

Long and Short Repeats of Sea Urchin DNA and Their Evolution

Gordon P. Moore¹, William R. Pearson², Eric H. Davidson and Roy J. Britten³

Division of Biology, California Institute of Technology, Pasadena, California 91125, U.S.A.; present addresses: ¹ Division of Biology, University of Michigan, Ann Arbor, Michigan 48109, USA.; ² Department of Microbiology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205; ³ Also staff member, Carnegie Institution of Washington, Washington, D.C. 20015, U.S.A.

Abstract. Repeated sequences cloned from the DNA of the sea urchin *S. purpuratus* were used as probes to measure the lengths of individual families of repeats. Some probes reassociated much more rapidly with preparations of long repeats than with short repeats while others reassociated more rapidly with short repeats than with long repeats. In this way two of five cloned repeats were shown to represent families with a great majority of sequences in the long class. One represented a family with similar numbers of long and short class members. Two were members of predominantly short class families. – The cloned repeats representing long class families, formed more precise duplexes than those representing short class families. Thermal stability measurements using *S. purpuratus* or *S. franciscanus* driver DNA showed that precise repetitive sequences have as great an interspecies sequence difference as the less precise repeats. Thus the precision of many families may result from recent multiplication rather than from selective pressure on the DNA sequences. Measurements of evolutionary frequency change show a clear correlation between the frequency change and the size of families of repeats in *S. purpuratus*. Comparison with *S. franciscanus* indicates that many of the large size families in *S. purpuratus* are those that have grown in size since these two species diverged.

Introduction

Repetitive sequences occur in all eukaryotic DNA. Although they undergo active evolutionary changes, the quantities and frequency distribution of repeats appear to be conserved. For example two sea urchins from different orders, *Lytechinus pictus* and *S. purpuratus* have about the same amount of repetitive sequence and the frequency distributions are similar. However, during the approximately 175 million years since they shared a common ancestor most of the repeated sequences appear to have been replaced (Moore et al., 1978). In other words, individual large families of repeats in the genome of each species are homologous to only a few repeat sequences in the genome of the other

species. Thus overall patterns of quantity of repeats and repetition frequencies are unchanged while individual repeats are replaced and different families become dominant.

Both short and long repeated sequences occur universally in eukaryotic genomes. The majority of the repetitive sequences of the DNA of the sea urchin *S. purpuratus* (*Sp*) are short, averaging about 300 nucleotide pairs in length, and are interspersed with single copy DNA (Graham et al., 1974). Long repeats make up about a third of the mass of the repetitive DNA, that is they reassociate to form duplex structures greater than 2,000 nucleotides (nt) in length which are resistant to digestion with S1 nuclease. The two classes may be prepared by reassociating moderately long DNA fragments to repetitive C₀t, digesting with S1 nuclease and separating them by exclusion chromatography (Britten et al., 1976). Measurements of long and short repeats have been made with this method for the DNA of a large number of eukaryotes (Davidson et al., 1975; Goldberg et al., 1975). The short class of repeats makes up the majority of repetitive sequences in most species that have been studied. The primary exceptions are certain diptera including *Drosophila* in which only a few short repeats are observed (Manning et al., 1975; Crain et al., 1976). Short repeats have not so far been seen in large quantities in the DNA of birds although measurements of low frequencies and imprecise relationships have yet to be made (Eden and Hendrick, 1978; Eden et al., 1978). Some short repeats are clearly present in the chicken genome and examples have been described in the region of a structural gene (Cochet et al., 1979).

Members of many repetitive sequence families are expressed in sea urchin egg RNA and nuclear RNA (Costantini et al., 1978). Repeat sequence transcripts are often present on long RNA molecules which include single copy sequences (Costantini et al., 1980). Both complements are present in the RNA, presumably as a result of asymmetric transcription of different members of a family of repeats. The number of copies present in the RNA varies widely from family to family and shows very large changes during development or differentiation (Costantini et al., 1978; Scheller et al., 1978). Transcription is observed for members of families which are predominantly long as well as for families that are predominantly short.

Previous measurements (Eden et al., 1977) have indicated that sea urchin long and short repetitive sequence families are principally non-overlapping. Only 10% of a long repeat DNA preparation reassociated with short repeats. It was not shown whether this 10% cross reaction was due to imperfect purification of the two fractions or due to the presence of short repeats in the genome which are homologous to long class families. In every species that has been examined the long class reassociates to give duplexes of relatively high thermal stability, implying that long repeat families are made up of sets of quite similar sequences. On the other hand most short repeats reassociate to yield duplexes of low thermal stability even after correction has been made for the duplex length.

Long repeats are defined as the sets of sequences which form long duplex regions after reassociation to repetitive C₀t and therefore there are a number of possibilities for their sequence organization. They may consist of extended

regions of precise tandem repetition or sets of long precisely homologous sequences scattered through the genome. There are probably also regions which consist of sets of shorter repeat sequences occurring in different (scrambled) arrangements (Wensink et al., 1979; Scheller et al., 1981).

Several of the repetitive sequence families examined in this paper have also been studied by screening DNA inserts from a recombinant lambda library of *S. purpuratus* (Anderson et al., 1981; Scheller et al., 1981). Restriction mapping, gel blots, reassociation kinetics, and electron micrographic heteroduplex analysis have yielded significant information about the sequence organization of the appropriate neighborhoods in the sea urchin genome. The sequences homologous to the repeat probes CSp2109 A and B appear to occur in environments of single copy DNA and thus clearly represent short class families (Anderson et al., 1981). The long repeat family represented by CSp2108 has an unexpected pattern of sequence relationships (Scheller et al., 1981). Individual members of this family have precise homology with a small number of other members for a length of about 5 kilobases (kb) and imprecise homology with a large set of other members. It appears that there is a "superfamily" containing a moderate number of different "subfamilies" each of which consist of 5 to 35 precisely related sequences. Recombinant λ library inserts containing members of the long class family CSp2034 often contain more than one homologous repeat. Thus members of this family occur in clusters in the genome. Other members are more widely spaced, but not as far apart as random expectation. These recent studies show that the repeats examined here are examples of a variety of different patterns of sequence organization.

Materials and Methods

1. Isolation of Long and Short Classes of Repetitive DNA. *Sp* DNA was extracted from sperm, sheared to 2,000 nt and sized as previously described (Britten et al., 1974). Long and short classes of repetitive DNA were prepared by a modification of the method of Eden et al. (1977). Fifteen mg of the sheared DNA was reassociated to C_{ot} 40 in 0.3 M NaCl, 0.01 M PIPES, pH 6.7. One volume 0.05 M sodium acetate, 0.0002 ZnSO₄, pH 4.5 was added and the mixture was adjusted to 0.025 M dithiothreitol (BioRad), and 0.1 ml/ml S1 nuclease prepared as described by Ando (1966). The mixture was incubated 2 h at 35° C so that about 80% of single strands were digested (Britten et al., 1976). The digestion products were chromatographed on a Sephadex G-100 column and 19% of the input DNA was excluded. The G-100 excluded material was chromatographed on a 70 ml Sepharose CL 2B column in 0.2 M sodium acetate at room temperature as shown in Figure 1. Excluded DNA (i.e., >800 nt) and chromatographed DNA (<400 nt) was pooled, concentrated and constituted the long and short classes of driver DNA used in the subsequent measurements. Final yields were 368 μ g or 32% of the recovered DNA in the long (excluded) class and 730 μ g or 68% in the short class. To eliminate the effect of driver length on the rate of reassociation (Britten et al., 1974), the long class repeat DNA was sheared at 50,000 rpm in 66% glycerol, 10⁻³ M EDTA, 0.01 M sodium acetate in an ethanol dry ice bath using a Virtis-60 homogenizer to a length of 300-400 nt.

2. Preparation, Labeling and Hybridization of Cloned Repeat Sequences. The individual cloned repetitive DNA sequences were those reported previously (Scheller et al., 1977) and further studied in some detail (Costantini et al., 1978; Klein et al., 1978; Scheller et al., 1978, 1981; Anderson et al., 1981; Posakony et al., manuscript submitted). They were prepared by C_{ot} 40 reassociation of total *Sp* DNA, S1 treatment, ligation of synthetic *Eco*RI decamers and cloning at the *Eco*RI site of the bacterial plasmid RSF2124. As described in the above references the cloned repeat

sequences were labeled with ^{32}P by the kinase method and strand separated to prevent tracer self reaction. Reassociation conditions and methodology of the hydroxyapatite (HAP) analysis were as described in Moore et al. (1978) and in the legend of Figs. 1 and 2.

3. *Blots.* *Sp* and *S. franciscanus* (*Sf*) DNA of length greater than 100 kb (as measured in the electron microscope) was prepared using a modification of the method of Blin and Stafford (1976). This DNA was digested with an excess of EcoRI (Boehringer-Mannheim). Twenty-five micrograms of digested DNA was placed in each track of 0.7% agarose vertical slab gels.

Parallel tracks contained an EcoRI digest of the λ phage Charon 4A and a HaeIII digest of the bacterial plasmid RSF2124 as molecular weight standards. The gels were stained with ethidium bromide and photographed. DNA was transferred from the gel to a nitrocellulose filter as described by Southern (1975). Filters were baked, washed, and prehybridized as described. Filters were hybridized overnight with individual cloned *Sp* repeats which had been prepared and labeled with ^{32}P . After hybridization the filters were washed, air dried and autoradiographed 1–7 days at -70°C using preflashed Kodak X-ray film and intensifying screens.

4. *Preparation of Low Divergence Repeat DNA Tracer.* *Sp* embryos were radioactively labeled *in vivo* with ^3H -thymidine and DNA isolated as described (Hinegardener, 1967; Moore, 1978). The specific radioactivity was 1×10^5 cpm μg^{-1} and the length was 610 nt as determined by alkaline sucrose velocity sedimentation. This tracer was twice reacted to C_{0t} 200 and bound to HAP to isolate repetitive DNA (20% of total DNA). Three percent of the DNA was removed as foldback by denaturation and rapid passage over HAP. About 2 μg of this repetitive tracer was precipitated and resuspended in 2.4 M tetraethyl ammonium chloride (TEACL). The T_m of precise duplexes 500 nt long is 61°C in 2.4 M TEACL. Therefore, repetitive duplexes which melted within 2°C of precise duplexes were selected by reassociating the tracer to C_{0t} 20 at 45°C (C_{0t} 100) followed by heating to 59°C for $1/2$ h. The sample was rapidly diluted 40-fold into 0.12 M phosphate buffer (PB) and passed over HAP at 60°C . The effective C_{0t} after melting was about 0.03 and some high frequency lower melting repeats could be present in the preparation. The bound fraction (17%) was eluted with 0.4 M PB, concentrated with butanol, passed over G-100 and precipitated with ethanol. The precipitate was resuspended in 10^{-4} M EDTA, brought to 2.4 M TEACL and heated again to 59°C for $1/2$ h. It was passed over HAP after a 40-fold dilution into 0.12 M PB. The final bound fraction (low divergence tracer) was 10% of the repetitive tracer or about 2% of the genome.

Results

1. Reassociation of Cloned Repeat Sequences and Long or Short Driver DNA

Long and short driver DNAs were prepared from unlabeled *Sp* sperm DNA as described in Materials and Methods. The insert in Figure 1 shows the effective separation of the two S1 resistant length classes. A large excess of these DNA preparations was reassociated with strand separated, labeled, cloned DNA fragments representing individual families of repeats. Reassociation was assayed by HAP as described in Materials and Methods. Figure 1 shows the kinetics of reassociation of two of the cloned fragments, CSp2101 and CSp2034. Clearly CSp2101 reassociates very much more rapidly with the short class than it does with the long class repetitive DNA. In contrast CSp2034 reassociates more rapidly with the long class than the short class DNA. These observations show that most members of the family represented by CSp2101 are in the short class and most members of the CSp2034 family are in the long class. Table I summarizes the results for the five cloned probes that were tested in this way and lists an estimate of the percentage of members of each family in the two classes.

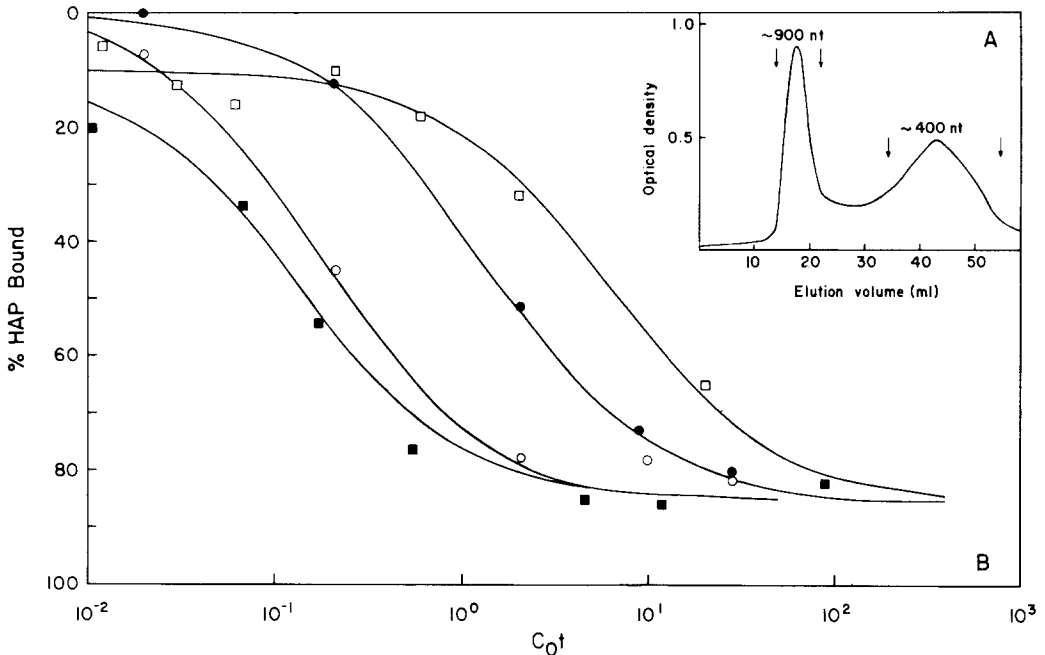


Fig. 1A and B. Reassociation of cloned repeat sequences CSp2101 and CSp2034 with the long and short classes of repetitive *Sp* DNA. **A** Chromatography of S1 treated repetitive DNA on Sepharose 2B. Total *Sp* DNA was sheared to 2,000 nt, reassociated to C_0t 40, treated with S1 nuclease and chromatographed on Sepharose 2B. The fractions taken as long and short classes of repetitive DNA are shown by arrows. **B** Cloned tracers reassociated with a 10^5 -fold weight excess of driver DNA: CSp2101: \circ short class, \bullet long class, and CSp2034: \square short class, \blacksquare long class. Reassociation was in 0.12 M or 0.41 M PB with 0.05% SDS and 10^{-3} M EDTA at 50°C with approximately 50 μg of driver and 500 cpm of tracer. The samples were diluted to 0.12 M PB and the fraction of molecules containing duplex regions was determined by chromatography on 0.4 g of HAP at 50°C . Unbound material was eluted with 0.12 M PB. The temperature was raised to 98°C and the bound DNA was eluted with 0.12 M PB. The curves shown are the least squares solutions assuming a single second order kinetic component

The kinetic data summarized in Table 1 shows the large majority of the members in two families (CSp2034 and CSp2108) to be in the long class and the majority of the members of the other two families (CSp2101 and CSp2137) in the short class. The large ratios between the rates of reassociation with the two classes in the most extreme case shows that good separation was achieved in the preparation of the long and short class driver DNAs. The ratio of rate constants is 45 for CSp2137, indicating that not more than 2% of the long driver preparation is made up of contaminating short class repeats. Thus any short repeat family which shows a larger fraction of long repeats than CSp2137 probably has some members which actually occur as long repeats. Such a conclusion appears certain for the CSp2007 family which shows almost as many members in the short class as in the long class. In the case of the CSp2102 family about 5% of the members appear to be in the long class. The data described in the next section support this conclusion for the CSp2101 family. It appears that repeat families usually fall predominantly into the short or long class and may have few or no members in the other class.

Table 1. Estimate of membership of repeats in long or short class by reassociation kinetics

Clone ^a	Reassociation rate constant ^b (M ⁻¹ s ⁻¹)		Percentage in length class ^c
	Long	Short	
CSp2034	7.4	0.16	95% Long
CSp2108	0.22 ^d	low ^e	>90% Long ^d
CSp2007	2.1	0.7	60% Long ^f
CSp2137	0.085 ^g	3.9	98% Short
CSp2101	0.83	5.5	92% Short

^a The characteristics of these clones derived from previous work are listed in Table 2

^b Second order rate constants by least squares analysis from measurements such as those shown in Figure 1

^c Calculated from estimates of the number of members in long class and short class using the equations: N (long)=K (long)×F (long)/K (single copy) or N (short)=K (short×F (short))/K (single copy). N (long) and N (short) are the respective calculated numbers of family members. K (long) and K (short) are the measured reassociation rate constants. F (long)=0.070 and F (short)=0.12 are the fraction of the genome in the long and short repeat classes estimated from the preparation of the driver DNAs. K (single copy)=0.00125 is the reassociation rate for single copy DNA as observed for this fragment size

^d Recent observations (Scheller et al., 1981) show that there are many sequences in the *S. purpuratus* genome which show low thermal stability sequence homology to CSp2108. These low thermal stability relatives apparently do not have an effect on the observed kinetics of reassociation. The data in the table and the classification are determined by the small set of long repeats which are closely related to the CSp2108 probe

^e The reaction did not go to completion and a rate constant cannot be accurately derived

^f The uncertainty is such that the number of members could easily be equal in the two classes

^g This slow reaction could be due to cross contamination

2. Repeat Family Organization Tested by Restriction Digests

The size distribution of fragments resulting from restriction enzyme digestion supplies information about the sequence organization of repeat families. For example a region of perfect tandem repetition has regularly spaced restriction sites. Repeated families with members in precise tandem regions show prominent bands after gel electrophoresis of appropriate restriction digests. Bands may also be observed for families whose members have two or more "internal" restriction sites in a precisely repeated sequence even though they occur in a variety of flanking sequences. Restriction fragments terminated by sites in a variety of flanking sequences naturally show a variety of lengths. Thus typical short interspersed sequence families yield broad distributions of fragment lengths for all restriction enzymes while long precisely repeated sequences yield one or a few bands for many restriction enzymes. Families which are predominantly short and interspersed but contain a few members in long precisely repeated regions are expected to show a small amount of DNA in bands while the majority will be in a broad distribution.

Tests of the restriction patterns of particular families were made by transferring a digest of total sea urchin sperm DNA to nitrocellulose and incubating

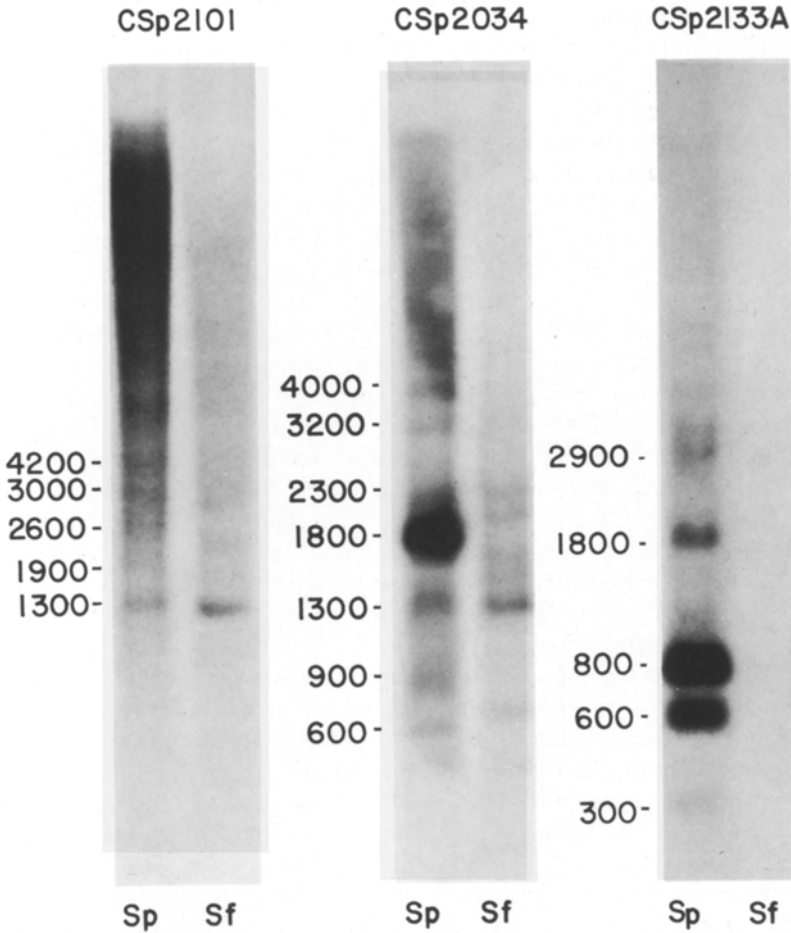


Fig. 2. Gel blot hybridization of three cloned *Sp* repeats with a total EcoRI digest of *Sp* and *Sf* genomic DNA. Twenty-five micrograms of genomic DNA was digested with EcoRI and placed in each lane of gels. The DNA was transferred to a nitrocellulose filter, hybridized with the ^{32}P -labeled *Sp* cloned repeats and autoradiographed

the filter with a highly labeled probe representing the family (Southern, 1975). Figure 2 shows the results for EcoRI digests with three probes. Two of these had been analyzed kinetically (Table 1), and one was shown to be long class (CSp2034) and the other short (CSp2101). The left-hand lanes in each pair contain *S. purpuratus* DNA digests. The probe CSp2034 shows a strong band at 1,800 nt and a set of weak bands. Clearly the CSp2034 family occurs in a variety of sequence environments and the strong band suggests a tandem array. Anderson et al. (1981), using selected inserts in a λ phage library of *Sp* DNA have shown that members of this family are clustered in some regions of the *Sp* genome, and have observed parts of tandem arrays. The broad distribution of radioactivity suggests that members of the CSp2034 family occur in a variety of sequence environments (Anderson et al., 1981). As a control we

note that such a distribution is absent for the CSp2133A probe. The majority of the CSp2034 family appears to be in the long class, in agreement with the kinetic measurements shown in Table 1.

The family represented by CSp2101 shows primarily a broad distribution of lengths in the *Sp* *Eco*RI digests, as expected for a principally short interspersed family of repeats. A few bands are present which suggest that a minority of the family is in the long class as suggested by the kinetic measurements (Table 1). The family in *Sp* DNA represented by CSp2133A is clearly in the long class because of the strong bands. Other evidence has not yet been obtained to confirm this observation.

Hybridization of these probes with digests of *S. franciscanus* DNA (right-hand lanes) gives an indication of the evolutionary changes that have occurred in these repetitive families since the two species diverged 10–20 million years ago. Previous observations (Moore et al., 1978) have shown that the families represented by these probes have changed greatly in frequency. They each have a much higher frequency in *Sp* DNA compared to *Sf* DNA (Table 2). The amount of DNA was identical in the *Sf* and *Sp* lanes and the *Eco*RI digestion was complete. Therefore the much lower radioactivity in the *Sf* lanes confirms the frequency change for these three families. For example the family represented by CSp2133A has a frequency in *Sp* 21 times as great as in *Sf* (Table 2) and very little if any hybridization is shown to the *Sf* DNA in Figure 2. The presence of several bands suggests a complex history of multiplication in the evolutionary line leading to *Sp*. The short interspersed members of the CSp2101 family seem to have undergone the greatest change. In addition the bands show that several multiplication and/or rearrangement events have occurred in this family.

The repeat families represented by CSp2034 show striking differences between the two species. For example, the band at 1,800 nt which dominates the *Sp* pattern is totally absent for the *Sf* pattern. However, a weaker band at 1,300 nt appears at about the same intensity in the DNA of the two species. Most of the other weaker bands seem to be distinct in the two species patterns. There has been a great deal of evolutionary change, and many events of multiplication and loss involving this family have occurred in both species.

3. Interspecies Relationships of Low Divergence Families of Repeats

Taken as a whole long class repeats form high thermal stability duplexes while short repeats yield low thermal stability duplexes (Britten et al., 1976). This correlation also holds for our set of individual families of repeats. Table 2 lists the length class and other data for these repeat families in order of the intrafamilial sequence divergence in the *Sp* genome. It is clear that the long class families all have less divergence among their members than do short class families. No examples have been observed of precisely repeated short class families or very divergent long class families.

The interspecies divergence between *Sp* and *Sf* of families which are precisely repeated within *Sp* was examined by means of a low divergence repetitive DNA tracer. This tracer was prepared by reassociating labeled repetitive DNA, heating to near the T_m of precise duplexes and collecting the fraction that did not

Table 2. Summary of measurements made with 13 cloned *Sp* repeat sequences^a

Clone	Divergence <i>Sp</i> (°C)	length class ^b	Method	Probe Length ^c (nt)	Frequency			T_m <i>Sp</i> (°C)	T_m <i>Sf</i> (°C)	T_m <i>Sp</i> - T_m <i>Sf</i> (°C)
					<i>Sp</i>	<i>Sf</i>	ratio			
CSp2007	3.6	both	K	1,100	400	45	9	69.5	64.0	5.5
CSp2034	3.7	long	K, B	510	1,000 (2,500) ^d	160	6-15	76.0	70.5	5.5
CSp2133 A	3.8	long	B	310	2,100	100	21	83.5		
CSp2108	5.7	long	K	204	20 ^e	30	0.8	83.5 ^f	77.50 ^f	6.0 ^f
CSp2101	7.5	short	K, B	320	700	55	13	70.8	70.0	0.8
CSp2109 A	20	short	B	180	1,000	50	20	65.3	59.0	6.3
CSp2096	large ^g			203	80	35	2.3	64.5	60.0	4.5
CSp2090	21			167	140	95	1.5	64.5	65.5	1.0
CSp2085	20			165	170	60	2.8	64.0	59.8	4.2
CSp2137	large ^g	short	K	226	530	185	3	64.0	58.0	6.0
CSp2109 B	25	short	B	110	200	50	4	56.8	58.3	-1.5
CSp2099				235	80	60	1.3			
CSp2111				155	3,000	1,100	2.7			

^a Data is drawn from refs. 19 and 21

^b Length class assignment was made on the basis on the kinetic data of Table 1 for CSp2137, CSp2108, CSp2007, CSp2101, and CSp2034. CSp2133 A was placed in the long category on the basis of the genome blot of Figure 2 which also confirmed that CSp2101 and CSp2034 are in the short and long classes respectively. Restriction and heteroduplex analysis (Anderson et al., 1981; Scheller et al., 1981) placed CSp2109 A and CSp2109 B in the short class and provide examples of a number of specific genomic repeats which include sequences of the CSp2034 and CSp2108 families

^c The length of CSp2108, CSp2096, CSp2090, CSp2109 B, CSp2085, CSp2137, CSp2034 and CSp2109 A are known from DNA sequence analysis (Posakony et al., submitted). The length of the other cloned repeats was estimated from electrophoretic mobility of the duplex fragment (Klein et al., 1978)

^d An estimate of 2,500 members in this family was obtained by Anderson et al. (1) from the kinetics of reassociation with *Sp* DNA of several members of the family which were isolated from λ library inserts. This is the best estimate of the actual size of the CSp2034 family but the ratio of the sizes between *Sp* and *Sf* is probably better from simultaneous measurements using the same probe (Moore et al., 1978)

^e See footnote *d* from Table 1

^f The measurements of thermal stability quoted are from Scheller et al. (1981) for a closely related member of the CSp2108 family isolated from a λ library insert

^g The T_m of the native cloned repeat was not determined. Thus the low thermal stability when reassociated with *Sp* DNA only shows that a large sequence divergence is present in these families

denature. The solvent 2.4 M TEACL suppressed the effect of base composition on melting temperature. The right hand curve in Figure 3 shows that the preparative procedure was effective since the resulting duplexes melt (by HAP assay) with a T_m of 82.5° C. When these duplexes are denatured and reassociated with total *Sp* DNA the T_m of the product drops to 78° C (center curve of Fig. 3). Presumably in making the tracer, duplexes were selected between the more closely related members even though some members of these families are more divergent in sequence. The T_m for the low divergence tracer reassociated with *Sf* DNA is 71° C. Thus there is a T_m reduction of 7° C due to the divergence between the families in *Sp* and the homologous families in *Sf*.

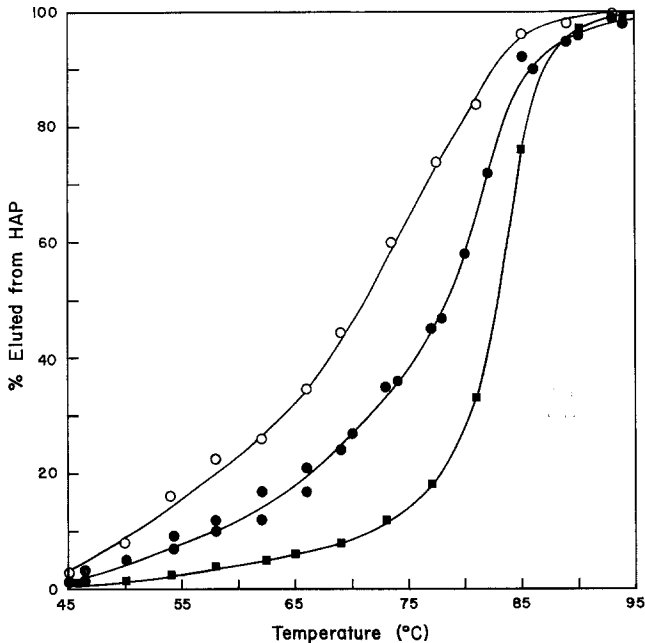


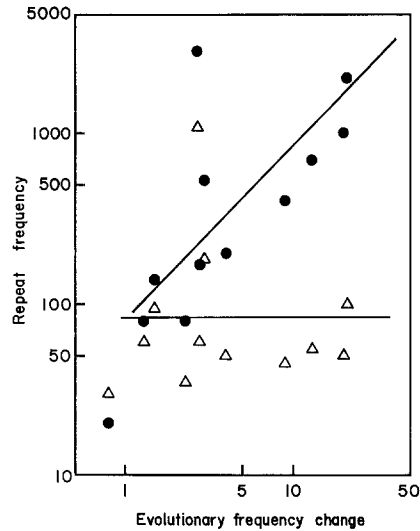
Fig. 3. Thermal stability of the low divergence *Sp* repeat tracer and duplexes formed with total *Sp* or *Sf* genomic DNA. Low divergence *Sp* repeats were prepared and labeled as described in Materials and Methods and hybridized to C_{ot} 100 with excess *Sp* (○) and *Sf* (●) 800 nt genomic DNA at 50° C in 0.12 M PB. The samples were passed over HAP at 45° C in 0.12 M PB and more than 90% was bound. The temperature was raised in 4° C increments and eluted with 0.12 M PB. Since the low divergence tracer was not denatured in the final step of preparation, its precision of base pairing was tested by binding to HAP and thermal elution as above (■)

The intermediate frequency repeat class shows T_m s of 70° C and 66.5° C when reassociated with *Sp* and *Sf* DNA (Moore et al., 1978). Thus the average T_m reduction is 4° C between repeat families in *Sp* and the homologous families *Sf*. Consequently the precisely repeated families change in sequence as much or more than the average repeated sequences. Why then are they more precisely repeated in *Sp*? The most likely explanation for families showing frequency changes is that events of multiplication of particular sequences have occurred in the *Sp* genome since the last common ancestor of the two species. Multiplication and deletion events and base substitutions together can maintain the precise sequence relationships among the members of a repeat family in the DNA of one species while the sequences of the families in the two species drift apart. This is an example of what has been termed "concerted evolution" (Hood et al., 1975; Zimmer et al., 1980). It is also possible that processes akin to gene conversion could maintain precise sequence homology.

Discussion

Measurements of the kinetics of reassociation of the cloned sequences with preparations of long and short class repeats establish three conclusions. First

Fig. 4. Relationship of frequency (number of members) of individual families of repeats to the net frequency change that has occurred since the last common ancestor of *Sf* and *Sp*. Data from Table 2. Solid circles represent the frequency in *Sp* and triangles the frequency in *Sf*. The abscissa is the ratio of these two frequencies. The upper line is the linear regression of the logarithms of the *Sp* frequencies and the frequency ratios. It has a slope of 1.015. The lower line is the same for the *Sf* frequencies and frequency ratios with a slope of 0.015



the long and short class repeat preparations are each nearly free of contamination by the other class. Second, some families of repeats have predominantly long class members and others predominantly short class. Finally some families contain large numbers of members in both classes. Identification of the length class of eight families (shown in Table 2) permits a test of the correlation between length and repeat family sequence divergence for individual families. The four examples of families which have many long class members have intrafamilial divergences ranging from a T_m reduction of 3°C to 6°C . The four examples of short class families have T_m reductions ranging from 7.5°C to 25°C .

A set of measurements of evolutionary frequency change of individual families of repeats is listed in Table 2 including some new measurements as well as previously published data (Moore et al., 1978). Figure 4 displays the correlation between repeat frequency and evolutionary frequency change. The number of members of each family is graphed as a function of the ratio of the number of members in *Sp* to those in *Sf*. Figure 4 shows clear trends in spite of considerable scatter. Families with many members in *Sp* show the largest frequency ratio between *Sp* and *Sf*. Statistical analysis suggests that the frequency in *Sp* is proportional to the frequency change that has occurred since evolutionary lines leading to these two species diverged. The regression line has a slope of almost exactly one on the logarithmic plot, implying proportionality between frequency in *Sp* and frequency change. When the frequencies of these families in *Sf* are analyzed no trend is observed as a function of interspecies frequency ratio. The regression line for these measurements has a slope very close to zero.

These families were selected from *Sp* DNA to represent a wide range of frequencies. It is very likely that if the families had been selected to cover a wide range in *Sf* a complementary result would have been obtained, with the *Sf* families showing a strong correlation between frequency and frequency change, as indicated by measurements with total repetitive tracers (Moore et al.,

1978). These correlations suggest that high frequency families of repeats result from evolutionary recent events of multiplication which occur for certain families in one species and for other different families of repeats in a closely related species. Such events of multiplication are probably balanced during long periods of evolution by deletion or loss of members of these or other repeat families.

An additional insight is supplied by the distribution of frequencies of the families in *Sf* shown on Figure 4. The number of members range from 30 to 200 with an average of about 75. These families have been identified in *Sf* DNA by a method which is more or less independent of the family size or the amount of DNA in each family. Thus such families are common in sea urchin genomes. The higher frequency families tend to make up a majority of the mass of repeats because the quantity of DNA in each family is large and are more likely to be cloned. However the typical family is probably small (20 to 200 members) and the total number of families is very large.

The patterns observed in Figure 2 for the size distribution of restriction fragments show some features of the evolution of CSp2101 and CSp2034 families. First, the patterns corroborate the interspecies frequency differences. Second, they show clearly that in the case of CSp2101 the frequency differences involve many short interspersed repeats. Finally patterns often contain bands in one species at positions that are absent from the other species. Since many very similar members of a family are required to form a band this shows that events of multiplication have occurred in both species. The presence of minor bands of about the same length in the two species may be due to conservation of precise subsets of the family in both species or to a coincidence.

The family represented by clone CSp2101 has undergone a large evolutionary frequency change (Table 2) (Moore et al., 1978). It is shown to be in the short class by both its reassociation with long and short repeats and by its restriction digest pattern. The intrafamilial degree of divergence in *Sp* is not large (7.5° C). The T_m of duplexes formed with *Sf* is very little lower than with *Sp* DNA. Thus the family has suffered little evolutionary change in sequence in the last 10–20 million years. The 13-fold change in frequency shows that many short interspersed copies have either been lost in *Sf* or gained in *Sp*. The CSp2101 family does not appear to have much, if any tandem sequence organization. Thus unequal crossover was probably not the primary mechanism of its frequency change. The mechanism of frequency change is likely to have been insertion or deletion, analogous to that observed in *Drosophila* (Potter et al., 1979; Strobel et al., 1979; Gehring and Paro, 1980). In contrast the family represented by CSp2034 which has also been subject to rapid frequency change may have a significant degree of tandem arrangement in the *Sp* genome (Anderson et al., 1981). Table 2 indicates no correlation between the frequency change and length class. Examples of both large and small frequency changes occur in both classes.

Taken together these observations indicate that more than one mechanism is involved in the repetitive frequency changes and thereby in the production and maintenance of repetitive sequences in the genome. We are at the initial stages of the study of complex and active processes of genomic change by multiplication, insertion and deletion of repeated sequences. These changes are

likely to be an important source of evolutionary variation not only in the DNA itself but also in the function of the genome.

Acknowledgements. This research was supported by NIH Grant GM-20927. G.P.M. was supported by an NIH postdoctoral training grant GM-07401 and by an NIH postdoctoral fellowship GM-07290. W.R.P. was supported by a Carnegie Institution postdoctoral fellowship.

References

- Anderson, D.M., Scheller, R.H., Posakony, J.W., McAllister, L.B., Trabert, S.W., Beall, C., Britten, R.J., Davidson, E.H.: Repetitive sequences in the sea urchin genome. I. Distribution of members of specific repetitive families. *J. molec. Biol.* (in press, 1981)
- Ando, T.: A nuclease specific for heat-denatured DNA isolated from a product of *Aspergillus oryzae*. *Biochim. biophys. Acta (Amst.)* **114**, 158–168 (1966)
- Blin, N., Stafford, D.W.: A general method for isolation of high molecular weight DNA from eukaryotes. *Nucleic Acids Res.* **3**, 2303–2308 (1976)
- Britten, R.J., Graham, D.E., Neufeld, B.R.: Analysis of repeating DNA sequences by reassociation. In: *Methods in enzymology* (L. Grossman and K. Moldave, eds.), vol. 29E, pp. 363–406. New York: Academic Press 1974
- Britten, R.J., Graham, D.E., Eden, F.C., Painchaud, D.M., Davidson, E.H.: Evolutionary divergence length of repetitive sequences in sea urchin DNA. *J. molec. Evol.* **9**, 1–23 (1976)
- Cochet, M., Gannon, F., Heu, R., Maroteaux, L., Perrin, F., Chambon, P.: Organization and sequence studies of the 17-piece chicken conalbumin gene. *Nature (Lond.)* **282**, 567–574 (1979)
- Costantini, F.D., Scheller, R.H., Britten, R.J., Davidson, E.H.: Repetitive sequence transcripts in the mature sea urchin oocyte. *Cell* **15**, 173–187 (1978)
- Costantini, F.D., Britten, R.J., Davidson, E.H.: Message sequences and short repetitive sequences are interspersed in sea urchin poly(A)⁺ RNAs. *Nature (Lond.)* **287**, 111–117 (1980)
- Crain, W.R., Eden, F.C., Pearson, W.R., Davidson, E.H., Britten, R.J.: Absence of short period interspersed repetitive and non-repetitive sequences in the DNA of *Drosophila melanogaster*. *Chromosoma (Berl.)* **56**, 309–326 (1976)
- Davidson, E.H., Galau, G.A., Angerer, R.C., Britten, R.J.: Comparative aspects of DNA organization in metazoa. *Chromosoma (Berl.)* **51**, 253–259 (1975)
- Eden, F.C., Hendrick, J.P.: Unusual organization of DNA sequences in the chicken. *Biochemistry* **17**, 5838–5844 (1978)
- Eden, F.C., Graham, D.E., Davidson, E.H., Britten, R.J.: Exploration of long and short repetitive sequence relationships in the sea urchin genome. *Nucleic Acids Res.* **4**, 1553–1567 (1977)
- Eden, F.C., Hendrick, J.P., Gottlieb, S.S.: Homology of single copy and repeated sequences in chicken, duck, Japanese quail, and ostrich DNA. *Biochemistry* **17**, 5113–5121 (1978)
- Gehring, W.J., Paro, R.: Isolation of a hybrid plasmid with homologous sequences to a transposing element of *Drosophila melanogaster*. *Cell* **19**, 897–904 (1980)
- 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., Davidson, E.H.: DNA sequence organization in the genomes of five marine invertebrates. *Chromosoma (Berl.)* **51**, 225–251 (1975)
- Graham, D.E., Neufeld, B.R., Davidson, E.H., Britten, R.J.: Interspersed repetitive and non-repetitive DNA sequences in the sea urchin genome. *Cell* **1**, 127–137 (1974)
- Hinegardner, R.: Echinoderms: In: *Methods in developmental biology* (F.H. Witt and N.K. Wessels, eds.), pp. 139–155. New York: Thomas Y. Crowell 1967
- Hood, L., Campbell, J.H., Elgin, S.C.R.: The organization, expression and evolution of antibody genes and other multigene families. *Ann. Rev. Genet.* **9**, 305–353 (1975)
- Klein, W.H., Thomas, T.L., Lai, C., Scheller, R.H., Britten, R.J., Davidson, E.H.: Characteristics of individual repetitive sequence families in the sea urchin genome studied with cloned repeats. *Cell* **14**, 889–900 (1978)
- Manning, J.E., Schmid, C.W., Davidson, N.: Interspersed repetitive and non-repetitive DNA sequences in the *Drosophila melanogaster* genome. *Cell* **4**, 141–155 (1975)
- Moore, G.P., Scheller, R.H., Davidson, E.H., Britten, R.J.: Evolutionary change in the repetition frequency of sea urchin DNA sequences. *Cell* **15**, 649–660 (1978)

- Potter, S.S., Brorein, W.J., Dunsmuir, P., Rubin, G.M.: Transposition of elements of the 412, copia and 297 dispersed repeated gene families in *Drosophila*. *Cell* **17**, 415-427 (1979)
- Scheller, R.H., Thomas, T.L., Lee, A.S., Klein, W.H., Niles, W.D., Britten, R.J., Davidson, E.H.: Clones of individual repetitive sequences from sea urchin DNA constructed with synthetic EcoRI sites. *Science* **196**, 197-200 (1977)
- Scheller, R.H., Costantini, F.D., Britten, R.J., Davidson, E.H.: Specific representation of cloned repetitive DNA sequences in sea urchin RNAs. *Cell* **15**, 189-203 (1978)
- Scheller, R.H., Anderson, D.M., Posakony, J.W., McAllister, L.B., Britten, R.J., Davidson, E.H.: Repetitive sequences of the sea urchin genome. II. Subfamily structure and evolutionary conservation. *J. molec. Biol.* (in press, 1981)
- Southern, E.M.: Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. molec. Biol.* **98**, 503-517 (1975)
- Strobel, E., Dunsmuir, P., Rubin, G.: Polymorphism in the chromosomal locations of elements of the 412, copia and 297 dispersed repeated gene families in *Drosophila*. *Cell* **17**, 429-439 (1979)
- Wensink, P.C., Shiro, T., Pachl, C.: The clustered and scrambled arrangement of moderately repetitive elements in *Drosophila* DNA. *Cell* **18**, 1231-1246 (1979)
- Zimmer, E.A., Martin, S.L., Berverley, S.M., Kan, Y.W., Wilson, A.C.: Rapid duplication and loss of genes coding for the α chain of hemoglobin. *Proc. nat. Acad. (Wash.)* **77**, 2158-2162 (1980)

Received February 26, 1981 / Accepted by J.H. Taylor
Ready for press April 17, 1981