Volume 4 Number 5 1977 Nucleic Acids Research

Exploration of long and short repetitive sequence relationships in the sea urchin genome

Francine C. Eden[1], Dale E. Graham[2], Eric H. Davidson and Roy J. Britten[3]

Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

Received 29 January 1977

ABSTRACT

Long and short repetitive sequences of sea urchin DNA were prepared by reassociation of 2000 nucleotide long fragments to Cot 4 and digestion with the single strand specific nuclease S1. The S1 resistant duplexes were separated into long repetitive and short repetitive fractions on Agarose A50. The extent of shared sequences was studied by reassociating a labeled preparation of short repetitive DNA with an excess of unlabeled long repetitive DNA. Less than 10% of the long repetitive DNA preparation was able to reassociate with the short repetitive DNA. Thus the long and short repetitive elements appear to be principally independent sequence classes in sea urchin DNA.

Precisely reassociating repetitive DNA was prepared by four successive steps of reassociation and thermal chromatography on hydroxyapatite. This fraction (3% of the genome) was reassociated by itself or with a great excess of total sea urchin DNA. The thermal stability of the products was identical in both cases (Tm=81°C), indicating that precisely repeated sequences do not have many imprecise copies in sea urchin DNA.

INTRODUCTION

The DNA of the sea urchin Strongylocentrotus purpuratus exhibits short period interspersion of repetitive and nonrepetitive sequences.[4] This highly organized pattern of DNA sequences is similar to that of Xenopus laevis[5,6] and most eukaryotes which have been studied.[7-10] Only the genomes of a few insects are known to have a strikingly different pattern.[11,12]

The features of the short period of Xenopus interspersion pattern[5] which are germane to this paper may be summarized as follows: (1) the majority of the DNA is single copy sequence (75% in the case of S. purpuratus DNA); (2) a broad spectrum of repetition frequencies is present, ranging from a few copies to many thousands of copies per genome; (3) repetitive and single copy sequences are interspersed, and single copy lengths range from less than a thousand nucleotides to several thousand nucleotides; (4) the interspersed repetitive sequences are typically short with a mode length near 300 nucleotides; (5) a smaller quantity of repetitive sequences occurs with lengths greater than 1000 or 2000 nucleotides; (6) the long repetitive sequences reassociate precisely to yield high thermal stability duplexes, while short repeated sequences are typically

divergent and reassociate to form duplexes of the same length and decreased thermal stability.[13]

It has long been recognized that both precisely repeated and imprecisely repeated sequences are present in eukaryote genomes, and that these classes are at least partially separate sets of sequences which do not reassociate with each other.[14] We suggested earlier that there is an evolutionary relationship in which long precisely repetitive regions may be precursors to short interspersed sequences which are more divergent but this issue remains unresolved.[14-16] The possibility of evolutionary and functional relationships among the various classes of repeated sequences raises a number of important and interesting issues. This paper describes new measurements of the sequence relationships between long and short repetitive sequences and between precisely and imprecisely repeated DNA sequences.

MATERIALS AND METHODS

Preparation of DNA

DNA was extracted from Strongylocentrotus purpuratus sperm or embryos as described by Graham et al.[4]

Fragmentation of DNA and Fragment Length Determinations

DNA fragments of desired lengths were prepared by shearing in a Virtis-60 homogenizer in a volume of 33 ml. 2000 nucleotide fragments were prepared in 0.12 M phosphate buffer (PB), pH 6.8, at 0°C at 12,000-15,000 rpm for 20 min. 300 nucleotide fragments were prepared by shearing in 66% glycerol, 0.01 M Na acetate, pH 6, in a dry ice-ethanol bath at 55,000 rpm for 20 min.

Fragment lengths were determined using isokinetic alkaline sucrose gradients[17] in an SW41 rotor (V_{mix} = 9.8 ml, C_{res} = 43% w/v, C_{flask} = 16% w/v in 0.1 M NaOH). Each length was determined in duplicate with two reference markers in each gradient. Reference marker lengths were determined by measurement of Kleinschmidt preparations in the electron microscope.

DNA Reassociation and Hydroxyapatite Chromatography

DNA fragments were reassociated in 0.12 M PB at 60°C or 0.4 M PB at 64°C. After the incubation the DNA was applied to hydroxyapatite columns (BioRad, DNA grade) in 0.12 M PB at 60°C. The single-strand fraction was eluted with 0.12 M PB; duplex-containing fragments were eluted with 0.4 M PB at 60°C or with 0.12 M PB at 100°C. Equivalent Cot for the 0.4 M PB incubations was calculated by multiplying Cot (moles/l sec) by a rate acceleration factor of 4.9.[18]

S1 Nuclease Digestion

DNA samples to be digested with S1 nuclease were reassociated in 0.3 M NaCl, 0.01 M PIPES buffer (Sigma), pH 6.7, at 64°C. Equivalent Cot was calculated by multiplying Cot by a rate of acceleration factor of 2.3. After the incubation, the DNA was

adjusted to 0.025 M Na acetate, pH 4.3, 0.1 mM $ZnSO_4$, 5 mM mercaptoethanol, 0.15 M NaCl; S1 nuclease (purified from Aspergillus in this laboratory)[19] was added at the desired ratio of enzyme to DNA[13] and incubated for 45 min at 37°C. The mixture was adjusted to 0.12 M PB and duplexes collected on hydroxyapatite at 60°C. The estimated fraction of the single strand DNA digested according to the criterion of inclusion by Sephadex G100 is termed DIG. Here we follow the method of calculation of Britten et al.[13]: $DIG = 1 - (1 + d/H)^{-1}$ where d is the ratio of enzyme to DNA multiplied by time used in the particular incubation in arbitrary units (min x µl per mg DNA). H is the ratio of enzyme to DNA multiplied by time in the same units required for half digestion of single-stranded DNA by the same criterion.

Agarose A50 columns (BioRad) (1 x 100 cm) were poured around a support of glass beads (3 mm) to prevent compression and improve flow rates. DNA duplex lengths were determined by chromatography in 0.12 M PB at room temperature. Long native DNA was used as an exclusion marker, and labeled orthophosphate was the inclusion marker. Double-stranded DNA which had been sheared and sized by alkaline sucrose gradients was used for size calibration.

In Vitro Labeling of DNA by "Gap Translation"

DNA was labeled in vitro by a modification of the method of Schachat and Hogness[20] as previously described.[21] DNase pretreatment was omitted and labeling by E. coli DNA polymerase I was initiated at gaps in the reassociated DNA structures. DNA was labeled to a specific activity of 3×10^6 cpm/µg.

Hydroxyapatite Thermal Chromatography

DNA was loaded on hydroxyapatite columns at 60°C in 0.12 M PB. The temperature was raised in intervals of 3°C and the column was washed with 8-10 bed volumes of 0.12 M PB at each temperature. DNA eluting from the column between 63°C and 98°C was included in the calculation of thermal stability. In the isolation of the high precision repetitive DNA fraction, reassociated DNA was bound to hydroxyapatite at 60°C, in 0.12 M PB. The column was then washed at 80°C with 0.12 M PB and the high thermal stability fraction was eluted from the column at 80°C with 0.4 M PB.

RESULTS

Preparation of Long and Short Repetitive DNA Fractions

Sea urchin DNA with a mode length of 2000 nucleotides was reassociated to Cot 4 and digested with the single strand specific nuclease S1[19], as described in Materials and Methods. The quantity of S1 was chosen so that 78% of the single-stranded DNA was digested by G100 assay (DIG = 0.78).[13] Such a digestion results in nearly complete removal of single-strand tails from duplex regions without appreciable digestion of the reassociated divergent repetitive duplexes bound by hydroxyapatite.[13]

After digestion the S1 resistant duplexes were bound to hydroxyapatite (at 60°C in 0.12 M PB) and then chromatographed on Agarose A50 to separate them into long and short fractions (Fig. 1). The fraction of the original total DNA bound to hydroxyapatite was 19.4% and the pooled fractions shown in Fig. 1 contain 5.6% and 9.7% of the genome. The Cot used here is 4 and the quantity binding to hydroxyapatite[13] is less than at Cot 20. The long and short repetitive fractions together comprised about 80% of the resistant DNA. Fragments of intermediate length and very short fragments were discarded. This procedure was repeated with labeled sea urchin DNA with quantitatively similar results, and equivalent fractions were pooled.

To show that the long and short S1 resistant duplex preparations exhibited the characteristics previous observed[13] they were melted in a spectrophotometer. The long fraction displayed a Tm less than 1°C below that of native DNA of the same length while the short duplex fraction displayed a Tm about 9°C lower. This is more than 7°C

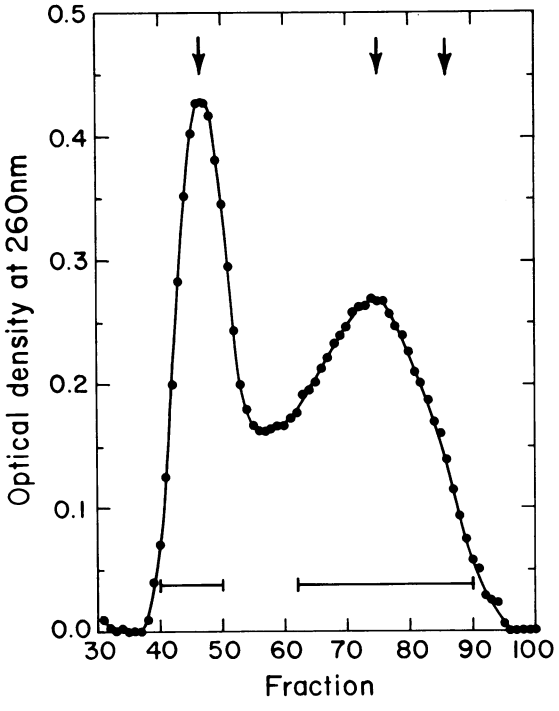

Figure 1. Agarose A50 Fractionation of S1 Nuclease Resistant Repetitive DNA. Sea urchin DNA was sheared to a mode length of 2000 nucleotides, reassociated to Cot 4, and digested with S1 nuclease (DIG = 0.78) as described in Materials and Methods. S1 resistant duplexes were collected on hydroxyapatite and chromatographed on Agarose A50 in 0.12 M PB at room temperature. The column was calibrated prior to use; from left to right the position of exclusion peak, a 300 ntp marker, and the inclusion peak are indicated by arrows. Fractions indicated by the bars were pooled as the long and short repetitive DNA preparation discussed in text.

below the Tm of native DNA duplexes of equal length (i.e., about 300 nucleotides). Samples of these two preparations were also redigested at two levels of S1 nuclease and rechromatographed on Agarose A50. With the same enzyme treatment as was used in their preparation (DIG = 0.78) very little change in the length and quantity of the short preparation was observed and the long preparation remained completely excluded from Agarose A50. When ten times more enzyme was used (DIG = 0.97) the long fraction remained 95% resistant and was still excluded. However under these conditions 30% of the short fraction was digested, as measured by its ability to bind to hydroxyapatite, and the mode length was slightly reduced. The difference in behavior is attributable to the fact that the short repetitive DNA duplexes are partially mismatched and under the more severe digestion conditions some of the mismatch sites are cut.

Reassociation Characteristics of the Long and Short Repetitive DNA Fractions

The labeled long and short repetitive DNA preparations were reassociated in the presence of a great excess of total DNA driver to display the repetition frequencies present. The results are shown in Fig. 2. For these measurements the long repetitive tracer preparation was sheared to about 300 nucleotides. The short DNA preparation

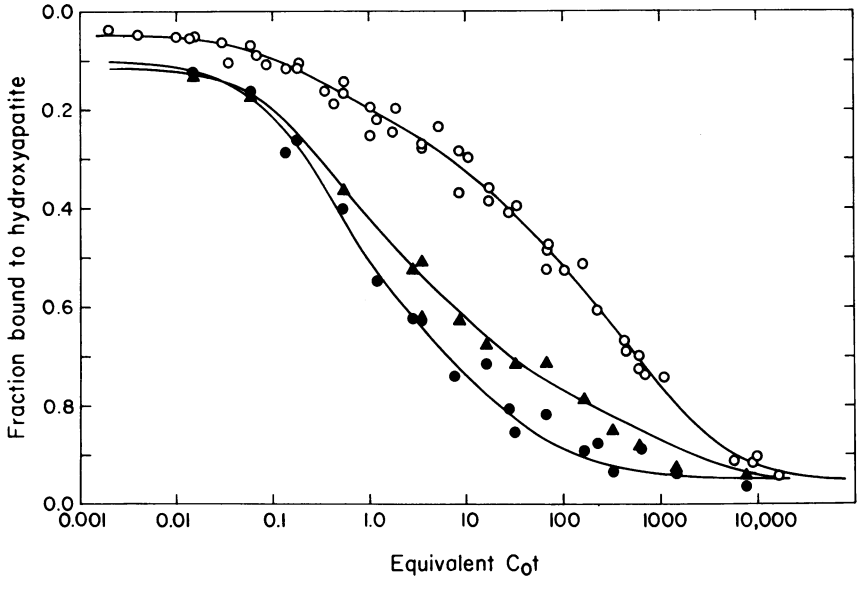

Figure 2. Measurement of Repetition Frequency Classes in the Long and Short DNA Fractions. Long and short repetitive fractions of ^{3}H-DNA were prepared as described in Fig. 1. The long repetitive DNA fraction was sheared to 300 ntp for this experiment. Each fraction was reassociated with at least a 1000-fold excess of unlabeled total DNA (450-550 ntp) in 0.12 M PB at 60°C or equivalent, and the reactions were assayed by hydroxyapatite chromatography at 60°C in 0.12 M PB; (o) total sea urchin DNA driver; (▲) long repetitive tracer; (●) short repetitive tracer.

was not sheared since the mode length of the fragments was already 300 nucleotides. The driver DNA was about 450 nucleotides in length. The lower frequency components are missing or greatly suppressed in the experiments shown in Fig. 2 due to the Cot 4 fractionation which was used in the preparative steps. The long tracer preparation appears to be a little less reactive than the short tracer preparation in Fig. 3.

The repetitive frequencies represented in the long and short preparations are clearly very similar. However the actual frequencies are difficult to assess due to the limited resolution of the second order kinetic measurements. There appear to be components which are half reassociated in the range of Cot 0.3 and others half reassociated around Cot 10. Thus repetition frequencies of around 3000 copies and around 100 copies per haploid genome occur in both the long and short repetitive fractions.

For the purpose of calculation we have chosen to represent these components with reassociation rate constants $k = 3\ M^{-1}sec^{-1}$ and $k = 0.1\ M^{-1}sec^{-1}$. Other components could be used for the analysis. The somewhat arbitrary choice of components,

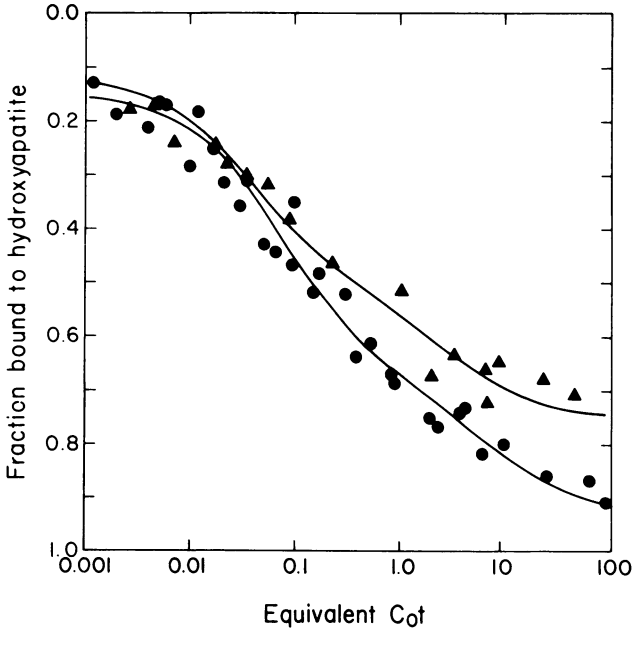

Figure 3. Self Reassociation of Long and Short Repetitive DNA Fractions. ^{3}H-labeled and unlabeled long and short repetitive DNA fractions were prepared as in Fig. 1 and the long repetitive fractions were sheared to 300 ntp. Tracers and drivers of the same sequence length class were mixed, denatured and reassociated in 0.12 M PB at 60°C or equivalent, and the duplex fraction was measured by binding to hydroxyapatite. (▲) long repetitive driver plus long repetitive tracer; (●) short repetitive driver plus short repetitive tracer. Table 1 shows the least squares solutions of the kinetics of reassociation.

however, does not affect the following conclusions. That is, the frequency components chosen represent reasonably well both the early part of the total DNA reassociation curve and the long and short tracer curves in Fig. 2. Least squares analysis of the reaction of the short tracer with the whole DNA in Fig. 2 shows that 51% of the reaction can be attributed to the assumed fast component ($k = 3\ M^{-1}sec^{-1}$) and 26% is due to the assumed intermediate component ($k = 0.1\ M^{-1}sec^{-1}$). The same sort of analysis shows that 39% and 25%, respectively, of the long tracer reassociation can be attributed arbitrarily to the fast and intermediate components. Such arbitrary components form a convenient method of calculation.

As mentioned above the first 51% of the reassociation of the short repetitive sequence preparation can be described by an average rate constant of $3\ M^{-1}sec^{-1}$ (driven by whole DNA). When this short DNA is allowed to reassociate by itself (Fig. 3) we ask what rate constant can be applied to the first 51% of that reaction and from this value we may calculate the amounts of different sequence with this average rate constant. We use the word "complexity" to refer to these estimates of amount of different sequence even though they may not represent the total complexity of repetitive DNA sequences. The self-reassociation data shown in Fig. 3 permit the precise calculation of the amount of different sequence present if the average "fast" and "intermediate" components accurately represent the repetitive DNA. We make use of the estimates derived from Fig. 2 for the quantity of these components in the long and short tracer preparations. Table 1 shows the rate constants calculated from the data of Fig. 3 when the quantities of the fast and intermediate components are fixed at these values. The results of this least squares analysis suggest that for both rate components the complexity of the short repetitive fraction is about twice that of the long repetitive class. Since the genome contains twice as much short repetitive DNA as long, this result is expected if the two classes of DNA are separate and independent sets of sequences. However further measurements were needed to establish this result; therefore the long and short fractions were reassociated with each other, as described in the text section.

Limited Reassociation between Long and Short Repetitive Sequences

Labeled short fragments were reassociated with a thousand-fold excess of long unlabeled driver which had been sheared to 300 nucleotides. The results are shown in Fig. 4. For comparison, Fig. 4 also shows the curve from Fig. 3 for the reaction of short repetitive ^{3}H-DNA tracer with short repetitive driver DNA. The least squares solutions for the kinetic components are given in Table 2. It is clear that the rate of reassociation of the short preparation tracer with long preparation driver is much reduced compared to the homologous tracer-driver reactions. This low rate of reassociation is a measure of the low concentration of sequence homologous to the repetitive sequences in the long repetition DNA fraction. The fact that the unlabeled long DNA fraction

Table 1. Apparent Kinetic Complexity of Long and Short Repetitive DNA Sequence Fractions

Component	Repetitive fraction	Fraction of component[b]	k (M^{-1} sec^{-1})[c]	k_{pure}[d]	Apparent complexity (ntp)[e]
Fast	Long[a]	0.385	26.7	69.4	1.3×10^4
	Short	0.515	14.2	27.6	3.3×10^4
Intermediate	Long[a]	0.248	0.39	1.55	5.8×10^5
	Short	0.260	0.18	0.672	1.3×10^6

[a] Sheared to 300 ntp.
[b] From analysis of the data of Fig. 2, see text.
[c] Fraction of each component was fixed at the values determined from the experiment shown in Fig. 2, and the best reassociation rate constants were determined by computer least squares analysis.[20] RMS deviations were 3.5% and 4.1% for long and short repetitive fractions, respectively.
[d] k_{pure} = observed k divided by the fraction of component.
[e] Graham et al.[4] presented an analysis of the kinetic components in sea urchin DNA which were similarly based on the arbitrary choice of "fast" and "intermediate" repetitive sequence classes. The calculated complexities reported were 2.4×10^4 and 1.8×10^6 ntp, respectively.

is an effective driver for such reassociation reactions is shown by the curve in Fig. 3 in which the long repetitive ^{3}H-DNA tracer reassociates with long repetitive driver DNA.

The least squares solutions shown in Table 2 indicate a 15- to 20-fold lower rate for the short repetitive ^{3}H-DNA tracer as compared to the long repetitive DNA driver. We may compare the rate of reassociation of the short tracer with long driver with that of the short tracer and its own labeled short driver preparation. The ratios in this case are 0.12 for the intermediate component and 0.14 for the fast. From these calculations we estimate that between 5 and 15% of the long repetitive DNA preparation is made up of sequences homologous to the short repetitive DNA preparation. However, the small portion of the long preparation which is homologous with the short duplex tracer appears to contain nearly all of the sequences in the latter since the reaction approaches completion. We have no way of knowing whether the small homologous fraction is due to cross contamination or to the occurrence in the genome of representatives of all of the short repetitive sequence families within the long repetitive regions.

The Existence of Precise Repeats without Divergent Homologues

As previously reported[13] long repetitive sequences are relatively precisely repeated

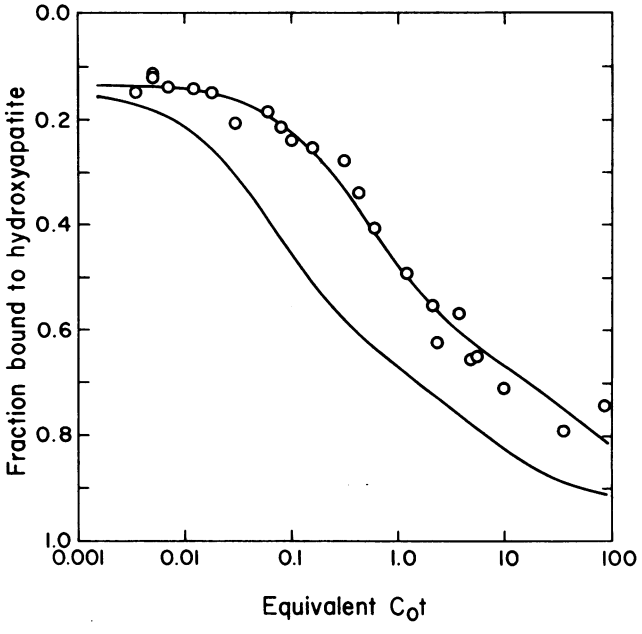

Figure 4. Reassociation of Labeled Short Repetitive Sequences with Long Repetitive DNA Driver. The same long repetitive DNA driver preparation and short repetitive ^{3}H-DNA tracer preparation as used for the experiment in Fig. 3 were mixed in a ratio of at least 150/1 and reassociated in 0.12 M PB at 60°C (o). The left hand curve is reproduced from Fig. 3 and shows the reaction of the short repetitive ^{3}H-DNA with short repetitive DNA driver.

Table 2. Sequence Homology Between Long and Short Repetitive DNA Sequence Fractions

Component	Fraction[a]	k_d[b]	k_t[c]	Percent homology[d]
Fast	0.515	26.7	2.02	7.6
Intermediate	0.260	0.385	0.022	5.7

[a] See text for estimate of the fraction of each component in short repetitive tracer.
[b] Rate constants for long tracer or long repetitive driver (Table 1).
[c] Rate constants for tracer were determined with quantities of components fixed at the values in column 2. RMS deviation for the fit was 3%.
[d] $k_t/k_d \times 100$.

while short repetitive sequences are relatively divergent. This correlation has been observed in the DNA of many species[7-9, 22-24], and is a general characteristic of repetitive DNA sequences. Since the data reported above show that the long and short repetitive sequence classes consist principally of separate nonhomologous sets of sequence we expect that high and low precision repetitive sequences would also be separate sets

of sequences. In other words high thermal stability repetitive fractions should not reassociate to any great extent with low thermal stability fractions of repetitive DNA. A simple and direct test of this idea is to prepare a high thermal stability repetitive sequence fraction, reassociate it with total DNA in great excess, and measure the thermal stability of the resultant duplexes. The thermal stability of these duplexes will be reduced if any measurable fraction of the high thermal stability repetitive sequence reassociates with divergent but homologous sequences.

To perform this test a high thermal stability fraction was prepared as described in Table 3. The first four steps consist merely of repeated incubations to repetitive Cot and isolation of the fraction of the DNA which remains duplex above 80°C on hydroxyapatite in 0.12 M PB. At this stage a fraction of the DNA was saved for driver and another fraction was labeled by "gap translation" with E. coli DNA polymerase I to 3 x 10^6 cpm/μg. The "foldback" fraction was removed from the tracer (step 5) and the high thermal stability fraction again isolated. The fragment sizes at this stage were (in nucleotides): high thermal stability driver, 980; gap translated high thermal stability tracer, 580; total DNA driver, 300.

Table 3. Isolation of High Thermal Stability Repetitive DNA Fraction

Fractionation[a]	Cot	% Eluted 60-80°C[b]	% Eluted above 80°C[b]	% of original DNA eluted >80°C
1	40	43	22	21.9
2	4	23	33	7.2
3	4	25	56	4.0
4	4	14	75[c]	3.0
5 (labeled DNA)	0.001 (84% unbound)			
6	4.9	28	33 (used for 7 and 8)	
7[d]	4 (high precision driver)	24	63	
8[d]	4 (total DNA driver)	21	64	

[a] For each fractionation step, DNA was denatured and reassociated in 0.12 M PB at 60°C. Duplexes were collected on hydroxyapatite at 60°C; the temperature was increased to 80°C and the column was washed with 0.12 M PB. DNA not denatured at 80°C was eluted with 0.4 M PB and used for the subsequent fractionation step. The percentages are averages of radioactivity and A_{260} assay which were closely concordant. The initial mixture consisted of 10 mg of S. purpuratus sperm DNA and about 10 mg (500,000 cpm) of S. purpuratus embryo DNA sheared to about 1000 nucleotides.
[b] Fraction of DNA incubated at each stage.
[c] A part of this fraction was "gap translated" to high specific radioactivity, and the remainder was used for driver in the final measurement.
[d] Hydroxyapatite thermal chromatogram at 3°C temperature steps.

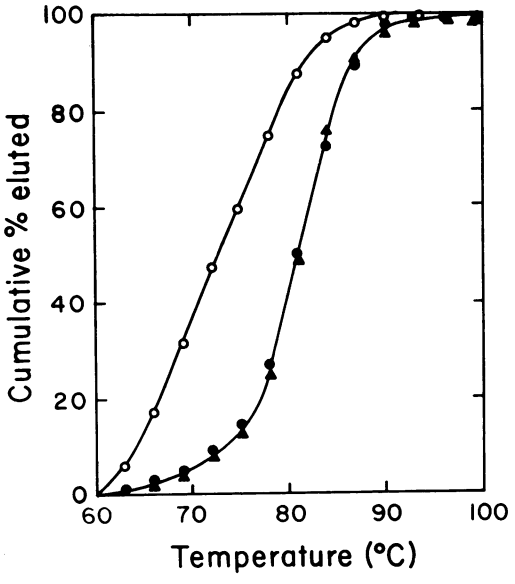

Figure 5. Absence of Reassociation Between High Thermal Stability Repetitive DNA and Divergent Related Sequences. A high thermal stability repetitive fraction of sea urchin DNA was prepared by successive low Cot incubations and isolation of the fraction eluting from hydroxyapatite above 80°C at each step (Table 3). The purified high thermal stability fraction (3% of the genome) was labeled in vitro with E. coli DNA polymerase I to a specific activity of 3 x 10^6 cpm/μg. The high thermal stability tracer was reassociated to Cot 4 either with homologous driver or with total DNA in excess. Duplexes were collected on hydroxyapatite at 60°C in 0.12 M PB and fractionated by thermal chromatography. For the reaction with total driver DNA, both the tracer radioactivity and the optical density of the driver were followed. (o) optical density of total driver DNA; (●) high thermal stability ^{3}H-tracer x homologous driver; (▲) high thermal stability ^{3}H-tracer x total driver.

Fig. 5 shows the thermal stability of the products of reassociation when this driver was reacted with total sea urchin DNA or the high thermal stability driver DNA from step (4) of Table 3. There is no detectable difference in thermal stability. Fig. 5 also shows the distribution of Tm of the total repetitive sequences which reassociated by Cot 4. The Tm of this DNA is about 11°C below that of the high thermal stability fraction.

These observations show directly that the precisely repeated DNA fraction does not reassociate with any measurable quantity of homologous DNA sequence of greater divergence. We conclude that few if any low precision homologues of these high precision repeats exist in the genome.

The high thermal stability fraction includes most of the total precisely repeated DNA sequence in the sea urchin genome. The actual yield at fractionation step 4 (Table 3) is 3% of the genome and probably some loss has occurred. Britten et al.[13] estimate that about 8% of the genome consists of long repeated sequences which reassociate

by Cot 20. Fig. 1 indicates that less of the genome is recognized at Cot 4 as long repetitive sequences (6% compared to 8% at Cot 20). Most of these sequences are of high thermal stability. Considering the preparative losses we feel that the high thermal stability fraction examined in this measurement includes more than half of all of the long high precision repetitive sequences of the sea urchin genome.

DISCUSSION

The experiments described here show that most of the long repetitive sequences which occur in the sea urchin genome differ from most of the short interspersed repetitive DNA sequences. There is evidence for some overlap between the long and short repetitive DNA preparations used in this work but we do not know whether this is due to cross contamination or to real aspects of sequence organization. For example a few copies of the short interspersed sequences could exist in tandem array or could be surrounded by other repeated sequences in long repetitive sequence elements. The precisely repeated DNA fraction isolated on the basis of its thermal stability lacks detectable divergent relatives in total sea urchin DNA.

We have discussed at length elsewhere some possible functional roles for the short interspersed repetitive sequences.[25-27] However, relatively little attention has so far been focused on the long repetitive sequence class. The main characteristic of the long repetitive sequence class is the precision of the repeats, indicated by the high thermal stability of their renaturation products. The long repetitive sequence class is highly complex, including a great many different individual sequences (Table 1). Therefore it is not simply a collection of a few satellite DNAs of low complexity. Satellites are not known in S. purpuratus DNA. Most of the examples of satellites known in other species would be classified as long precisely repetitive sequences, although their complexity is much less than the long repetitive sequences described here.

We do not know if the bulk of the long repetitive sequences are precisely repeated owing to sequence-specific function. The long repetitive sequences could either be restrained from evolutionary sequence divergence by selective pressure, or they could be nondivergent simply because they are of recent origin, or both. A long repetitive sequence set which provides an example of the first alternative is the repetitive DNA family constituting the ribosomal genes, and an example of the second alternative is provided by the species specific ribosomal spacer DNAs.[28-31] Sequences belonging to the latter class could be maintained as a set of relatively precise repeats within a species, through special mechanisms such as rapid unequal crossover. By this means new variants may be multiplied and substituted for the tandem arrays of preexistent variants.[32] Since sequences of this sort seem to be fairly widespread[29-40] they must perform some function, but this function is evidently not sequence specific in the same sense as are most transcribed gene sequences. The long repetitive sequence class of

sea urchin DNA surely includes ribosomal and histone genes and also the spacer DNA belonging to these structural genes. However our measurements of repetition frequency and sequence complexity show that the known examples of repetitive structural genes and spacers can constitute only a minor fraction of the long repetitive sequence class in sea urchin DNA. Thus in our present state of knowledge we cannot identify or ascribe function to most of the sequences in this class.

We are faced with two possibilities for the evolutionary relationship of the long and short repetitive sequence classes. They could evolve in complete independence of each other. That is, the evolutionary appearance and disappearance from the genome of long repetitive sequences and of short interspersed repetitive sequences could be unrelated sets of events, and might occur at widely different rates. On the other hand, the short interspersed repetitive sequences might derive during evolution from long repetitive sequences by processes of translocation, followed by sequence divergence.[14, 16, 25, 26] The present measurements imply an important restriction on their evolutionary relationship. If short repetitive sequences descend from long repetitive sequences, the rate of breakup and translocation of the long repetitive sequence subelements must be much higher than the basic rate of sequence divergence. Otherwise we would expect to see widespread sequence homology between short and long sequence classes, particularly of inexactly matched sequences.

Our measurements show that the long repetitive sequences are in general little diverged, suggesting that they are either newly formed or are subject to strong selection. In contrast the short repetitive sequences are considerably divergent and in general are unrelated to the long repetitive sequences. Thus for long repetitive sequences to serve as ancestors of short ones the following type of process would have to be envisioned: new long repetitive sequence elements would appear, probably in some form of tandem array, and some of these would be very quickly broken up and distributed around the genome, where they subsequently diverge at the usual rate. This process would occur at a rate sufficient to account for the appearance of new interspersed repeats, and would result in the disappearance of the ancestral long repetitive sequence. Thus at any given time the extant long repetitive sequence class would consist of long repetitive sequences which might serve in the future, as progenitors of short repetitive sequences, plus any others which have special gene or spacer functions. This model has testable implications. It predicts a rapid flow in and out of the long repetitive sequence class, and species specific distribution for the ancestor class of long repetitive sequences. It seems certain that important insights into the nature of genomic evolution will derive from further analysis of this class of repetitive sequences.

ACKNOWLEDGEMENTS

This research was supported by a grant from the National Science Foundation,

BMS 75-07359, and by grants from the National Institutes of Health HD 05753 and GM-20927. FCE was the recipient of a fellowship from the American Cancer Society. DEG was supported by fellowship grants from the Damon Runyon-Walter Winchell Cancer Fund and the Carnegie Institution of Washington.

REFERENCES

1 Present address: National Cancer Institute, NIH, Bethesda, Md. 20014
2 Present address: Department of Biology, Purdue University, Lafayette, Indiana 47907
3 Also Staff Member, Carnegie Institution of Washington, Washington, D.C. 20005
4 Graham, D. E., Neufeld, B. R., Davidson, E. H. and Britten, R. J. (1974) Cell 1, 127-137
5 Davidson, E. H., Hough, B. R., Amenson, C. S. and Britten, R. J. (1973) J. Mol. Biol. 77, 1-23
6 Chamberlin, M. E., Britten, R. J. and Davidson, E. H. (1975) J. Mol. Biol. 96, 317-333
7 Goldberg, R. B., Crain, W. R., Ruderman, J. V., Moore, G. P., Barnett, T. R., Higgins, R. C., Gelfand, R. A., Galau, G. A., Britten, R. J. and Davidson, E. H. (1975) Chromosoma 51, 225-251
8 Crain, W. R., Davidson, E. H. and Britten, R. J. (1976) Chromosoma, in press
9 Efstratiadis, A., Crain, W. R., Britten, R. J. and Davidson, E. H. (1976) Proc. Nat. Acad. Sci., USA 73, 2289-2293
10 Angerer, R. C., Crain, W. R., Hough-Evans, B. R., Galau, G. A., Klein, W. H., Davis, M. M., Wold, B. J., Smith, M. J., Britten, R. J. and Davidson, E. H. (1976) in The Molecular Biology of the Mammalian Genetic Apparatus, Vol. II, Part A, Assoc. Scientific Publishers, in press
11 Manning, J. E., Schmid, C. W. and Davidson, N. (1975) Cell 4, 141-155
12 Crain, W. R., Eden, F. C., Pearson, W. R., Davidson, E. H. and Britten, R. J. (1976) Chromosoma 56, 309-326
13 Britten, R. J., Graham, D. E., Eden, F. C., Painchaud, D. M. and Davidson, E. H. (1976) J. Molec. Evol., in press
14 Britten, R. J. and Kohne, D. E. (1968) Science 161, 529-540
15 Britten, R. J. and Davidson, E. H. (1976) Fed. Proc. 35, 2151-2157
16 Britten, R. J. and Kohne, D. E. (1967) Carnegie Inst. Wash. Year Book 65, 78-106
17 Noll, H. (1967) Nature 215, 360-363
18 Britten, R. J., Graham, D. E. and Neufeld, B. R. (1974) in Methods in Enzymology, Vol. 29, Part E, pp. 363-418, Academic Press, New York
19 Ando, T. (1966) Biochim. Biophys. Acta (Amst.) 114, 158-168
20 Schachat, F. H. and Hogness, D. S. (1973) Cold Spring Harbor Symp. Quant. Biol. 38, 371-381
21 Galau, G. A., Klein, W. H., Davis, M. M., Wold, B. J., Britten, R. J. and Davidson, E. H. (1976) Cell 7, 487-505
22 Angerer, R. C., Davidson, E. H. and Britten, R. J. (1975) Cell 6, 29-39
23 Pearson, W. R., Personal Communication
24 Davidson, E. H., Graham, D. E., Neufeld, B. R., Chamberlin, M. E., Amenson, C. S., Hough, B. R. and Britten, R. J. (1973) Cold Spring Harbor Symp. Quant. Biol. 38, 295-301
25 Britten, R. J. and Davidson, E. H. (1969) Science 165, 349-358
26 Britten, R. J. and Davidson, E. H. (1971) Quart. Rev. Biol. 46, 111-138
27 Davidson, E. H., Klein, W. H. and Britten, R. J. (1977) Dev. Biol. 55, in press
28 Brown, D. D., Wensink, P. C. and Jordan, E. (1971) Proc. Nat. Acad. Sci., USA 12, 3175-3179
29 Brown, D. D., Wensink, P. C. and Jordan, E. (1972) J. Mol. Biol. 63, 57-73
30 Wellauer, P. K., Reeder, R. H., Carroll, D., Brown, D. D., Deutch, A., Higashinakagawa, T. and Dawid, I. B. (1974) Proc. Nat. Acad. Sci., USA 71, 2823-2827

31 Wellauer, P. K. and Reeder, R. H. (1975) J. Mol. Biol. 94, 151-161
32 Smith, G. P. (1973) Cold Spring Harbor Symp. Quant. Biol. 38, 507-513
33 Patterson, J. B. and Stafford, D. W. (1971) Biochemistry 10, 2775-2779
34 Clarkson, S. G., Birnstiel, M. L. and Purdom, I. F. (1973) J. Mol. Biol. 79, 411-429
35 Birnstiel, M., Telford, J., Weinberg, E. and Stafford, D. (1974) Proc. Nat. Acad. Sci., USA 71, 2900-2904
36 Gall, J. G. and Rochaix, J. D. (1974) Proc. Nat. Acad. Sci., USA 71, 1819-1823
37 Forsheit, A. B., Davidson, N. and Brown, D. D. (1974) J. Mol. Biol. 90, 301-314
38 Kedes, L. H., Cohn, R. H., Lowry, J. C., Chang, A. C. Y. and Cohen, S. N. (1975) Cell 6, 359-369
39 Schaffner, W., Gross, K., Telford, J. and Birnstiel, M. (1976) Cell 8, 471-478
40 Clarkson, S. G. and Kurer, V. (1976) Cell 8, 183-195