Deoxyribonucleic Acid Sequence Organization in the Mung Bean Genome†

Michael G. Murray,* Jeffrey D. Palmer,‡ Richard E. Cuellar,‡ and William F. Thompson‡

ABSTRACT: Deoxyribonucleic acid (DNA) reassociation kinetics were used to characterize repeated and single-copy sequences in the mung bean (*Vigna radiata*) genome and to analyze the interspersion of single-copy and repeated sequences. Experiments with 300 nucleotide long fragments of total DNA as well as with a purified single-copy fraction show that the most slowly reassociating sequences do so with a rate of 1.6×10^{-3} when assayed with hydroxylapatite. This rate is ~105 times slower than that observed for *Escherichia coli* DNA included as an internal standard, and thus the kinetic complexity of the mung bean genome is $\sim 4.7 \times 10^8$ nucleotide pairs or 0.48 pg/haploid nucleus. This value is in good agreement with cytophotometric measurements and makes the mung bean genome the smallest yet characterized among higher plants. About 35% of the total root DNA fragments contain repetitive sequences. However, total leaf DNA preparations contain an additional 11% rapidly renaturing fraction, which is attributable to chloroplast DNA. The reassociation kinetics of repetitive DNA can be modeled with a single theoretical second-order component, but the data do not specify a unique solution. The reassociation kinetics of isolated high- and low-frequency fractions indicate that repetitive sequence families probably cover a range from about fifty to several thousand copies per haploid genome. Long tracer reassociation kinetics at three different fragment lengths indicate that ~35% of the single-copy sequences are interspersed with repetitive sequences at distances between 300 and 1200 nucleotide pairs, and an additional 18% of the single-copy sequences are interspersed between 1200 and 6700 nucleotide pairs. About 46% of the single-copy sequences are not interspersed with repeats at a distance of 6700 nucleotide pairs. While the distances at which the latter sequences might be interspersed are unknown, it appears that sequence organization in the mung bean covers the entire spectrum from short- to long-period interspersion. The pattern of sequence repetition and arrangement in mung bean DNA differs dramatically from that previously observed in the garden pea, which is a member of the same subfamily.

Studies on the linear arrangement of single-copy and repetitive sequences have contributed to our current concepts on the evolution and function of the eucaryotic genome. DNA from a majority of the organisms examined to date exhibits the same general pattern of short-period interspersion, wherein short repetitive elements (modal length ~300 NTP[1]) are interspersed among single-copy sequences at distances of less than ~2500 NTP (Davidson et al., 1975; Walbot & Goldberg, 1978; Thompson & Murray, 1979). The widespread occurrence of this short-period pattern has led to the hypothesis that it may have some functional significance in the coordinate regulation of transcription (Britten & Davidson, 1969; Davidson & Britten, 1979). However, since only a small fraction of the single-copy sequences in most eucaryotes may actually function as structural genes [e.g., Galau et al. (1976)], it is reasonable to suppose that at most only a small fraction of the short-period interspersion is actually important for gene regulation (Davidson et al., 1977). Moreover, the short-period interspersion pattern is not universal. An increasing number of organisms are now being found whose genomes are organized in long-period interspersion patterns. In the classic case of long-period interspersion, single-copy sequences in *Drosophila* DNA may extend for lengths exceeding 13 000 NTP without interruptions by repeated sequences (Manning et al., 1975). Most other early examples of long-period interspersion were also found among insect species [e.g., Crain et al. (1976)], but similar patterns have recently been reported for fungal (Hudspeth et al., 1977), nematode (Schachat et al., 1978), and avian (Arthur & Strauss, 1978; Epplen et al., 1978) genomes.

Plant genomes differ from those of most animals in that they are generally larger (even on a haploid basis) and contain a larger fraction of repetitive sequences. Studies of sequence interspersion have shown that the tobacco (Zimmerman & Goldberg, 1977), soybean (Gurley et al., 1978; Goldberg, 1978), parsley (Kiper & Herzfeld, 1978), wheat and rye (Flavell & Smith, 1976; Smith & Flavell, 1977), and pea (Murray et al., 1978) genomes generally conform to the short-period interspersion pattern. In fact, we have shown (Murray et al., 1978) that the pea genome is more extensively interspersed at very short intervals than has yet been documented for any other genome.

However, we cannot generalize that all higher plants show primarily short-period interspersion. Experiments reported here show that the DNA of the mung bean (*Vigna radiata*) differs dramatically in the pattern of sequence organization from that of the pea, even though both are in the same subfamily (Faboideae) of the legumes. The mung bean genome has a haploid (1C) DNA content of ~0.48 pg or 4.7×10^8 NTP, making it the smallest plant genome that has yet been extensively characterized.[2] About 35% of the mung bean

† From the Department of Plant Biology, Carnegie Institution of Washington, Stanford, California 94305. *Received May 30, 1979.* This work was supported in part by grants from the National Science Foundation (PCM 7705656) and the Competitive Grants Office of the Science and Education Administration, U.S. Department of Agriculture (5901-0410-8-0009-0). R.E.C. received support from a Ford Foundation predoctoral fellowship. J.D.P. received support from a National Science Foundation predoctoral fellowship.

‡ Also Department of Biological Sciences, Stanford University, Stanford, CA 94305.

[1] Abbreviations used: NT, nucleotides; NTP, nucleotide pairs; C_0t, the product of molar concentration of DNA nucleotides and time of incubation (values throughout paper are given in units of mol s L^{-1}); K, the reassociation rate constant (values throughout paper are given in units of L mol^{-1} s^{-1}); T_m, the temperature at which half of the nucleotides are unpaired; Pipes, 1,4-piperazinediethanesulfonic acid; PB, an equimolar mixture of NaH_2PO_4 and Na_2HPO_4 (pH 6.8).

[2] Wimpee & Rawson (1979) have recently estimated the haploid genome size of pearl millet [*Pennisetum americanum* (L.) K. Schum. (= *P. glacum* auct.)] to be 0.22 pg, but his value is in serious disagreement with cytophotometric measurements of 2.5 pg (Bennett & Smith, 1976).

0006-2960/79/0418-5259$01.00/0

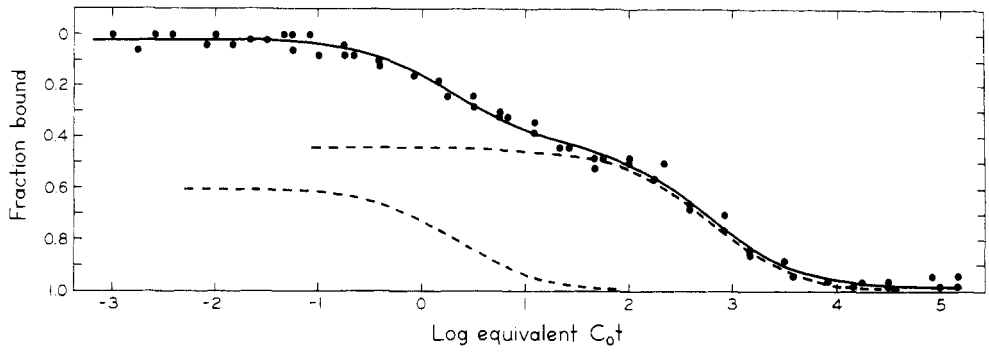

FIGURE 1: Reassociation kinetics of 300-NT mung bean leaf DNA fragments. Samples of unlabeled DNA were reassociated to various equivalent C_0t values and fractionated on hydroxylapatite as described in the text. *E. coli* tracers were included in all reassociation mixtures to provide an internal kinetic standard (see text). The solid line is the simplest least-squares fit to the data using two theoretical second-order components (dashed lines), the elements of which are summarized in Table IA.

genome is composed of repetitive sequences, which is considerably less than in most other higher plants [reviewed by Thompson & Murray (1979)]. Although a portion of the mung bean genome does show short-period interspersion, almost half of the single-copy sequences are at least 6700 NTP in length.

Experimental Procedures

Details of DNA purification, preparation of radioactive labeled tracers, shearing, and fragment length measurement have been described previously (Murray et al., 1978). "Leaf" DNA was isolated from light-grown seedling tissue consisting of about 80% primary leaves and 20% epicotyls (on a fresh weight basis). In vivo ^{3}H-labeled root DNA was isolated from seedlings germinated aseptically in the presence of [^{3}H]thymidine. Dry seeds with unbroken seed coats were surface sterilized by sequential treatment with 10% chlorox–1% sodium dodecyl sulfate for 3 min, followed by 95% ethanol for 1 min. Tissue was washed extensively with 1% sodium dodecyl sulfate prior to DNA extraction. Upon isopycnic centrifugation the radioactivity profile coincided precisely with the absorbance profile, indicating that the tracer was free of significant bacterial contamination (Bendich, 1972).

Chloroplast and Nuclear DNA Isolation. Chloroplasts were prepared from 20-day-old seedling leaves by using sucrose flotation as described by Manning et al. (1971). DNA was purified from chloroplast lysates by two rounds of CsCl–ethidium bromide isopycnic centrifugation, followed by RNase, Pronase, and hydroxylapatite chromatography as described previously (Murray et al., 1978). Optical measurements of reassociation kinetics in 0.12 M PB at 60 °C were used to assess the purity of the chloroplast DNA. Ninety-three percent of the 400 NT long chloroplast DNA reassociated as a single theoretical component with a rate of 5.91 (root mean square error = 0.019). This rate is 30.5 times faster than that of 400 NT long *Escherichia coli* DNA under our conditions. Taking the size of the *E. coli* genome to be 4.5×10^6 NTP (Cairns, 1963; Klotz & Zimm, 1972), we calculated the kinetic complexity of the chloroplast DNA to be 1.48×10^5 NTP. This value is in good agreement with previous estimates for the kinetic complexity of chloroplast DNA in other leguminous species of $(1.45–1.47) \times 10^5$ NTP (Kolodner & Tewari, 1975).

During the preparation of chloroplasts, the material sedimenting through 1.7 M sucrose (80000*g*, 2 h) was retained for the preparation of a nuclear fraction. This pellet was washed 3 times with 3% Triton X-100 (1000*g*, 10 min) before lysis and DNA purification as described for chloroplasts.

Reassociation and Fractionation. Reassociation was carried out in 0.12 or 0.4 M PB or in 1.0 M NaCl–0.01 M PB at 60, 66, or 70 °C, respectively (T_m –25 °C). Single-stranded DNA was separated from duplex-containing fragments by chromatography on hydroxylapatite (Bio-Rad HTP, lot no. 16020) as described previously (Britten et al., 1974; Murray et al., 1978). Reassociation of unlabeled DNA was assayed by measuring the A_{260nm} after correction for light scattering at 340 nm. Radioactivity in each fraction was measured by scintillation counting in a Triton–toluene cocktail [13.3 g of Ominfluor (Packard), 1 L of Triton X-100, and 2 L of toluene].

The reassociation kinetics of in vivo labeled *E. coli* DNA fragments, included as an internal standard in all reassociation mixtures, have been used to correct for the effects of viscosity on the reassociation rate (Wetmur & Davidson, 1968) by normalizing so that the observed internal standard rate is equal to that for *E. coli* DNA reassociated alone (K = 0.17 for 300-NT fragments). All C_0t values for buffers other than 0.12 M PB have been corrected to the equivalent C_0t for 0.12 PB at 60 °C (Britten et al., 1974; Angerer et al., 1976). The results were analyzed in terms of theoretical second-order kinetics by using the computer program described by Pearson et al. (1977), and data are presented in the form of C_0t curves (Britten & Kohne, 1968).

Thermal Stability of Reassociated Repeats. 300 NT long fragments were reassociated in 2.4 M tetraethylammonium chloride (Melchior & Von Hippel, 1973; Chang et al., 1974). Optical thermal denaturation profiles were recorded in the same solvent with a Gilford 2527 thermoprogrammer. First-derivative profiles were obtained as described previously (Cuellar et al., 1978).

Repeat Length Measurements. Samples of 8100 NT long in vivo ^{3}H-labeled mung bean root DNA fragments in 1.0 M NaCl and 20 mM Pipes (pH 6.7) were reassociated at 70 °C to an equivalent C_0t value of 10, samples were adjusted to 25 mM sodium acetate (pH 4.4), 0.1 mM $ZnCl_2$, 150 mM NaCl, and 5 mM 2-mercaptoethanol (Britten et al., 1976) and digested for 45 min at 37 °C with 5 units of S-1 nuclease (Miles Laboratoreis) per μg of DNA. A unit is defined as the amount of enzyme which renders 1 μg of single-stranded DNA acid-soluble in 30 min at 37 °C. S-1 resistant duplexes were isolated by hydroxylapatite chromatography and sized on neutral agarose gels as described previously (Murray et al., 1978) by using the *Hae*III restriction nuclease fragments of ϕX-174 DNA as markers.

Results

Mung Bean DNA Reassociation Kinetics. The reassociation kinetics of 300 NT long total leaf DNA fragments are shown in Figure 1. The data were analyzed with a least-squares

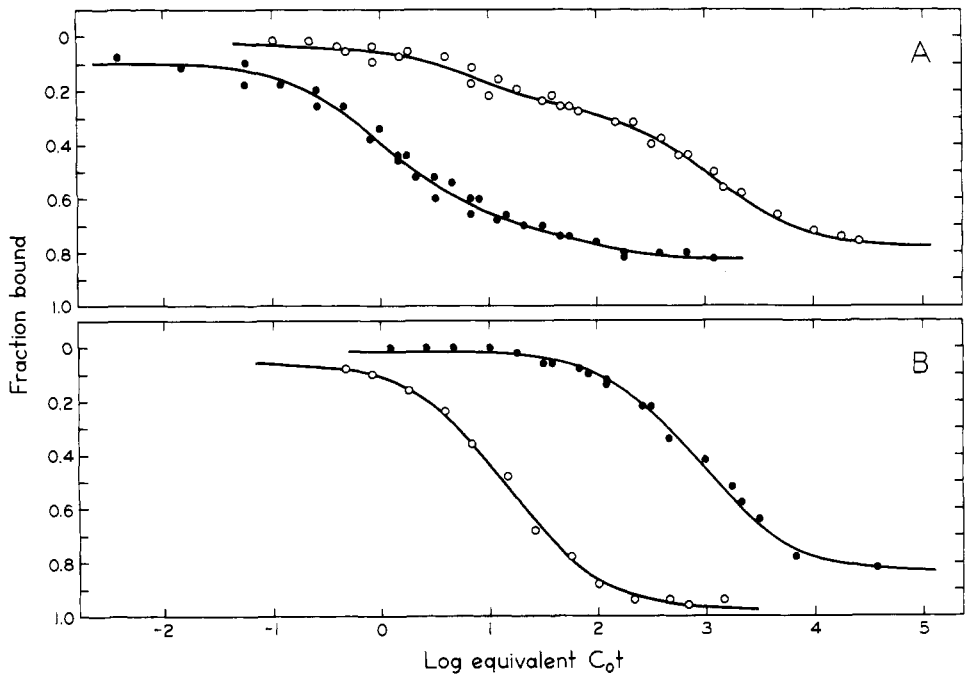

FIGURE 2: Reassociation kinetics of isolated kinetic fragments of the mung bean leaf DNA. (A) Reassociation kinetics of sequences renaturing before and after C_0t = 5. 300-NT DNA fragments were separated into two fractions by reassociation and fractionation on hydroxylapatite. ^{14}C-Labeled tracers with a modal length of 200 NT were prepared by using *E. coli* DNA polymerase I as described previously (Murray et al., 1977a). The reassociation kinetics of each tracer were followed in the presence of excess unlabeled 300-NT total leaf DNA (as described in Figure 1). Lines represent the best least-squares fit to the data, the elements of which are summarized in Table IB. (●) C_0t = 5 bound, 40%; (○) C_0t = 5 unbound, 60%. (B) Reassociation of a single-copy tracer. A single-copy-enriched fraction with the following reassociation history was prepared: C_0t = 5 unbound (60%), 100 unbound (73%), and 300 unbound (57%). ^{14}C-Labeled tracer (modal length 200 NT) prepared from this fraction was reassociated again to C_0t = 300 in the presence of a 33-fold sequence excess of total leaf DNA and the unbound fraction again isolated. The final tracer represented 12% of the starting DNA. Reassociation of this tracer (●) was followed in the presence of a 300-fold sequence excess of 300-NT total leaf DNA as described in Figure 1. The elements of the best least-squares solution to the data are shown in Table IB. Open symbols (○) show the reassociation kinetics of the internal *E. coli* kinetic standard.

computer program (Pearson et al., 1977) designed to fit theoretical second-order components to the observed reassociation kinetics.

About 2% of the reactable fragments bind to hydroxylapatite before $C_0t = 10^{-3}$ and no further binding is observed until C_0t values approach 10^{-1}. While the initial binding will include any nonspecific binding to hydroxylapatite, it probably results primarily from the rapid intramolecular reassociation of inverted repeat or palidromic sequences (Huguet et al., 1975; Bazetoux et al., 1978). Forty-two percent of the reactable fragments reassociate more rapidly (K = 0.50) than would be expected for sequences present only once per haploid genome and thus must contain repetitive elements. The remaining 56% of the reactable fragments reassociate with a rate of 0.0018.

While the data in Figure 1 can be fit by using only two theoretical second-order components, a somewhat better solution (by the criterion of root mean square error) can be obtained by using three theoretical second-order components (Table IA). In this solution there is little change in the most slowly reassociating component, which now includes 54% of the reactable fragments with a rate constant of 0.0016. However, the repetitive fraction is now described by two theoretical second-order components: a 16% fast repeat fraction reassociating with a rate of 2.4 and a 29% slow repeat component reassociating with a rate of 0.20. In addition to the two- and three-component solutions to total DNA reassociation kinetics, additional solutions could be obtained which were nearly as good; thus, the data do not specify a unique solution.

To determine more precisely whether the repeated sequences in the mung bean genome are composed of one, two, or many different repetition frequency classes, we have followed the reassociation of kinetically fractionated tracers. Leaf DNA fragments (300 NT) were successively reassociated and fractionated on hydroxylapatite to yield a repetitive fraction reassociating before C_0t = 5 (hereafter called 5 BD) and a second fraction reassociating after C_0t = 5 (5 UBD). A portion of the 5 UBD fraction was further fractionated (see Figure 2 legend) to yield a fraction (C_0t = 300 UBD) that should be nearly pure single-copy sequences. Radioactively labeled tracers were prepared from each of the above fractions as described previously (Murray et al., 1977a), and each tracer was reassociated in the presence of a 2000-fold excess of unlabeled total leaf DNA.

Figure 2A shows that while the fractionation at C_0t = 5 did achieve a reasonable separation of rapidly and slowly reassociating sequences, both the 5 BD and 5 UBD tracers still show heterogeneous kinetics. Neither fraction can be adequately modeled with a single theoretical component. The reassociation kinetics are within experimental error of the behavior predicted from the three-component solution to the data of Figure 1 (see Table IA). However, even this solution is probably an oversimplification. Whereas the 300 UBD fraction should have been nearly pure single-copy based on the three-component solution for total DNA, ~24% of this tracer reassociates as low-frequency (~50 copies) repetitive sequence (Table IB). Taken with the fact that none of the reassociation kinetics clearly specified unique solutions, this observation makes it likely that mung bean genome actually includes a broad range of different repetition frequencies.

The reassociation kinetics of a pure single-copy probe were used to obtain an accurate estimate of the size of the mung bean genome. This probe was prepared from the 300 UBD tracer by further fractionation after reassociation to C_0t = 300 in the presence of a 33-fold sequence excess of unlabeled total

Table I: Reassociation Kinetics of Total Mung Bean Leaf DNA and Kinetic Fractions

	DNA prepn	component	fraction[a]	rate[b]	rms[c]
A	total leaf (2 component)	early binding	0.02 ± 0.01		
		repeat	0.42 ± 0.01	0.50 ± 0.07	0.0259
		single copy	0.56 ± 0.01	0.0018 ± 0.0002	
	total leaf (3 component)	early binding	0.01 ± 0.01		
		highly repetitive	0.16 ± 0.12	2.4 ± 2.5	
		middle repetitive	0.29 ± 0.11	0.20 ± 0.14	0.0245
		single copy	0.54 ± 0.02	0.0016 ± 0.0002	
B	5 BD	highly repetitive	0.78 ± 0.05	1.60 ± 0.21	0.0279
		middle repetitive	0.22 ± 0.05	0.030 ± 0.024	
	5 UBD	middle repetitive	0.32 ± 0.03	0.20 ± 0.04	
		single copy	0.68 ± 0.03	0.0012 ± 0.002	0.0205
	300 UBD	middle repetitive	0.24 ± 0.03	0.082 ± 0.022	
		single copy	0.76 ± 0.03	0.0013 ± 0.0002	0.0205
	single copy	single copy	1.0	0.0016 ± 0.0001	0.0175

[a] Fractions have been corrected to 100% reactable DNA. Reactivity was 98% for total leaf DNA and 78–84% for the kinetic fraction tracers. [b] All data have been corrected for the effects of viscosity on reassociation rate. Rates for kinetic fractions have been standardized for 300 NT long fragments. [c] Root mean square error of the least-squares solution to the data.

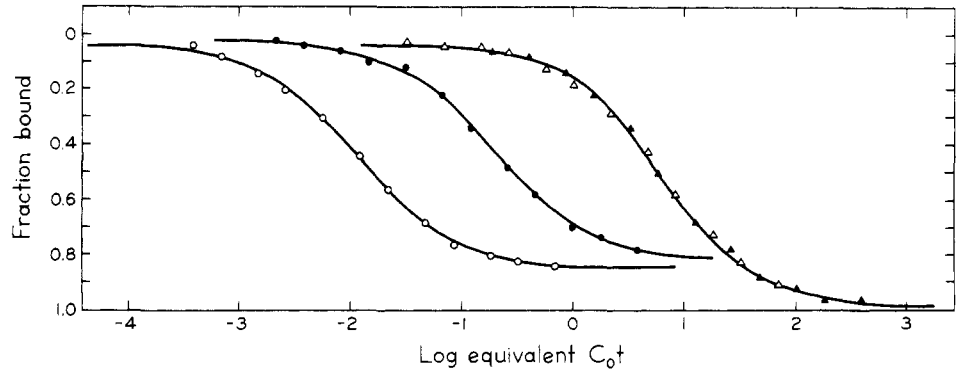

FIGURE 3: Determination of the concentration of chloroplast sequences in leaf DNA. Chloroplast DNA (○) was labeled in vitro with ^3H (see text) and mixed with ^{14}C-labeled *E. coli* DNA (△). Aliquots of the mixture were reassociated in the presence (open symbols) or absence (closed symbols) of excess unlabeled leaf DNA in 0.4 M PB at 66 °C and fractionated on hydroxylapatite. The chloroplast tracer reassociated with a rate of 87.5 and 5.2 in the presence or absence of leaf driver DNA, respectively. The leaf driver–tracer mixture contained 55.4 μg/mL leaf DNA and 0.365 μg/mL ^3H-labeled chloroplast probe. Chloroplast probe reassociations are plotted against chloroplast probe equivalent C_0t, while *E. coli* reassociations are plotted against *E. coli* equivalent C_0t. All DNAs were 300 NT in length.

leaf DNA. The reassociation kinetics of the resulting unbound tracer were followed in the presence of an additional 300-fold sequence excess of total leaf DNA (Figure 2B and Table IB). These data can be fit with a single theoretical second-order component with a reassociation rate of 0.0016. Since this rate is 105 times slower than that of the internal *E. coli* DNA kinetic standard, the kinetic complexity of the mung bean genome is ~105 times greater than that of *E. coli* DNA. Taking the size of the *E. coli* genome to be 4.5×10^6 NTP (Cairns, 1963; Klotz & Zimm, 1972), we calculate that the mung bean contains ~4.7×10^8 NTP or ~0.48 pg of DNA per haploid nucleus.

Contribution of Chloroplast Sequences to the "Repetitive" DNA of Leaves. During the course of this investigation, some tissue-specific differences in the reassociation kinetics of various DNA preparations were observed which suggested that a significant fraction of the repeated sequences in total leaf DNA preparations might be attributable to chloroplast DNA. Preliminary evidence supporting this possibility was obtained from detailed analysis of the reassociation kinetics of total leaf DNA, leaf nuclear DNA, and the 5 BD fraction. Reassociation at various temperatures was monitored continuously by optical hyperchromicity. Even at only 5 °C below the T_m, a point at which only precisely paired duplexes may form, 10–17% of the total leaf DNA could be described as reassociating with the kinetic complexity of chloroplast DNA (Preisler & Thompson, 1978; M. G. Murray unpublished experiments). This component was not observed in nuclear DNA preparations.

Direct measurements of the concentration of chloroplast sequences in leaf DNA preparations were made by following the reassociation kinetics of a chloroplast probe in the presence or absence of a known concentration of leaf DNA (Figure 3). ^3H-Labeled chloroplast tracer was prepared from purified chloroplast DNA by using *E. coli* DNA polymerase I as described previously (Murray et al., 1977a). ^3H-Labeled chloroplast tracer was mixed with ^{14}C-labeled *E. coli* DNA, which served as a nonaccelerated internal kinetic standard. The absolute concentration of chloroplast tracer as well as the ratio chloroplast/*E. coli* DNA was calculated from the reassociation kinetics of the tracer mixture alone. A portion of the same tracer solution was diluted with leaf DNA to a total leaf DNA concentration of 55.4 μg/mL. From the *E. coli* reassociation rate in this mixture, the chloroplast tracer concentration was calculated to be 0.365 μg/mL. Chloroplast probe reassociation was accelerated by a factor of 16.8, and thus the leaf preparation contributed 5.77 μg/mL [= (16.8 × 0.365) − 0.365] chloroplast DNA. The concentration of chloroplast DNA in the leaf DNA preparation is thus 10.4% (= 5.77/55.4). In a second experiment a value of 11.8% was obtained, giving an average of 11% for the contribution of chloroplast DNA to the leaf preparation. By similar analysis, the nuclear preparation shown in Figure 4 was found to contain 1% contamination with chloroplast DNA (Palmer et al., 1979).

Table II: Reassociation Kinetics of Long Mung Bean Tracers

fragment length[a] (NT)	early binding fraction[b]	highly repetitive		middle repetitive		single copy		
		fraction	rate	fraction	rate	fraction	rate[c]	rms[d]
300	0.02	0.05 ± 0.02	16.0 ± 22.0	0.27 ± 0.02	0.28 ± 0.26	0.65 ± 0.01	0.0016	0.0156
1200	0.07	0.13 ± 0.03	7.5 ± 6.0	0.38 ± 0.02	0.11 ± 0.03	0.42 ± 0.03	0.0065	0.0197
6700	0.19	0.18 ± 0.02	27.6 ± 12.5	0.32 ± 0.04	0.30 ± 0.12	0.30 ± 0.06	0.036	0.0197

[a] To compensate for strand scission, we determined tracer lengths on parallel samples incubated to the approximate $C_0t_{1/2}$ of the slowest reacting component in each series. [b] Binding observed prior to $C_0t = 10^{-3}$. Fractions have been normalized to 100% of the DNA reactable. Reactivity was 95-98%. [c] Single-copy rates were fixed at those predicted for 1200- and 6700-NT single-copy sequences, as described in the text. [d] Root mean square error of the least-squares solution to the data.

The chloroplast sequences in total leaf DNA preparations have a significant effect on the thermal stability profiles of reassociated repeats (Figure 4). 300 NT long leaf, root, or nuclear DNA fragments were reassociated and denatured in 2.4 M tetraethylammonium chloride. Since this solvent abolishes the effect of base composition on thermal stability (Melchior & Von Hippel, 1973; Chang et al., 1974), the denaturation profiles reflect only the effects of mismatch and fragment length. All three samples show a broad range in thermal stability, indicating that there is a range in the mismatch in reassociated sequences. However, while all were reassociated to the same equivalent C_0t value of 40, the leaf preparation was 31% base-paired while the root and nuclear preparations were only ~20% base-paired. Most of the additional reassociated sequences in the leaf preparation appear as a prominent high-stability peak. Since this peak occurs at nearly native T_m for fragments of this length (60.7 °C), these sequences are nearly perfectly paired, as would be expected for chloroplast DNA.

Interspersion of Repetitive and Single-Copy DNA. The linear arrangement of single-copy and repetitive sequences was examined by comparing the reassociation kinetics for ^{3}H-labeled root DNA tracers of differing lengths in the presence of excess unlabeled 300-NT leaf DNA. The reassociation kinetics of 300-NT ^{3}H-labeled root DNA fragments in the presence of excess unlabeled 300-NT leaf DNA are shown in Figure 5, and the elements of the best three-component solution are summarized in Table II. As in the case of leaf DNA, ~2% of the tracer binds to hydroxylapatite before $C_0t = 10^{-3}$. About 65% of the root tracer reassociates with the single-copy rate of 0.0016, and the remainder can be described by two repetitive components. In comparing these data to those for total leaf DNA, we note that the total apparent repetitive sequence content in root DNA is ~11% lower than in leaf preparations of the same length, which is consistent with the presence of more chloroplast DNA in leaf preparations.

To determine the fraction of the long tracer fragments containing purely single-copy sequences, long tracer reassociation kinetics were analyzed by using the rate constants expected for single-copy sequences at the indicated lengths. The predicted reassociation rate for 1200 NT long single-copy DNA is 0.0064 [= (1200/300) × 0.0016], and that for 6700 NT long single-copy sequences is 0.036 [= (6700/300) × 0.0016] (Davidson et al., 1973; Hinnebusch et al., 1978; Chamberlin et al., 1978; Murray et al., 1978). As tracer fragment length is increased from 300 to 1200 NT, the fraction of the fragments reassociating with the predicted single-copy kinetics drops from 65 to 42%. Consequently, 23% (= 65% − 42%) of the 1200-NT fragments contain single-copy sequences contiguous with repeated sequences, and we may calculate that 35% (= 23%/65%) of the single-copy DNA is interspersed with repetitive sequences at distances between 300 and 1200 NTP. At a fragment length of 6700 NT, 30% of the fragments reassociate with the predicted single-copy rate and thus 18% [= (42% − 30%)/65%] of the single-copy sequences are interspersed with repeats at distances between 1200 and 6700 NT. The remaining 46% (= 30%/65%) of the single-copy sequences are not interspersed at distances of 6700 NT.

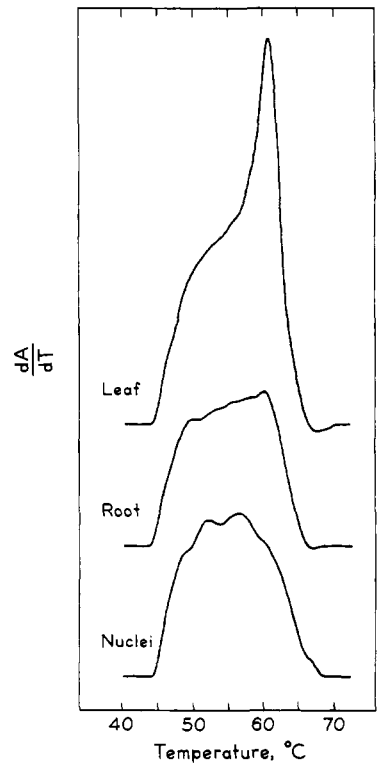

FIGURE 4: Thermal stability of reassociated repeats in leaf, root, and nuclear DNA preparations. 300-NT fragments of total leaf, root, or nuclear DNA were reassociated in 2.4 M tetraethylammonium chloride at 45 °C to an equivalent C_0t value of 40, prior to thermal denaturation. First derivatives of optical thermal denaturation profiles were obtained as described previously (Cueller et al., 1978) and have been scaled such that the area under each curve is proportional to the total hyperchromicity. The hyperchromicity of reassociated DNA (9.8, 6.8, and 7.0% of the denatured A_{260} for leaf, root, and nuclear DNA, respectively) was compared with that of the native DNA (27.5%), allowing for the 1.6% hyperchromic shift observed for single strands under our conditions. The fraction base-paired was 31.7% [= (9.8 − 1.6)/(27.5 − 1.6)], 20.1% [= (6.8 − 1.6)/(27.5 − 1.6)], and 20.8% [= (7 −1.6)/(27.5 − 1.6)] for leaf, root, and nuclear preparations, respectively. Native T_m for 300-NT fragments in this solvent was 60.7 °C.

Length of Repetitive Sequence Elements. Having established that about half of the single-copy sequences in mung beans are greater than 6700 NT in length, we wished to determine whether the repeated sequences might also be quite long. Figure 6 shows the size distribution of the S-1 resistant duplexes obtained after reassociating long (8100 NTP) ^{3}H-labeled root DNA fragments to an equivalent C_0t of 10. The mass average length after S-1 digestion is ~1260 NTP, and the number average is 550 NTP. Five units of S-1 nuclease per μg of DNA was found previously to give complete digestion

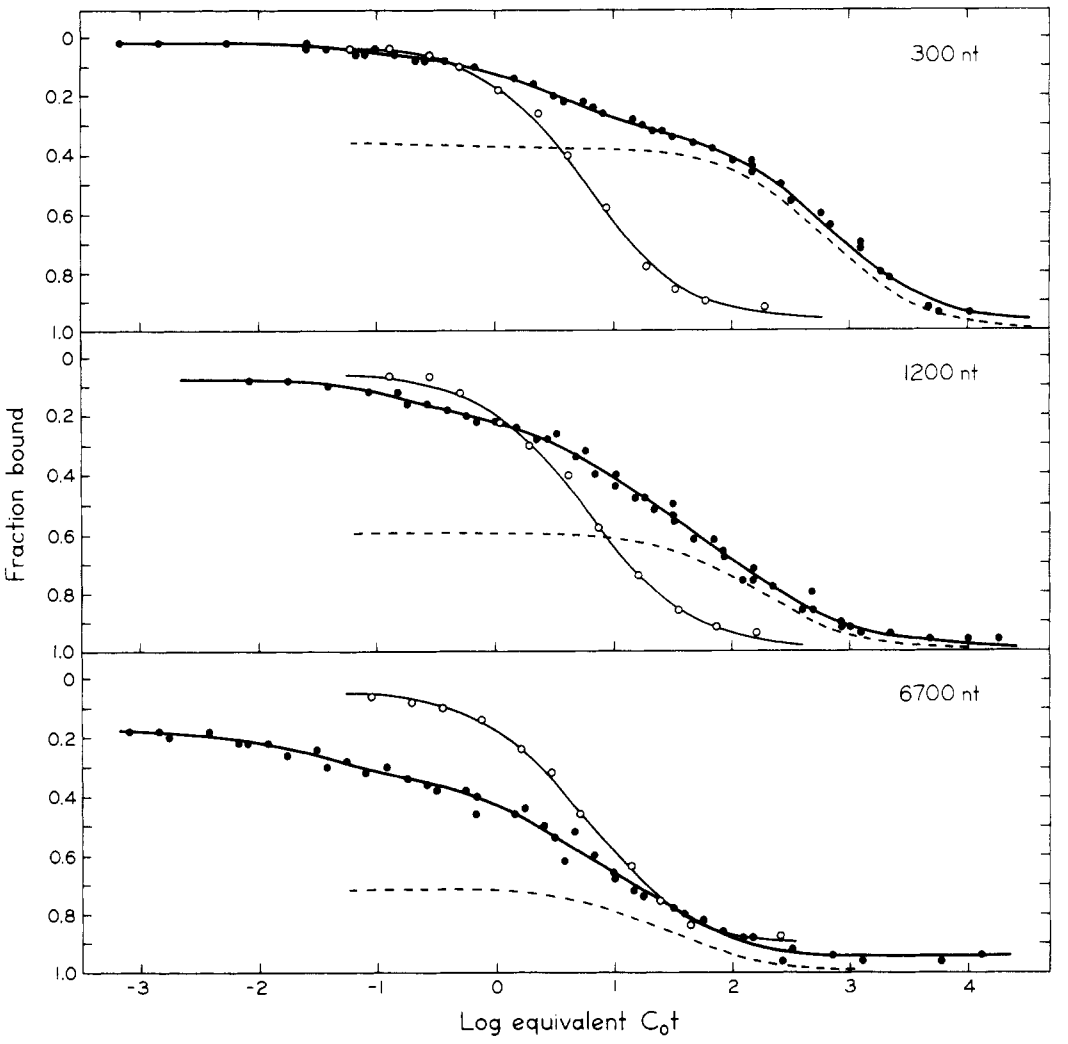

FIGURE 5: Reassociation of mung bean DNA fragments of differing lengths. ^{3}H-Labeled mung bean DNA was isolated from seedling roots grown as described in the text. Tracers with starting lengths of 300 or 1200 NT were prepared as described previously (Murray et al., 1978). Tracer with a starting length of 8100 NT was prepared by five passes through a 26-gauge needle. The reassociation of each tracer was followed in the presence of excess unlabeled 300-NT leaf DNA as described in Figure 1. In each panel, the heavy line (●) represents the best least-squares fit to the data, the elements of which are summarized in Table II. The single-copy component for each tracer is indicated with a dashed line, and the light line (○) shows the reassociation of the *E. coli* internal kinetic standard. Indicated fragment lengths were measured after incubation to the approximate $C_0t_{1/2}$ of the single-copy component in each series to compensate for size degradation upon incubation. Degradation was significant only in the longest tracer.

of single-stranded DNA without producing noticeable shortening of long native DNA fragments (Murray et al., 1978).

Discussion

Figures 1 and 2 show that the single-copy sequences in total mung bean DNA reassociate ~105 times more slowly than *E. coli* DNA fragments in the same reassociation mixture, and thus the haploid mung bean genome contains $\sim 4.7 \times 10^8$ NTP or 0.48 pg of DNA. This value is in good agreement with cytophotometric measurements of 0.53 pg/haploid nucleus performed by M. D. Bennett (personal communication) on samples of our material. Both of these values are considerably smaller than our own preliminary kinetic data suggested (Murray et al., 1977b). We believe that the earlier estimate is in error primarily because too few data points were included in the later part of the C_0t curve and because the lack of an internal kinetic standard did not permit corrections for the reduction in reassociation rate attributable to viscosity effects in concentrated DNA solutions (Wetmur & Davidson, 1968).

The reassociation kinetics for total leaf DNA may be described with only one repetitive component in addition to the single-copy component (see Figure 1). While this is the simplest solution to the data, further experiments with isolated kinetic fractions show that a minimum of two repetitive components are required for an adequate description of the reassociation kinetics (Figure 2A and Table I). Even this solution is probably an oversimplification, however, inasmuch as the

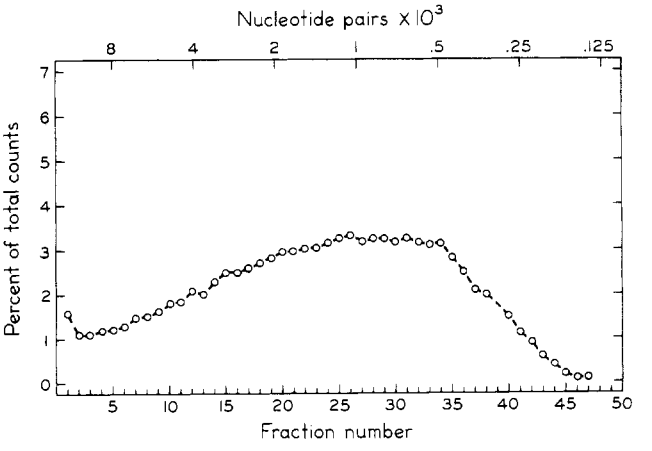

FIGURE 6: Size of S-1 nuclease resistant repeated sequences in mung bean DNA. 8100 NT long ^{3}H-labeled root DNA fragments were reassociated to an equivalent C_0t value of 10 and the size of the S-1 resistant duplexes measured by neutral agarose electrophoresis as described in the text.

kinetic data did not specify unique solutions and additional low-frequency repeats were detected in a more slowly reassociating fraction isolated after incubation to C_0t = 300 (Table IB). Thus, it seems probable that there is in fact a spectrum of repetition frequencies in mung bean DNA, ranging from about fifty to several thousand copies per haploid genome. We note that a recent study of cloned sea urchin repeated sequences demonstrates the presence of a broad range of repetition frequencies (Klein et al., 1978), even though previous work had shown that the sea urchin genome can be described in terms of a small number of simple kinetic components [e.g., Graham et al. (1974)].

Unidentified chloroplast sequences in a DNA preparation can seriously complicate analysis of sequence repetition and interspersion in plant DNA. Because pure preparations of plant nuclei are not generally available in quantity, several other plant studies have used leaf or shoot DNA [e.g., Murray et al. (1978) and Goldberg (1978)]. Our data illustrate the extent to which chloroplast sequences could contribute to the putative repetitive fraction in leaf DNA and emphasize the need for direct measurements in future work of this kind. By direct measurement using a purified chloroplast probe, our mung bean leaf DNA preparation was found to contain ~11% chloroplast sequences. Thus, about one-fourth of the apparent repeats in the mung bean leaf DNA preparation are actually of chloroplast origin. We consider the value of 11% for the contribution of chloroplast sequences to leaf DNA to be a minimal estimate since our tissue included ~20% epicotyl tissue, and mature leaves might contain considerably more chloroplast DNA. The problem of chloroplast contamination may be more acute in smaller genomes and/or in genomes where the fraction of repetitive sequences is small, as it is in mung beans. Indeed, the pea genome is some 9 times larger than the mung bean genome and an analogous pea leaf preparation was found to contain ~2.4% chloroplast DNA (Palmer et al., 1979).

Perhaps the most significant problem which could arise from the presence of unidentified chloroplast sequences involves the significance in plant genomes of very long, well matched repeated sequences. A popular hypothesis to explain the presence of such repeats in animal genomes supposes that they are mostly the products of recent amplification events and have thus not yet accumulated many base substitutions nor been translocated throughout the genome. Chloroplast sequences in mung bean leaf DNA clearly contribute a significant fraction of well matched duplexes that might easily be misconstrued as new repeated sequences (Figure 4). Thus, in higher plants, the significance of sequences capable of forming long, well matched duplexes in leaf preparations cannot be properly evaluated until the possible contribution of chloroplast sequences is accurately determined.

From the reassociation kinetics of long tracers shown in Figure 5, it is evident that there is considerable range in the size of single-copy sequences in the mung bean genome. By comparing the fraction of each tracer reassociating with single-copy kinetics, we calculate that ~35% of the single-copy DNA is bounded by repetitive sequences at intervals ranging between 300 and 1200 NTP and that ~18% is contained in sequences between 1200 and 6700 NTP in length. However, 46% of the single-copy sequences are longer than 6700 NTP.

Measurements of length-dependent increases in binding to hydroxylapatite at a single repetitive C_0t value are often used in attempts to analyze the distribution of single-copy sequence lengths in more detail. However, in the case of mung bean such experiments were unsatisfactory. While length-dependent increases in hydroxylapatite binding were consistently observed [Murray et al. (1977b) and unpublished data], the curves did not differ significantly from those which would result merely from the effect of tracer fragment length on single-copy reassociation rate in the absence of interspersion (Hinnebusch et al., 1978; Chamberlin et al., 1978). Similar effects may explain two conflicting reports on sequence organization in the chicken genome. On the basis of complete long tracer reassociation kinetics, Arthur & Strauss (1978) concluded that the chicken genome is organized in a long-period interspersion pattern. In contrast, Eden & Hendrick (1978) interpreted binding data at a single C_0t value as indicating the presence of a short-period component. However, we calculate that the binding data in the latter paper could result merely from the effects of fragment length on the reassociation rate of long single-copy sequences and thus cannot be construed as evidence for interspersion. Complications inherent in such experiments are discussed at more length elsewhere (Murray et al., 1978; Thompson & Murray, 1979).

It is noteworthy that the mung bean genome is the smallest higher plant genome yet characterized and also contains the largest fraction of long single-copy sequences.[2] While we do not know the minimum distance at which all of the single copy would be interspersed with repeats, it is clear that about half of the single-copy DNA (46% ≥ 6700 NTP) is organized in a pattern more closely resembling long-period than short-period interspersion. Heretofore, a significant amount of long-period single-copy DNA has been detected in the DNA of only one other higher plant. About 20% of the cotton genome is composed of single-copy sequences ≥ 4000 NTP long (Walbot & Dure, 1976). Cotton has a haploid genome size of ~0.8 pg. The remaining well characterized higher plant genomes (including several legumes) contain more than 1 pg of DNA and conform more closely to the short-period interspersion pattern [reviewed by Thompson & Murray (1979)]. An extreme example is the garden pea, where most single-copy sequences have a model length of ~300 NTP and essentially none are longer than 1000 NTP (Murray et al., 1978).

Some correlation between genome size and the pattern of sequence interspersion exists in the animal kingdom also. Several birds with genome sizes ranging from 1 to 1.65 pg have been shown to have predominantly long single-copy sequences (Arthur & Strauss, 1978; Epplen et al., 1978). All of the other organisms known to exhibit long-period interspersion have genome sizes smaller than 1 pg (Thompson & Murray, 1979). The most striking correlation occurs in dipteran insects, where fruit flies (0.1 pg) show exclusively long-period interspersion while house flies (0.8 pg) have extensive short-period interspersion (Crain et al., 1976; Manning et al., 1975).

The presence of different patterns of sequence interspersion in related species is most readily explained by large-scale events such as sequence amplification and translocation. Flavell and colleagues (Smith et al., 1976; Flavell et al., 1977) have shown that these processes have played an important role in the evolution of repetitive DNA in cereal grains. Several groups of repeated sequences have been identified which emerged at various times during the evolution of modern wheat. Individual members of these families have both diverged and been translocated, so that they are now interspersed throughout the genome.

Comparisons of genome size and single-copy sequence content are consistent with the view that the larger plant genomes have become so primarily by addition of repeated sequences [reviewed by Thompson & Murray (1979)]. One would predict that random translocation of short sequences in a large, predominantly repetitive genome such as that of the pea plant

would result in extensive interspersion of single-copy and repetitive sequences at fairly short intervals. However, in a small, less repetitive genome such as that of the mung bean a similar amount of translocation would intersperse single-copy sequences with other single-copy DNA more often than with repetitive elements and thus would not produce extensive repeat–single copy interspersion.

The foregoing evolutionary processes might be quite common in DNA which serves so strongly sequence-dependent function, "secondary DNA" in the terminology of Hinegardner (1976). Particularly in higher plants, most of the DNA might be secondary in this sense. For example, even a small higher plant genome such as that of the mung bean contains about 4 times more DNA on a haploid basis than *Drosophila*. If one assumes that mung beans do not require significantly more DNA than fruit flies for basic coding and regulatory functions, at least 75% of the mung bean genome might be secondary DNA. This figure may exceed 97% in an average-sized angiosperm genome such as that of the pea. Consequently, we consider most of the available information on genome organization in higher plants to be more directly relevant to the questions of genome evolution than to control of gene activity. Patterns of sequence arrangement that might be relevant to gene regulation may be obscured during analysis of total genome organization.

Future work will attempt to determine whether, in spite of overall differences in pea and mung bean genome organization, there are common patterns of sequence organization in the transcribed regions.

References

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

Arthur, R. R., & Strauss, N. A. (1978) *Can. J. Biochem. 56*, 257.

Bazetoux, S., Jouanin, L., & Huguet, T. (1978) *Nucleic Acids Res. 5*, 750.

Bendich, A. J. (1972) *Biochim. Biophys. Acta 272*, 494.

Bennett, M. D., & Smith, J. B. (1976) *Philos. Trans. R. Soc. London, Ser. B 274*, 227.

Britten, R. J., & Kohne, D. E. (1968) *Science 161*, 529.

Britten, R. J., & Davidson, E. H. (1969) *Science 165*, 349.

Britten, R. J., Graham, D. E., & Neufeld, B. R. (1974) *Methods Enzymol. 29E*, 363.

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

Cairns, J. (1963) *Cold Spring Harbor Symp. Quant. Biol. 28*, 43.

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

Chang, C. T., Hain, T. C., Hutton, J. R., & Wetmur, J. G. (1974) *Biopolymers 13*, 1874.

Crain, W. R., Davidson, E. H., & Britten, R. J. (1976) *Chromosoma 59*, 1.

Cuellar, R. E., Ford, G. A., Briggs, W. R., & Thompson, W. F. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 6026.

Davidson, E. H., & Britten, R. J. (1979) *Science 204*, 1052.

Davidson, E. H., Hough, B. R., Amenson, C. S., & Britten, R. J. (1973) *J. Mol. Biol. 77*, 1.

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

Davidson, E. H., Klein, W. H., & Britten, R. J. (1977) *Dev. Biol. 55*, 69.

Eden, F. C., & Hendrick, J. P. (1978) *Biochemistry 17*, 5838.

Epplen, J. T., Leipoldt, M., Engel, W., & Schmidtke, J. (1978) *Chromosoma 69*, 307.

Flavell, R. B., & Smith, D. B. (1976) *Heredity 37*, 231.

Flavell, R. B., Rimpau, J., & Smith, D. B. (1977) *Chromosoma 63*, 205.

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

Goldberg, R. B. (1978) *Biochem. Genet. 16*, 45.

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

Gurley, W. B., Hepburn, A. G., & Key, J. L. (1978) *Biochim. Biophys. Acta 561*, 167.

Hinegardner, R. (1976) in *Molecular Evolution* (Ayala, J., Ed.) p 179, Sinauer Associates, Sunderland, MA.

Hinnebusch, A. G., Clark, V. E., & Koltz, L. C. (1978) *Biochemistry 17*, 1521.

Hudspeth, M. E. S., Timberlake, W. E., & Goldberg, R. B. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 4332.

Huguet, T., Jouanin, L., & Bazetoux, S. (1975) *Plant Sci. Lett. 5*, 379.

Kiper, M., & Herzfeld, F. (1978) *Chromosoma 65*, 335.

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

Klotz, L. C., & Zimm, B. H. (1972) *J. Mol. Biol. 72*, 779.

Kolodner, R., & Tewari, K. K. (1975) *Biochim. Biophys. Acta 402*, 372.

Manning, J. E., Wolstenholme, D. R., Ryan, R. S., Hunter, J. A., & Richards, O. C. (1971) *Proc. Natl. Acad. Sci. U.S.A. 68*, 1169.

Manning, J. E., Schmid, C. W., & Davidson, N. (1975) *Cell 4*, 141.

Melchior, W. B., & Von Hippel, P. H. (1973) *Proc. Natl. Acad.* Sci. *U.S.A. 70*, 298.

Murray, M. G., Belford, H. S., & Thompson, W. F. (1977a) *Carnegie Inst. Washington, Yearb. 76*, 262.

Murray, M. G., Preisler, R. S., & Thompson, W. F. (1977b) *Carnegie Inst. Washington, Yearb. 76*, 240.

Murray, M. G., Cuellar, R. E., & Thompson, W. F. (1978) *Biochemistry 17*, 5781.

Palmer, J. D., Murray, M. G., & Thompson, W. F. (1979) *Carnegie Inst. Washington, Yearb.* (in press).

Pearson, W. R., Davidson, E. H., & Britten, R. J. (1977) *Nucleic Acids Res. 4*, 1727.

Preisler, R. S., & Thompson, W. F. (1978) *Carnegie Inst. Washington, Yearb. 77*, 323.

Schachat, F., O'Connor, D. J., & Epstein, H. F. (1978) *Biochim. Biophys. Acta 520*, 688.

Smith, D. B., & Flavell, R. B. (1977) *Biochim. Biophys. Acta 474*, 82.

Smith, D. B., Rimpau, J., & Flavell, R. B. (1976) *Nucleic Acids Res. 3*, 2811.

Thompson, W. F., & Murray, M. G. (1979) in *The Biochemistry of Plants: A Comprehensive Treatise VI* (Marcus, A., Ed.) Academic Press, New York (in press).

Walbot, V., & Dure, L. S. (1976) *J. Mol. Biol. 101*, 503.

Walbot, V., & Goldberg, R. (1978) in *Nucleic Acids in Plants* (Davies, J. W., & Hall, T., Eds.) CRC Press, Cleveland, OH.

Wetmur, J. G., & Davidson, N. (1968) *J. Mol. Biol. 31*, 349.

Wimpee, C. F., & Rawson, J. R. Y. (1979) *Biochim. Biophys. Acta 562*, 192.

Zimmerman, J. L., & Goldberg, R. G. (1977) *Chromosoma 59*, 227.