 55-99%. The "nonreactive" radioactivity was shown to be of low molecular weight and is almost certainly γ²⁵⁰-ATP persisting from the kinase reaction. The γ²⁵⁰-DNA was stored in 0.12 M PB at −20°C. DNA reactivities

Figure 5. Strand Separation of Cloned Repetitive Sequences by Polyacrylamide Gel Electrophoresis

Supercritical 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 γ²⁵⁰-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). The gel was 6% 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 faster 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 2040. These are 560 and 190 nucleotides long, respectively. The 560 nucleotide fragment was stranded on a 4.5% gel and the 190 nucleotide fragment on a 6% gel, respectively.
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 studied 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 DNA complexes could not be compared to the complementary strands had been enzymatically released together; specific activities were expected to be equal and were found to be so (see Figure 4C of the accompanying paper by Costantini et al., 1979). The absolute specific activity could be calculated from the intensity of the reactions between each pair of complementary strands, given the cloned repeat fragment length (that is, the fragment com生态文明). The intensity of the reaction for the reaction with total sea urchin DNA determines the reaction 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 2106 are shown in Figure 1. The reactions between complementary strands terminate at about 90% and follow second-order kinetics. The same terminations are observed when the cloned fragments are reacted with excess urchin DNA (Klein et al., 1978). The incomplete (100%) reaction of the tracer is due to the persistence of a minute fraction of the 32P-ATP originally present in the kinase labeling reactions.

**RNA Preparations**

**Sagittal Nuclear RNA**

1.3 x 10^8 Strongylocentrotus purpuratus eggs were suspended in three liter jars of seawater containing penicillin and streptomycin (Gibco, 1974) and fertilized. After 2.5 hr, 95% of the eggs were at first cleavage, 3.5% were at second cleavage and 3.7% were uninfertilized. After 36 hr (early gastrula stage), the embryos were washed in cold Ca-Mg-free seawater, pH 7.5, and collected by centrifugation at 5000 rpm for 5 min. The embryo pellets were resuspended in 200 ml of cold 2 mM MgCl2, 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 treated with 7 M urea and extracted with phenol-cresol-Senag solution (0.15/0.65/1.0). The interface was reextracted and combined with the aqueous phase, which was reextracted twice with an equal volume of Senag 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, then resuspended in 10 mM NaOH and dialyzed against 0.1 M NaCl. The RNA was dialyzed in 5 mM sodium acetate (pH 5.8) at 20°C. Yields from a preparation of this size ranged from 25-35 μg. The complexity and pseudo-first-order rate constants obtained in reactions of the nuclear RNA with yeast RNA polymerase 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 treated with 0.5 M sucrose cushion and the RNA was extracted as above.

**Blastula Polyoma RNA**

Blastula polyoma RNA was prepared from 24 hr swimming blastulae as described by Gaisau et al. (1978).

**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 NaCl). The mixtures contained 1 mM EDTA, 0.05% SDS and calf thymus DNA carrier. RNA-DNA ratios were 5 x 10^4 to 1 x 10^7 for reactions in which RNA excess kinetics were to be measured. Titrations were carried out in 88P-DNA sequence excess—that is, at lower RNA/DNA ratios. Reaction mixtures were sealed in 5-50 μl capillaries and boiled at 103°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 Hought and Davidson (1972) and Gaisau, 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 RNA A incubated for 1 hr at 37°C, extracted with Senag 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. Five columns of hydroxyapatite. Six 2 ml washes were collected at 30°C and three 2 ml washes were collected at 98°C. The fractions were counted by liquid scintillation, and the fraction of radioactivity eluted at 98°C represented the fraction of DNA in hybrid structures. The DNA-DNA doublet content was always very small (<5%) compared with the DNA-RNA hybrid portion of the reaction.

Tirancon 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 Beal. This research was supported by a grant from the NSF N-V.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**


Receptive Sequence Representation in RNAs

203


CHAPTER 7
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 PBS2124, labeled, strand-separated and reassociated with 800–900 nucleotide long unlabelled 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 reassociated 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 eu-
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 for 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 nuclelease 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 purposes 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 reassocciated 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 thermal stability when reassociated with S. franciscanus DNA is only 0.8°C lower than when reassocciated with S. purpuratus DNA. The Tm 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 Tm 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 Tm, 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 interspecies 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 interspecies 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

<table>
<thead>
<tr>
<th>Clone</th>
<th>Native T_m</th>
<th>Driver DNA</th>
<th>T_m (50°C)</th>
<th>S. purpuratus T_m</th>
<th>S. franciscanus T_m</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS2007</td>
<td>72.7</td>
<td>S. purpuratus</td>
<td>48.5</td>
<td>5.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>S. franciscanus</td>
<td>64.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS2034</td>
<td>79.4</td>
<td>S. purpuratus</td>
<td>76.0</td>
<td>5.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>S. franciscanus</td>
<td>70.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>L. pictus</td>
<td>57.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS2090</td>
<td>83.1</td>
<td>S. purpuratus</td>
<td>64.5</td>
<td>-1.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>S. franciscanus</td>
<td>65.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS2101</td>
<td>76.7</td>
<td>S. purpuratus</td>
<td>70.8</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>S. franciscanus</td>
<td>70.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>L. pictus</td>
<td>63.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS2108</td>
<td>83.4</td>
<td>S. purpuratus</td>
<td>56.8</td>
<td>-1.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>S. franciscanus</td>
<td>58.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS2137</td>
<td></td>
<td>S. purpuratus</td>
<td>64.0</td>
<td>6.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>S. franciscanus</td>
<td>58.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Frequency Repetitive S. purpuratus DNA (Figure 4A)</td>
<td></td>
<td>S. purpuratus</td>
<td>74.5</td>
<td>4.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>S. franciscanus</td>
<td>70.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intermediate Repetitive S. purpuratus DNA (Figure 5)</td>
<td></td>
<td>S. purpuratus</td>
<td>70.0</td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>S. franciscanus</td>
<td>66.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The native T_m is that of the non-denatured or renatured cloned DNA and is a function of the length and base composition. These values are from Klein et al. (1978).
* The driver DNA lengths were approximately 850 nucleotides in each case. The reaction was run to at least 90 times their kinetically determined 
  C0 values, usually to Cot 1000. Termination points varied depending upon trace reactivity, but were usually 70-90%. The cloned repetitive
  sequence segments were reacted to termination with 100 μg of the indicated driver DNA in 0.4 M NaCl, 0.05% SDS, 1 mM EDTA at 50°C.
  The reaction mixture was diluted to 1.2 M NaCl 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 2.7 ml aliquots of 0.12 M NaCl. An internal standard,
  consisting of native S. purpuratus M-DNA of 450 nucleotides, was added to the cloned repetitive H-F-DNA reaction mixture, and the
  thermal elution profiles of both DNA were obtained simultaneously. The native DNA standards had a mean T_m of 83.3°C ± 0.3°C (SD).
  The T_m 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 <0.3°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 10^4 fold 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.
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 hydroxypatite and the number of members of the family it represents may be calcu-
Table 2. Frequencies Determined by Resassociation Kinetics

<table>
<thead>
<tr>
<th>Clone</th>
<th>Length (bp)</th>
<th>Frequency* (S. purpuratus)</th>
<th>Frequency* (S. franciscanus)</th>
<th>Ratio</th>
<th>Corrected Ratio[a]</th>
<th>Frequency* (L. pictus)</th>
<th>Ratio</th>
<th>Corrected Ratio[a]</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS2007</td>
<td>1100</td>
<td>400</td>
<td>45</td>
<td>9</td>
<td>7</td>
<td>low</td>
<td>100</td>
<td>25</td>
</tr>
<tr>
<td>CS2004</td>
<td>580</td>
<td>1000</td>
<td>160</td>
<td>6</td>
<td>4</td>
<td>10</td>
<td>100</td>
<td>25</td>
</tr>
<tr>
<td>CS2000</td>
<td>220</td>
<td>140</td>
<td>95</td>
<td>1.5</td>
<td>1.5</td>
<td>low</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CS2009</td>
<td>235</td>
<td>80</td>
<td>60</td>
<td>1.3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CS2101</td>
<td>320</td>
<td>700</td>
<td>55</td>
<td>13</td>
<td>13</td>
<td>2</td>
<td>350</td>
<td>100</td>
</tr>
<tr>
<td>CS2106</td>
<td>190</td>
<td>20</td>
<td>30</td>
<td>0.7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CS2109</td>
<td>125</td>
<td>200</td>
<td>50</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CS2133A</td>
<td>310</td>
<td>2100</td>
<td>102</td>
<td>21</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CS2137</td>
<td>190</td>
<td>530</td>
<td>185</td>
<td>3</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

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

* Resaturation frequency is calculated as \( S = K_0 / K_{2,4} \), where \( K_0 \) is the second-order rate constant for the reaction of the cloned \( ^{3}H-DNA \) fragment with sea urchin DNA, and \( K_{2,4} \) is the second-order rate constant for the reaction of the \( ^{3}H-DNA \) single-copy tracer with the same sea urchin DNA doer. Second-order rate constants were obtained by a least-squares method (Pearson et al., 1977). Since factors such as the mean differences in single-copy DNA, driver DNA and clones DNA fragment lengths, thermal stability of the reaction products, salt concentrations, and other unknown variables affect the absolute rate of the reactions, the frequency 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 850 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: \( K_0 = K_{2,4} (LT/L_C) \), where \( K_{2,4} \) is the length corrected second-order rate constant for the reaction of the cloned fragment with the driver DNA, \( K_0 \) is the observed rate constant for the reaction, \( L_C \) is the length of the cloned fragment and \( L_T \) 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 resaturation frequency of these cloned sequences in the \( S. \) purpuratus genome only were made simultaneously by Klein et al. (1978) and ourselves. 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 CS2000 and CS2009, agreement was within a factor of three.

[a] For the corrected ratio, the resaturation frequency of the cloned sequences in the \( S. \) franciscanus and \( L. \) pictus genomes was corrected for the effect of sequence divergence on resaturation using the equation \( F_{2,4} = \exp (-2L^2) \), which is equivalent to a 2 fold rate retardation for each \( 10^2 \) change in sequence divergence (Bonner et al., 1979), where \( L^2 \) is the difference in \( L_0 \) for \( S. \) purpuratus DNA compared to the driver of the other species.

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 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 columns 6 are very close to the
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 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 Ressociation 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 higher 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 tracer to target (400,000:1) so that even one recognizable copy was 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

<table>
<thead>
<tr>
<th>Clone and Length (Nucleotides)</th>
<th>Species</th>
<th>Q*</th>
<th>Apparent Frequency</th>
<th>Ratio</th>
<th>Length Corrected Frequency</th>
<th>Frequency by Kinetics</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS2101 L = 320</td>
<td>S. purpuratus</td>
<td>4,000</td>
<td>1,250</td>
<td>1</td>
<td>781</td>
<td>996</td>
</tr>
<tr>
<td></td>
<td>S. franciscanus</td>
<td>39,000</td>
<td>128</td>
<td>10</td>
<td>75</td>
<td>82</td>
</tr>
<tr>
<td></td>
<td>L. pictus</td>
<td>444,000</td>
<td>11</td>
<td>110</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>CS2034 L = 560</td>
<td>S. purpuratus</td>
<td>1,050</td>
<td>2,721</td>
<td>1</td>
<td>971</td>
<td>1,080</td>
</tr>
<tr>
<td></td>
<td>S. franciscanus</td>
<td>4,630</td>
<td>617</td>
<td>4.4</td>
<td>220</td>
<td>126</td>
</tr>
<tr>
<td>CS2007 L = 1100</td>
<td>S. purpuratus</td>
<td>318</td>
<td>4800</td>
<td>1</td>
<td>838</td>
<td>465</td>
</tr>
<tr>
<td></td>
<td>S. franciscanus</td>
<td>3,780</td>
<td>283</td>
<td>12</td>
<td>71</td>
<td>77</td>
</tr>
</tbody>
</table>

* Ratio of sea urchin DNA to cloned DNA tracer at half-reaction using least-squares solution.

* Calculation using frequency = 2Q/G, where Q is the genome size (8 x 10^6 nucleotides), L is the cloned repetitive sequence fragment length and G is from column 3.

* Empirical length corrected frequency obtained by dividing column 4 by (L/195).

* 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 Tm 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

<table>
<thead>
<tr>
<th>Frequency Ratio</th>
<th>&lt;1°C</th>
<th>&gt;1°C</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤2</td>
<td>CS2000 (1°C, 1.6%)</td>
<td>CS2127 (6°C, 2)</td>
</tr>
<tr>
<td>&gt;2</td>
<td>CS2101 (6°C, 13)</td>
<td>CS2007 (3.5°C, 7)</td>
</tr>
</tbody>
</table>

* The difference in Tm from Table 1 for the cloned DNA reassocia-
ted with S. purpuratus DNA compared to that with S. franci-
canus DNA. The five cases are those for which the Tm is far
enough from the criterion of precision for the Tm difference to be
considered reliable.
* The ratio of the number of members of the family with which the
cloned DNA reassociates in S. purpuratus DNA compared with S.
franciscanus and S. purpuratus.

The measured Tm difference and the frequency ratio are listed
individually in the parentheses.

It appears 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. francisci-
canus and S. purpuratus.

Families of repeated sequences are apparently
recognizable over long evolutionary periods al-
though 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 do-
nominate families in modern genomes are principi-
taly 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 repeat 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, un-
published 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 pro-
duct 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 frequency changes of
frequency changes in the evolution of repeated
DNA. Measurements have been made in two in-
sects (P. Dunmur, 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. simu-
lans and D. melanogaster). An observation that is
probably due to frequency change has been made
in rodents. The mouse satelitte is observed in three
closely related species (Sutton and McGaugh, 1972; Rice and Straus, 1973) in upwards of a
million fairly precise copies which reassociate to
form strand pairs whose Tm is within 5°C of that of
perfect duplex. Large interspecies reassocation is
observed and the interspecies dupplexes melt
sharply at about 25°C below perfectly complemen-
tary 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 spe-
cies.

Many measurements of interspecies repeated
sequence reassocation 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 550 nucleotides and the driver/tracer DNA mass ratios were approximately $2 \times 10^9$, 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 (E) S. purpuratus, (C) S. franciscanus, (B) L. pictus or (F) self-reacted. The S. purpuratus and S. franciscanus driven reactions terminate at 88.7 and 79.5%, with rate constants of $8.48 \times 10^{-1}$ and $4.3 \times 10^{-1}$ M$^{-1}$ sec$^{-1}$, respectively. (B) S. franciscanus tracer reassociated with (C) S. franciscanus or (E) S. purpuratus driver DNA. The reactions terminate at 91.3 and 91.0% with rate constants of 2.2 and 3.93 M$^{-1}$ sec$^{-1}$, respectively. (C) L. pictus tracer reassociated with (B) L. pictus or (E, C) S. purpuratus DNA, or (F) self-reacted.

Figure 5. Reassociation of Middle Repetitive S. purpuratus Tracer with Homologous and Heterologous Driver DNA
S. purpuratus middle repetitive tracer was reacted with (E) S. purpuratus, (C) S. franciscanus, (B) L. pictus, (C) self DNA or (F) self-reacted. Weight mean driver DNA fragment length was approximately 550 nucleotides and the driver/tracer DNA weight ratio was approximately $2 \times 10^9$. 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 93.0% and had a 0.1% component with a rate constant of 0.022 M$^{-1}$ sec$^{-1}$, and a 41.9% component with a rate constant of 0.31 M$^{-1}$ sec$^{-1}$. The S. franciscanus driven reaction fitted 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 appears the L. pictus driven reaction terminates at 88.3% with a rate constant of $2.8 \times 10^{-1}$ M$^{-1}$ sec$^{-1}$. For some S. purpuratus points (C), the driver/tracer DNA mass ratio was increased to $4 \times 10^9$. 

The L. pictus driven reaction terminates at 88.3% with a rate constant of $2.8 \times 10^{-1}$ M$^{-1}$ sec$^{-1}$. For some S. purpuratus points (C), the driver/tracer DNA mass ratio was increased to $4 \times 10^9$. 
cies repetitive DNA cross-reactions among pri-
mates, 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 MacGreg-
or (1974) and Flavell and Smith (1976) have con-
cluded that frequency changes or saltatory repli-
cations have occurred in amphibians and mono-
cots. 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 com-
plementary 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 x 10^4 = 2.5 x 10^-4 per year. There are
approximately 5 x 10^4 interpersed repetitive se-
quence 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 in-
cludes many different families which may have
quite different rates and could be affected by a few
large changes in individual families termed sallar-
tory 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 or-
dovician (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 x 10^8 nucleotide pairs. This
argument suggests a long-term balance of the
processes of addition and deletion, but the ques-
tion 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 ma-
ture oocyte RNA (Costantini et al., 1978) and in
nuclear RNA in early development (Scheller et al.,
1978). High concentrations of transcripts from dif-
f erent families were shown to occur at different
stages of development. Thus transcription of mem-
bers 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 observation that repetitive con-
dence 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 fami-
lies. 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 inser-
ion 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 differently labeled (L5, L6, L7, L8,
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 un-
altered. We therefore propose the following hy-
pothesis: during evolution most families of re-
peated sequences are replaced with others which
perform slightly modified functions leading to phe-
notypic 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. Evo-
olutionary changes in genomic sequence organiza-
tion might exhibit coordinated characteristics and
could provide a significant source of morphologi-
ical 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 Angerter et al. (1978). DNA preparations were made from 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 (A260 at 98°C/A260 at 60°C/C∞ at 98°C = 100) in 0.12 M Na (Na phosphate (pH 8.3)).

Preparation and Sizing of DNA Fragments

DNA fragments of the indicated sizes were produced by homog

Preparation and Labeling of Tracers

For the preparation of radioactive labeled S. purpuratus DNA, embryos were raised to the 4 cell stage as described by Hingmard

Hepatitis A (HAP) Chromatography

The fraction of DNA fragments in molecules containing duaplexes after reassociation was determined by chromatography on HAP (BioRad, DNA grade lot No. 18288). Samples were diluted to 0.12 M PB and applied to water-precipitated 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 30°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 fractions for the tracer), the column was presaturated with 20 µg of calf thymus DNA. HAP was also used to concentrate DNA solutions by dilution of the sample to 0.02 ml PB, application to the column and finally elution with 0.4 M PB in a small volume.

Saturation Estimate of Frequency

When the amount of radioactive DNA in the virus is varied at a constant molar ratio of the two DNA species, the saturation can be expressed as a 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, Davi-
Acknowledgments

The knowledge and assistance of P. Dunham, P. Bingham and M. Masek with Drosophila species was helpful in recognizing the generality and significance of the results. Unpublished measurement by Dr. Robert A. Anger 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 Roan. The 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


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\textsuperscript{+} 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 repetitive 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 \( \lambda \) genome libraries a number of \( \lambda \)-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 MgCl₂, 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 7.8), 10 mM MgCl₂, 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 10⁴ 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 10⁵ and 2.5 x 10⁵ different phage, respectively. The first of these libraries (denoted SpλR₁A) was derived from DNA of the sperm of 5 males and the second (denoted SpλR₁B) 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 ≥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 $7.8 \times 10^5$ 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 $C_0 t 40$ 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 $-20^\circ C$. 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 $37^\circ C$. 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 $60^\circ C$ and passed over a 1 ml column of hydroxyapatite at $50^\circ C$. 
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 $2 \times 10^4$ recombinant phage were plated on $4 \times 10^8$ 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, $\sim 2 \times 10^5$ cpm tracer ($\sim 10^7$ 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 MgCl₂). 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 x 10¹⁰ 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 C₀t values quoted in the text are equivalent C₀t (Britten et al., 1974). For example, C₀t in 0.41 M phosphate buffer was converted to equivalent C₀t by multiplying by 5. Renaturation kinetic analyses carried out in this work included an internal single copy DNA rate standard. Single copy ³H-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 bacterio-
phage 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 >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 occurrence 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 $\lambda$-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 $\lambda$ 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 $2 \times 10^4$ 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 $\lambda$ 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 repli-
cation 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 cri-
terion 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 mul-
tiple 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 $^{32}\text{P}$, and reacting them with excess sea urchin DNA. Single copy $^{3}\text{H}$-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 organi-
zation 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 sub-
fragments 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 $2 \times 10^7$ nt of DNA. Thus, unless the $\lambda$ 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**

<table>
<thead>
<tr>
<th>Family&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Length of probe sequence (nt)&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Repeat length class&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Family size: Genomic reiteration frequency&lt;sup&gt;d&lt;/sup&gt;</th>
<th>Intrafamilial thermal stability:&lt;sup&gt;e&lt;/sup&gt; $T_m$ (°C)</th>
<th>Representation of family in RNA:&lt;sup&gt;f&lt;/sup&gt; Oocyte RNA</th>
<th>Gastrula mRNA</th>
<th>Intestine mRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>2034</td>
<td>498</td>
<td>long</td>
<td>2500</td>
<td>3.7</td>
<td>0.2</td>
<td>0.9</td>
<td>48</td>
</tr>
<tr>
<td>2108</td>
<td>204</td>
<td>long</td>
<td>20</td>
<td>5.7</td>
<td>590</td>
<td>95</td>
<td>55</td>
</tr>
<tr>
<td>2109 A</td>
<td>180</td>
<td>long</td>
<td>900</td>
<td>19.7</td>
<td>0.3</td>
<td>1</td>
<td>22</td>
</tr>
<tr>
<td>2109 B</td>
<td>110</td>
<td>short</td>
<td>1000</td>
<td>25.3</td>
<td>5</td>
<td>64</td>
<td>1.8</td>
</tr>
</tbody>
</table>

<sup>a</sup> 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).

<sup>b</sup> Lengths of the cloned repeat fragments have been obtained from their primary sequence (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).
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_{0}t 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).

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).

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
The $\Delta T_m$ 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.

\[ \text{% representation} = \frac{T_c}{F_c \times T_{sc}} \times 100, \]

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 \( \lambda \)-genome library screens

with cloned repeat sequence probes\(^a\)

<table>
<thead>
<tr>
<th>Repeat family</th>
<th>Positive plaques per plate</th>
<th>Approximate expectation if repeat sequences occur singly(^c)</th>
<th>Total No. ( \lambda ) recombinants plaque-purified(^d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2034</td>
<td>304</td>
<td>920</td>
<td>7</td>
</tr>
<tr>
<td>2108(^b)</td>
<td>9</td>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td>2109 A probe(^b)</td>
<td>343</td>
<td>370</td>
<td>13</td>
</tr>
<tr>
<td>2109 B probe</td>
<td>375</td>
<td>330</td>
<td>27</td>
</tr>
</tbody>
</table>

\(^a\) 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 \( 2 \times 10^4 \) pfu per 15 cm plate. Only representative data are shown; a large number of different screening experiments was actually carried out with the several \( \lambda \)-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
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 \( \lambda \) recombinants, for which we take \( 1.5 \times 10^4 \) ntp; \( R \) is the repetition frequency of the repeat family, from Table 1; and \( G \) is the haploid genome size for \( S. purpuratus \), or \( 8.1 \times 10^8 \) ntp.

\( ^d \)This column lists the total number of positive clones ultimately selected from each family for further analysis.
<table>
<thead>
<tr>
<th>Family</th>
<th>isolate&lt;sup&gt;a&lt;/sup&gt;</th>
<th></th>
<th>Enzyme&lt;sup&gt;b&lt;/sup&gt;</th>
<th></th>
<th>Estimated number of family sequences per λ recombinant&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>2109</td>
<td>λ2109A-6</td>
<td>1</td>
<td>HaeIII, HhaI, HpaII, HinfI</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>λ2109B-6</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Family</td>
<td>isolate&lt;sup&gt;a&lt;/sup&gt;</td>
<td>HaeIII</td>
<td>HhaI</td>
<td>HpaII</td>
<td>HinfI</td>
</tr>
<tr>
<td>--------</td>
<td>---------------------</td>
<td>--------</td>
<td>------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>2109</td>
<td>λ2109B-16</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>52</td>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2034</td>
<td>λ2034-3</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>2</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>5</td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Family</td>
<td>isolate&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Enzyme&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td>Estimated number of family sequences per λ recombinant&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>------------------</td>
<td>------------------</td>
<td>---</td>
<td>-------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HaeIII</td>
<td>Hhal</td>
<td>HpaII</td>
<td>HinfI</td>
</tr>
<tr>
<td>2034</td>
<td>λ2034-13</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>2108</td>
<td>λ2108-16</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>high</td>
<td>32</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>criterion</td>
<td>33</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>selection</td>
<td>34</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>λ2108-8</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>low</td>
<td>12</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>criterion</td>
<td>15</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>selection</td>
<td>17</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>
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 SpλR₁B. The other λ recombinants listed were isolated from library SpλH₃C.

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

<table>
<thead>
<tr>
<th>λ recombinant</th>
<th>Digest(^a)</th>
<th>Fragment length</th>
<th>(K^b)</th>
<th>corrected(^c)</th>
<th>Genomic reiteration frequency(^d)</th>
<th>Restriction map(^e)</th>
</tr>
</thead>
<tbody>
<tr>
<td>λ2034-9</td>
<td>HaeIII</td>
<td>640</td>
<td>4.4</td>
<td>4.1</td>
<td>3280</td>
<td>331 3280 3280</td>
</tr>
<tr>
<td></td>
<td></td>
<td>640</td>
<td>4.4</td>
<td>4.1</td>
<td>3280</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>560</td>
<td>4.0</td>
<td>4.1</td>
<td>3310</td>
<td></td>
</tr>
<tr>
<td>λ2034-17</td>
<td>HhaI(^f)</td>
<td>1400</td>
<td>3.7</td>
<td>1.6</td>
<td>1260</td>
<td></td>
</tr>
<tr>
<td>HinfI</td>
<td></td>
<td>1300</td>
<td>3.5</td>
<td>1.6</td>
<td>1260</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>750</td>
<td>1.5</td>
<td>1.2</td>
<td>950</td>
<td></td>
</tr>
<tr>
<td>λ2034-18</td>
<td>HhaI</td>
<td>1380</td>
<td>1.5</td>
<td>0.60</td>
<td>780</td>
<td>1400 780 2160</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>1.6</td>
<td>1.8</td>
<td>2160</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>380</td>
<td>0.90</td>
<td>1.13</td>
<td>1400</td>
<td></td>
</tr>
<tr>
<td>λ2108-11</td>
<td>HinfI</td>
<td>640</td>
<td>0.015</td>
<td>0.014</td>
<td>13</td>
<td>19 13 12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>560</td>
<td>0.02</td>
<td>0.021</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>480</td>
<td>0.012</td>
<td>0.013</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>λ2108-12</td>
<td>HaeIII</td>
<td>720</td>
<td>0.004</td>
<td>0.0033</td>
<td>3</td>
<td>3 4 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>680</td>
<td>0.004</td>
<td>0.0035</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>520</td>
<td>0.003</td>
<td>0.0032</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>λ recombinant</td>
<td>Digest&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Fragment length</td>
<td>K&lt;sup&gt;b&lt;/sup&gt;</td>
<td>K corrected&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Genomic reiteration frequency&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Restriction map&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>---------------</td>
<td>----------------</td>
<td>----------------</td>
<td>------------</td>
<td>----------------</td>
<td>---------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>λ2108-15</td>
<td>Hinfl</td>
<td>570</td>
<td>0.008</td>
<td>0.009</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>510</td>
<td>0.02</td>
<td>0.021</td>
<td>21</td>
<td>21 9 6 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>400</td>
<td>0.005</td>
<td>0.006</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>315</td>
<td>0.001</td>
<td>0.0013</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>λ2108-16</td>
<td>Hinfl&lt;sup&gt;+&lt;/sup&gt; Hpall</td>
<td>600</td>
<td>0.02</td>
<td>0.02</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hpall</td>
<td>580</td>
<td>0.052</td>
<td>0.053</td>
<td>53</td>
<td>40 20 70 53</td>
</tr>
<tr>
<td></td>
<td></td>
<td>240</td>
<td>0.025</td>
<td>0.04</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>160</td>
<td>0.036</td>
<td>0.07</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>λ2108-21</td>
<td>Hpall</td>
<td>1170</td>
<td>0.10</td>
<td>0.051</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>780</td>
<td>0.047</td>
<td>0.036</td>
<td>36</td>
<td>17 36 51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>650</td>
<td>0.018</td>
<td>0.017</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>λ2109A-6</td>
<td>Hpall&lt;sup&gt;+&lt;/sup&gt; Hinfl</td>
<td>995</td>
<td>0.0027</td>
<td>0.0016</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hinfl</td>
<td>680</td>
<td>0.0064</td>
<td>0.0056</td>
<td>4</td>
<td>410</td>
</tr>
<tr>
<td></td>
<td></td>
<td>600</td>
<td>0.53</td>
<td>0.53</td>
<td>410</td>
<td></td>
</tr>
<tr>
<td>λ2109A-8</td>
<td>Hpall</td>
<td>260</td>
<td>0.0001</td>
<td>0.00015</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>950</td>
<td>0.0025</td>
<td>0.0013</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>700</td>
<td>1.4</td>
<td>1.2</td>
<td>920</td>
<td></td>
</tr>
<tr>
<td>λ recombinant</td>
<td>Digest$^a$</td>
<td>Fragment length</td>
<td>$K^b$</td>
<td>corrected$^c$</td>
<td>Genomic reiteration frequency$^d$</td>
<td>Restriction map$^e$</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------</td>
<td>----------------</td>
<td>------</td>
<td>-------------</td>
<td>---------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>$\lambda$2109A-11</td>
<td>HpaII</td>
<td>670</td>
<td>0.0024</td>
<td>0.0021</td>
<td>2</td>
<td>$2X$ $4X$ $1X$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>450</td>
<td>0.0045</td>
<td>0.0052</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>360</td>
<td>0.0012</td>
<td>0.0015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda$2109A-12</td>
<td>HaeIII+</td>
<td>900</td>
<td>0.0011</td>
<td>0.00073</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>HpaII</td>
<td></td>
<td>560</td>
<td>0.0011</td>
<td>0.0011</td>
<td>1</td>
<td>$1X$ $3X$ $1X$ $1X$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>440</td>
<td>0.0033</td>
<td>0.0039</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>400</td>
<td>0.00071</td>
<td>0.0009</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>$\lambda$2109B-9</td>
<td>HaeIII+</td>
<td>440</td>
<td>1.4</td>
<td>1.63</td>
<td>680</td>
<td>$1$ $680$ $1$ $1$</td>
</tr>
<tr>
<td>HinfI</td>
<td></td>
<td>360</td>
<td>0.0016</td>
<td>0.0021</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>340</td>
<td>0.0013</td>
<td>0.0017</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>210</td>
<td>0.002</td>
<td>0.0033</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>$\lambda$2109B-11</td>
<td>HhaI+</td>
<td>1120</td>
<td>0.0038</td>
<td>0.0020</td>
<td>1</td>
<td>$1$ $3$</td>
</tr>
<tr>
<td>HinfII</td>
<td></td>
<td>305</td>
<td>0.0044</td>
<td>0.0061</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>$\lambda$2109B-14</td>
<td>HpaII+</td>
<td>850</td>
<td>0.16</td>
<td>0.11</td>
<td>70</td>
<td>$1$ $73$ $2$ $2$</td>
</tr>
<tr>
<td>HaeIII</td>
<td></td>
<td>800</td>
<td>0.005</td>
<td>0.0037</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>λ recombinant</td>
<td>Digest^a</td>
<td>Fragment length</td>
<td>K^b</td>
<td>K corrected^c</td>
<td>Genomic reiteration frequency^d</td>
<td>Restriction map^e</td>
</tr>
<tr>
<td>--------------</td>
<td>---------</td>
<td>----------------</td>
<td>-----</td>
<td>---------------</td>
<td>-------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>λ2109B-14</td>
<td></td>
<td>300</td>
<td>0.0035</td>
<td>0.0049</td>
<td>3 73 2 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>150</td>
<td>0.001</td>
<td>0.002</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>λ2109B-17^e</td>
<td>HpaII+ HinfI</td>
<td>1070</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>380</td>
<td></td>
<td></td>
<td>250 7 250</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td></td>
<td></td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>λ2109B-18</td>
<td>HaeIII</td>
<td>500</td>
<td>0.158</td>
<td>0.173</td>
<td>160 250 160 25</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>440</td>
<td>0.24</td>
<td>0.280</td>
<td>250</td>
<td></td>
</tr>
</tbody>
</table>
NOTES TO TABLE 4

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 $M^{-1} \text{sec}^{-1}$.

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).

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 $M^{-1} \text{sec}^{-1}$.

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 $2 \times 10^4$ pfu from library SpλR$_1$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 $2 \times 10^4$ pfu from library SpλH$_3$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.
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'-'-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'-'-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'-'-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'-'-P-2109A probe (A–C) or with 5'-'-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'-32P-2034 probe.

G, HhaI subfragment of λ2108-16, digested with HpaII, and hybridized with 5'-32P-2108 probe.
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, Hinfl; 5, HincII; 6, TaqI; 7, Aval; 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.
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.
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.
References


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.


