J Mol Evol 17, 31–42 (1981)

Journal of Molecular Evolution


# Ancient Repeated Sequences in the Pea and Mung Bean Genomes and Implications for Genome Evolution

Michael G. Murray, Debra L. Peters, and William F. Thompson

Department of Plant Biology, Carnegie Institution of Washington, Stanford, California 94305, USA

**Abstract.** Essentially all of the sequences in the pea (*Pisum sativum*) genome which reassociate with single copy kinetics at standard ($T_m$ -25°C) criterion follow repetitive kinetics at lower temperatures (about $T_m$ -35°C). Analysis of thermal stability profiles for presumptive single copy duplexes show that they contain substantial mismatch even when formed at standard criterion. Thus most of the sequences in the pea genome which are conventionally defined as "single copy" are actually "fossil repeats" – that is, they are members of extensively diverged (mutuated) and thus presumably ancient families of repeated sequences. Coding sequences as represented by a cDNA probe prepared from polysomal poly(A) + mRNA reassociate with single copy kinetics regardless of criterion and do not form mismatched duplexes. The coding regions thus appear to be composed of true single copy sequences but they cannot represent more than a few percent of the pea genome. Ancient diverged repeats are present, but not a prominent feature of the smaller mung bean (*Vigna radiata*) genome. An extension of a simple evolutionary model is proposed in which these and other differences in genome organization are considered to reflect different rates of sequence amplification or genome turnover during evolution. The model accounts for some of the differences between typical plant and animal genomes.

**Key words:** Pea – Mung bean – Genome organization – Evolution – Amplification – Repetitive DNA – Single copy DNA

*Offprint requests to:* Michael G. Murray

## Introduction

Eukaryotic genomes vary considerably both in haploid DNA content and in the fashion in which single copy and repeated sequences are arranged. It is striking that organisms of apparently similar biological and developmental complexity may show dramatically different genome organization. For example, in one genus of the Legume family, *Vicia*, there is at least a 7-fold variation in haploid DNA content (Bennett and Smith 1976). Fruit flies and house flies differ by 8-fold in genome size; the former shows long period interspersion while the latter shows short period interspersion (Crain et al. 1976; Manning et al. 1975). The pea and mung bean genomes differ in size by 9-fold and the mung bean genome contains many elements of the long period interspersion pattern while the pea shows only extremely short period interspersion (Murray et al. 1978; 1979). It does not seem reasonable to postulate that these differences in genome size and organization directly reflect fundamental differences in the way in which these organisms fulfill basic coding and regulatory requirements. Instead, we believe that many of the differences are more likely to involve DNA which serves no strongly sequence-dependent function – "secondary DNA" in the terminology of Hinegardner (1976).

Analysis of DNA reassociation kinetics has proved invaluable in elucidating the structure of eukaryotic genomes. Most reassociation experiments are performed at a temperature about 25°C below the melting temperature of native DNA, a criterion at which optimal reassociation rates are generally achieved. It is possible to estimate the fraction of the genome composed of "repetitive" and "single copy" sequences. Repeated sequences may then be further subdivided in terms of frequency. While it is a common practice to speak in

0022–2844/81/0017/0031/$ 02.40

terms of "middle" repetitive, "highly" repetitive, and etc., there are probably no such clear cut divisions in most genomes. Further analysis of repeat subfractions often reveals additional kinetic heterogeneity (e.g., Murray et al. 1978; 1979). Klein et al. (1978) have shown that while total sea urchin DNA reassociation may be modeled in terms of a few discrete repeat frequency classes, there is in reality a broad range in repeat frequency. It is generally believed that following amplification events which give rise to a family of repeats, the individual members diverge from one another with time as they accumulate random base substitutions (e.g., Britten et al. 1976). Divergence coupled with reamplification of certain repeat family members can produce a very broad spectrum in the degree of homology between members of a repeat family (Flavell et al. 1977; Preisler and Thompson 1981). Since mismatch affects reassociation rate (see below), divergence also contributes to apparent frequency heterogeneity. Thus it may be more appropriate to view repeated sequences as comprising a frequency/divergence continuum rather than belonging to discrete classes.

In a few cases, the fraction of the genome which consists of recognizably repetitive elements has been shown to be a function of the reassociation temperature. Most of these observations were made with plant DNA (e.g., Bendich and McCarthy 1970; Bendich and Anderson 1977; McCarthy and Farquhar 1972). When the reassociation temperature is reduced below that of a standard criterion such as $T_m$ -25°C to a more permissive criterion where more mismatched duplexes are allowed to form, the fraction of DNA which is scored as repetitive increases. Similar experiments have recently lead to the identification of some extensively diverged repeats in the presumptive single copy fraction of mouse and primate genomes (Ivanov and Markov 1978; Deininger and Schmidt 1979). Thus, as Kohne pointed out in 1970, the terms repetitive and single copy are largely operational definitions.

The operational definitions used in this paper are as follows. We will use the term "repetitive" for sequences which are present in more than a few (1–3) copies per haploid genome as determined kinetically at standard ($T_m$ -25°C) criterion. The term "single copy" refers to sequences which reassociate at the same criterion with kinetics consistent with their presence in only one copy per haploid genome. "Fossil repeats" are sequences which reassociate with single copy kinetics at standard criterion but which nevertheless form mismatched duplexes and can exhibit repetitive reassociation kinetics at lower temperatures (about $T_m$ -35°C). Finally we will term "true single copy"any sequences which follow single copy kinetics at both standard and permissive criterion and which do not form mismatched duplexes. We stress that these definitions are still completely arbitrary and that no clear cut divisions or discrete fractions may exist in reality. Even "true" single copy DNA may be very old repeat.

Only about 30% of the pea genome reassociates with single copy kinetics at $T_m$ -25°C, and sequence interspersion effects reduce this value to about 19% when assays are carried out with hydroxylapatite using DNA fragments 300 NT in length (Murray et al. 1978). Data presented here show that only a very small fraction of even this 19% may be classified as true single copy DNA since essentially all of it can be shown to behave as fossil repeats. Coding sequences in the pea genome as represented by a cDNA probe do appear to be true single copy sequences and thus the data indicate that coding sequences must constitute less than a few percent of total pea DNA. In contrast, in the smaller mung bean genome, fossil repeats are not as prominent and the majority of the "single copy" DNA does meet the above definition for "true" single copy sequences. These results are discussed in terms of a model for genome evolution in which different rates of sequence amplification can explain these differences in pea and mung bean as well as many other differences in plant and animal genomes.

## Experimental Procedures

*General.* Details of DNA purification, preparation of radioactive tracers, shearing, fragment length determination, and analysis of reassociation kinetics have been described previously (Murray et al. 1978; 1979).

*Single Copy Tracers.* Single copy-enriched pea and mung bean tracers were prepared by successive reassociation and fractionation on hydroxylapatite at standard criterion as detailed previously (Murray et al. 1978; 1979). Briefly, a fraction corresponding to 13% of the pea genome which remained single stranded after successive incubations to $C_0t$[1] 200 and $C_0t$ 1000, as well as an analogous fraction corresponding to 25% of the mung bean genome which was single stranded after successive incubations to $C_0t$ 5, $C_0t$ 100 and $C_0t$ 300, were isolated. 220 NT-long tracers were prepared from each by using calf oligodeoxyribonucleotides to prime second strand synthesis by *E. coli* DNA polymerase I. In the presence of excess homologous driver DNA (300 NT) about 60% of the reactable pea and 76% of the reactable mung bean tracers reassociated with the expected single copy kinetics at standard criterion (Murray et al. 1978; 1979).

Since neither tracer displayed homogeneous single copy reassociation kinetics, further purification was undertaken. Under the assumption that the residual repetitive sequences in each tracer might result from the failure of some of the repeats to form hydroxylapatite-bindable duplexes after a single round of reassociation (Davidson et al. 1973) each tracer was subjected

---

[1] Abbreviations used are: NT, nucleotides; NTP, nucleotide pairs; $C_0t$, the product of molar concentration of DNA nucleotides and time of incubation (mol s/L); $T_m$, the temperature at which half of the nucleotides are unpaired; $T_{m,i}$, the temperature at which half of the complementary strands are completely separated; PIPES, 1,4, Piperazinediethane sulfonic acid; PB, an equimolar mixture of $NaH_2PO_4$ and $Na_2HPO_4$ (pH 6.8)

to an additional reassociation in the presence of fresh unlabeled homologous total DNA. About 37% of the input $^3$H-labeled mung bean tracer remained single stranded after additional reassociation ($C_0t$ 300 at standard criterion in the presence of a 33-fold sequence excess of total unlabeled DNA). Essentially all of this material followed the predicted single copy kinetics upon subsequent analysis at standard criterion in the presence of 300-fold sequence excess of total mung bean DNA (Murray et al. 1979). Analogous treatment of the pea single copy enriched-tracer led to about 55% of the tracer remaining single stranded after additional incubation (to $C_0t$ 1000 in the presence of a 2.5-fold sequence excess of unlabeled DNA). The reassociation kinetics of the resulting tracer in the presence of a 700-fold sequence excess of unlabeled total pea DNA are presented below.

*Pea cDNA.* cDNA was prepared from poly(A) + mRNA isolated from light grown pea seedling leaves. Polysomes were isolated according to Gray and Cashmore (1976) with two modifications. The KCl in the homogenization buffer was reduced from 400 to 50 mM and 20 mM $VOSO_4$-UMP complex was included as a ribonuclease inhibitor (Egberts et al. 1977). 20 mM $VOSO_4$-UMP was found to give complete inhibition of ribonuclease in this buffer system (see Peters et al. 1979). Polysomal RNA was purified by CsCl centrifugation (Glisen et al. 1974) and poly(A) + RNA purified on poly(U) sepharose (Pharmacia) by a modification of the procedure of Ihle et al. (1974). Total polysomal RNA was loaded onto poly(U) sepharose in 100 mM NaCl, 20 mM Tris-HCl (pH 7.5), 1 mM EDTA and 0.1% SDS at 45°C and the column washed extensively with the same buffer. Bound RNA (5.0% of the total) was eluted at 55°C with the same buffer, but in the absence of NaCl. After a second poly(U) sepharose column step, during which 2.5% of the starting material remained in the bound fraction, the RNA was judged essentially free of rRNA contamination as judged by electrophoresis in glyoxal gels (McMaster and Carmichael 1977).

$^3$H-labeled cDNA with a specific activity of 9 x $10^6$ cpm/ug was prepared using AMV reverse transcriptase (kindly supplied by Dr. J.W. Beard) and oligo(dT) primer according to Meyers and Spiegelman (1978). This procedure includes 4 mM Na pyrophosphate to inhibit the synthesis of self-complementary cDNA. In our hands, self complementarity was reduced from about 60% to 6% of the incorporated label. Residual self-complementary and non-denaturable (foldback) cDNA was removed by pre-incubation and passage over hydroxylapapatite, using a cDNA $C_0t$ some 20 times greater than that attained in any subsequent experiments.

*Reassociation Kinetics.* The reassociation kinetics of pea and mung bean single copy tracers were followed in 1 M NaCl, 10 mM PB, 0.1 mM EDTA at 70°C or 60°C for standard ($T_m$ -25°C) or permissive ($T_m$ -35°C) criterion respectively (Angerer et al. 1976). While Angerer et al. (1976) reported that 1.0 M NaCl at 70°C gives 25-fold faster reassociation than 0.12 M PB at 60°C, we consistently observe only an 8-fold acceleration. Preliminary pea cDNA reassociation kinetics were performed in the 1.0 M NaCl buffer, but the experiments shown here were performed in 2 M ammonium sulfate, 20% ethanol, 10 mM PIPES (pH 6.9), 0.1 mM EDTA (D. Kohne, pers. comm.). We now prefer this buffer system to the NaCl buffer because it permits more rapid reassociation and at lower temperatures. Detailed analysis of the optical reassociation kinetics of *E. coli* and *B. subtilis* DNA as a function of temperature showed that optimal reassociation rates were achieved at 50°C, about 19°C below the $T_m$ for 300 NT fragments of either bacterial DNA (Murray, unpublished). 45°C and 35°C were chosen for standard ($T_m$ -25°C) and permissive ($T_m$ -35°C) criterion in ammonium sulphate-ethanol respectively. While the rates at 45°C are slightly suboptimal for well-matched DNA, the system still gives about 25-fold faster reassociation than does 0.12 M PB at 60°C ($T_m$ -25°C). Because the thermal stability profiles for pea single copy duplexes formed in 1.0 NaCl at 60°C or in 2 M ammonium

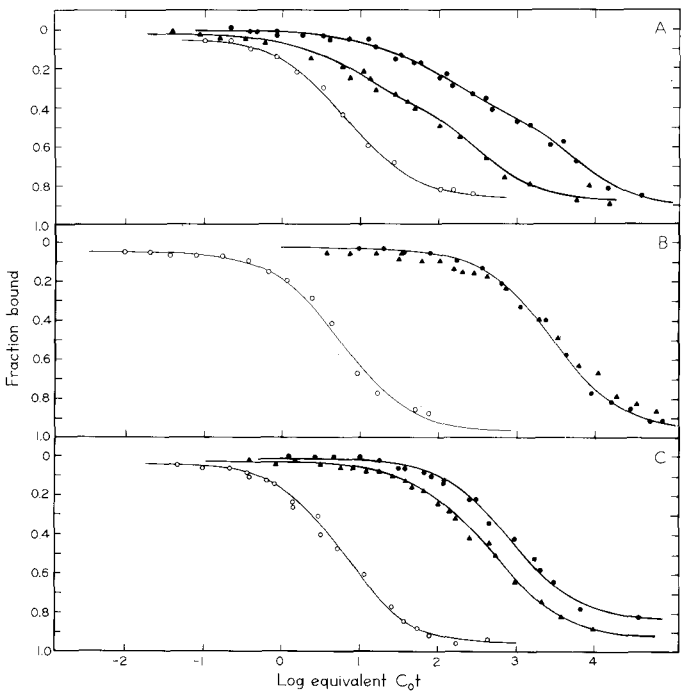


**Fig. 1 A-C.** Reassociation kinetics of pea and mung bean single copy tracers and pea cDNA driven by excess unlabeled total DNA at standard or permissive criteria. In each case, tracers were reassociated with 300 NT driver DNA fragments to the indicated equivalent $C_0t$ values prior to fractionation on hydroxylapatite as described in Experimental Procedures. Reassociation kinetics were followed at standard (•) and permissive (▲) criteria. All mixtures included 300 NT $^3$H- or $^{14}$C- *E. coli* DNA (○) fragments to provide internal kinetic standards. Data were corrected for the effects of viscosity and temperature by normalizing the observed rates for these standards to that for *E. coli* DNA reassociated alone at standard criterion. Reassociation kinetics were analyzed in terms of theoretical second order components as shown in Table 1. **A** Pea $^{14}$C-single copy-enriched tracer (220 NT) was reassociated with a 700-fold sequence excess of total pea DNA in 1.0 M NaCl, 10 mM PB, 0.1 mM EDTA at either 70°C or 60°C. Solid lines show the simplest least squares solutions consisting of 2 theoretical second order components. **B** $^3$H-pea cDNA was reassociated with a 2,000-fold mass excess of total pea DNA in 2 M ammonium sulfate-20% ethanol at 45°C or 35°C (see Experimental Procedures). cDNA lengths were 450 and 350 NT for standard and permissive criterion experiments respectively. Solid lines show the best least squares fits consisting of one theoretical second order component. **C** Mung bean $^3$H-single copy tracer (200 NT) was reassociated with a 300-fold sequence excess of total mung bean DNA in 1.0 M NaCl, 10 mM PB, 0.1 mM EDTA at 70°C or 60°C. Solid lines show the best least squares fits consisting of one theoretical second order component

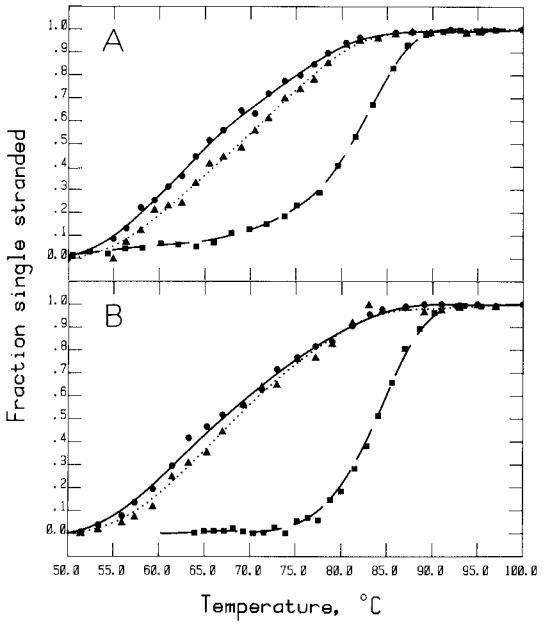


**Fig. 2 A and B.** Analysis of the thermal stability of pea single copy duplexes formed at permissive criteria. Single copy-enriched tracers (220 NT) were reassociated with a 700-fold sequence excess of 300 NT total pea DNA. All data were normalized to 100% and smooth lines fit to the data using a cubic spline computer program. **A** Single copy-enriched tracer was reassociated to either $C_0t$ 25 or $C_0t$ 500 in 1.0 M NaCl at 60°C prior to thermal denaturation in 0.12 M PB. The denaturation profiles are shown for the duplexes formed at $C_0t$ 500 before (●, *solid line*) or after (▲, *dotted line*) subtracting the profile for duplexes formed at $C_0t$ 25. Duplexes forming between $C_0t$ 25 and 500 represent 40% of the total duplexes. A thermal stability profile for pea cDNA (350 NT) reassociated in 2 M ammonium sulfate-20% ethanol at 35°C is also shown (■, *dashed line*). **B** Single copy tracer taken prior to the final round of reassociation with fresh total DNA (see Experimental Procedures) was reassociated to $C_0t$ 500 in either 2 M ammonium sulfate-20% ethanol at 35°C (●, *solid line*) or 1.0 M NaCl at 60°C (▲, *dotted line*) prior to thermal denaturation in 0.12 M PB. The thermal stability profile for 300 NT native pea is also shown (■, *dashed line*)

sulfate -20% ethanol at 35°C are nearly superimposable (Fig. 2B) we consider these conditions to provide equivalent permissive criteria. Finally the low temperature seems to reduce thermal degradation of DNA. We have seen little change in tracer reactivity in up to 4 weeks of incubation at 45°C (J.D. Palmer, unpublished).

Reassociation was monitored using hydroxylapatite as described previously (Britten et al. 1974; Murray et al. 1978). Single strands were eluted at 50°C or 60°C for standard and permissive criteria respectively, while duplex containing structures were eluted with 0.12 M PB at 100°C[2]. Single strands were eluted equally well with 0.12 PB at either 50°C or 60°C. This can be seen by comparison of the earliest $C_0t$ points in Fig. 1, and was also routinely observed during thermal denaturation experiments with the preheating technique described below. All reassociation mixtures included either $^3H$- or $^{14}C$- *E. coli* DNA (labeled in vivo) as an internal kinetic standard with which plant DNA data were corrected for the effects of temperature and viscosity on reassociation rate (see Murray et al. 1978).

Data are presented in the form of $C_0t$ curves (Britten and Kohne 1968). All $C_0t$ values have been corrected to the equivalent $C_0t$ for 0.12 M PB at 60°C using the *E. coli* standards. Results were analyzed in terms of theoretical second order components with the computer program described by Pearson et al. (1977).

*Thermal Stability Studies.* Samples of the same mixtures used for reassociation kinetic measurements were incubated as above and diluted into a large volume of cold 0.12 M PB. Individual aliquots were then preheated for 15 min at various temperatures using a linear thermal gradient block (Britten et al. 1978) prior to fractionation on hydroxylapatite at 50°C as described above. This technique was chosen instead of more conventional thermal elution procedures in order to avoid uncertainties in maintaining an elution "window" which cleanly discriminates between single-stranded and duplex containing structures over the entire temperature range (Martinson and Wagenaar 1977). Therefore we use $T_{m,i}$ instead of $T_E$ to distinguish data obtained with the gradient block from that obtained by thermal elution.

A cubic spline program was used to fit a smooth curve to the resulting data. Examples of the fit lines and actual data points for pea single copy-enriched tracer duplexes and native pea DNA (labeled in vivo) are shown in Fig. 2. In the case of the pea single copy-enriched tracer it is desirable to compare the thermal stabilities of the most slowly reassociating duplexes at each criterion. This was achieved by obtaining denaturation data at low and high $C_0t$ values ($C_0t$ 250 and 5000 for standard criterion and $C_0t$ 25 and 500 for permissive criterion). In principle, the fraction of low $C_0t$ duplexes was subtracted from the high $C_0t$ value for each temperature point. In practice this involved subtracting the fraction single stranded for each temperature in the high $C_0t$ sample from the corresponding value for low $C_0t$ duplexes. The resulting data were normalized to 100% and fit with a smooth curve using the spline program. First derivatives of the spline-fit lines were computed as described by Cuellar et al. (1978).

## Results

### *Pea Single Copy-Enriched Tracer*

The reassociation kinetics of pea single copy-enriched tracer are shown in Fig. 1A and summarized in Table 1. Under standard criterion conditions about 60% of the reactable tracer reassociated with a rate of 0.00022, in reasonable agreement with the predicted rate of 0.00015 for 220 NT single copy sequences (Murray et al. 1978). Lowering the temperature of reassociation by 10°C leads to quite different results. Under this more permissive criterion, virtually all of the pea tracer reassociates with repetitive kinetics. In the simplest 2 component solution to the data, the slowest component reassociates some 19 times faster than does the slowest component observed at standard criterion. Even when

[2]Ammonium sulfate concentrations of less than 5 mM do not affect hydroxylapatite fractionation. If the ammonium sulfate concentration will be higher when individual samples are diluted into PB the molarity of the later should be lowered to maintain 0.18 M total cation.

**Table 1.** Summary of reassociation kinetics at standard and permissive criteria for pea single copy, pea cDNA, and mung bean single copy tracers[a]

| DNA tracer | | Standard criterion | | | Permissive criterion | | | Permissive Rate |
|---|---|---|---|---|---|---|---|---|
| | | Fraction[b] | Rate | RMS[c] | Fraction[b] | Rate | RMS[c] | Standard Rate |
| Pea single copy | | | | | | | | |
| 2 Components | 1 | 0.41 | 0.014 | | 0.38 | 0.16 | | 11.4 |
| | 2 | 0.59 | 0.00022 | 0.015 | 0.62 | 0.0041 | 0.023 | 18.6 |
| 3 Components | 1 | 0.41 | 0.014 | | 0.32 | 0.21 | | 23.0 |
| | 2 | 0.46 | 0.00022 | | 0.60 | 0.0058 | | 26.3 |
| | 3 | 0.13 | 0.00022 (fixed) | 0.015 | 0.08 | 0.00022 (fixed) | 0.022 | 1.0 (fixed) |
| Pea cDNA | | | | | | | | |
| 1 Component | 1 | 1.0 | 0.00035 | 0.022 | 1.0 | 0.00037 | 0.017 | 1.4[d] |
| 2 Components | 1 | 0.10 | 0.0027 | | 0.10 | 0.017 | | 8.0[d] |
| | 2 | 0.89 | 0.00029 | 0.020 | 0.90 | 0.00031 | 0.012 | 1.4[d] |
| Mung bean single copy | | | | | | | | |
| 1 Component | 1 | 1.0 | 0.0012 | 0.018 | 1.0 | 0.0025 | 0.021 | 2.1 |
| 2 Components | 1 | 0.17 | 0.0088 | | 0.29 | 0.012 | | 1.4 |
| | 2 | 0.83 | 0.0008 | 0.012 | 0.71 | 0.0012 | 0.013 | 1.5 |

[a]DNA fragment lengths were as follows: all drivers, -300 NT; pea and mung bean single copy tracers, 220 NT; pea cDNA for standard criterion, 450 NT; pea cDNA for permissive criterion, 350 NT

[b]Fractions were normalized to 100% after subtracting residual foldback and unreactable tracer. Residual foldback ranged from 0 to 3%. Reactivity was 88–97% for pea single copy, 92–97% for pea cDNA, and 84–94% for mung bean single copy

[c]Root mean square error of the least squares solution to the data

[d]cDNA rates are not corrected for differences in fragment length. However, length differences are taken into account in the calculated ratio of the permissive to standard criterion rate

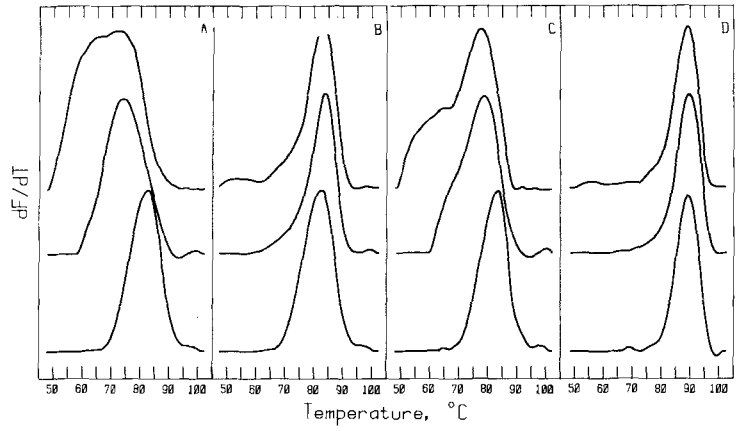


**Fig. 3 A-D.** Thermal stability profiles for pea and mung bean single copy, pea cDNA, and *E. coli* DNA duplexes formed at standard or permissive criteria. Portions of the reaction mixtures used in Fig. 1 were subjected to thermal denaturation as described in Experimental Procedures and illustrated in Fig. 1. First derivatives of lines fit to the integral data using a cubic spline computer program are shown here. In each panel the thermal stability profiles for native duplexes (*bottom*), duplexes formed at standard criterion (*middle*), and duplexes formed at permissive criterion (*top*) are shown. **A** Most slowly reassociating fraction of the pea single copy-enriched tracer; see Fig. 1 for details. **B** $^3$H-pea cDNA (350 NT), 80% double-stranded at standard criterion ($C_0t$ 21,000) and 70% double- stranded at permissive criterion ($C_0t$ 11,000); **C** $^3$H-mung bean single copy (220 NT), 61% double-stranded at standard criterion ($C_0t$ 2,000) and 73% double-stranded at permissive criterion ($C_0t$ 2,000); **D** $^3$H- *E. coli* (300 NT), 95% double-stranded at both standard and permissive criterion ($C_0t$ 200). Modes and $T_{m,i}$s for these profiles are summarized in Table 2

**Table 2.** Summary of thermal stability data for pea single copy and cDNA, mung bean single copy, and total *E. coli* DNA duplexes formed at standard and permissive criteria[a]

| DNA | Criterion | Mode | Δ Mode | $T_{m,i}$ | $\Delta T_{m,i}$ |
|---|---|---|---|---|---|
| Pea | Native | 83.0 | --- | 83.9 | --- |
| Single copy[b] | Standard | 76.0 (75.2) | -7.0 | 75.0 (74.2) | -8.9 |
| Single copy[b] | Permissive | 73.5 (72.7) | -9.5 | 70.0 (69.2) | -13.9 |
| cDNA | Standard | 84.2 (84.5) | +1.2 | 82.2 (82.5) | -1.7 |
| cDNA | Permissive | 83.2 (83.5) | -0.2 | 80.7 (81.0) | -3.2 |
| Mung bean | Native | 82.5 | --- | 82.5 | --- |
| Single copy | Standard | 80.3 (79.5) | -2.2 | 78.3 (77.5) | -4.2 |
| Single copy | Permissive | 80.5 (79.7) | -2.0 | 75.8 (75.0) | -6.7 |
| *E. coli* | Native | 89.7 | --- | 88.6 | --- |
| | Standard | 90.0 | +0.3 | 89.5 | +0.9 |
| | Permissive | 89.5 | -0.2 | 88.0 | -0.6 |

[a]Fragment lengths for all driver DNAs and *E. coli* tracers was 300 NT. Pea and mung bean single copy tracers were 220 NT long. Pea cDNA tracer was 350 NT long. Observed data for tracers (*parenthesis*) have been normalized to the values expected for 300 NT fragments using the relationship $\Delta T_m$ or Δmode = 650/L (Britten et al. 1974), where L is the tracer fragment length in NT measured by alkaline agarose gel electrophoresis

[b]Pea single copy data are for the most slowly reassociating sequences (see Fig. 1 legend and Experimental Methods)

the permissive criterion data are analyzed in terms of three theoretical second order components, with the rate constant for the slowest component being fixed at the value expected for single copy sequences, we can model only about 8% of the tracer as reassociating at the pea single copy rate. Since the entire tracer represents only about 13% of the pea genome, it is clear that only a small fraction of the pea genome can be classified as true single copy DNA.

Thermal stability profiles of the duplexes formed by this tracer (Fig. 3A, Table 2), provide additional evidence that most pea DNA sequences conventionally classified as a single copy are actually members of diverged repetitive DNA families (fossil repeats). Since a substantial repetitive fraction was observed by reassociation kinetics at standard criterion, it was desirable to remove the influence of the early reassociating sequences from the thermal stability data. This was achieved by taking the difference between profiles observed at $C_0t$ 5000 and those observed at $C_0t$ 250 as described in Experimental Procedures and Fig. 2A. The net effect of this procedure was an increase of about 3 and 2.5°C in the thermal stability of the most slowly reassociating duplexes formed at permissive and standard criterion respectively. Since we cannot exclude the possibility that repetitive duplexes readjust or "mature" to form more stable structures upon prolonged incubation (Beckmann and Daniel 1974; Ivanov et al. 1978), the thermal stabilities resulting from the subtraction procedure might be somewhat overestimated. Mismatch values for slowly reassociating DNA are thus minimum estimates. The most slowly reassociating duplexes at standard criterion show a modal thermal stability of 76°C and a $T_{m,i}$ of 75°C after applying corrections for small differences in fragment length (see Table 2). Upon reassociation at permissive criterion, the thermal stability profile becomes quite broad with a length-corrected $T_{m,i}$ of 70°C and very few duplexes melting in the range of native pea DNA. Comparing these data to that for 300 NT native pea DNA we estimate that the slowly reassociating duplexes formed at standard and permissive criteria contain about 9% and 14% mismatch respectively, assuming a 1°C reduction in thermal stability per 1% mismatch (Wetmur 1976).

Thermal stability for the *E. coli* DNA (Fig. 3D, Table 2) standards included in these reassociation mixtures confirm that the observed $T_{m,i}$ depression is not an experimental artifact, since the reassociated *E. coli* DNA samples melt with profiles nearly identical to that of native *E. coli* DNA.

### *Mung Bean Single Copy Tracer*

Reassociation of the mung bean single copy tracer can be adequately described with one theoretical second order component regardless of the reassociation criterion (Fig. 1C, Table 1). At standard criterion all of the single copy tracer reassociates with a rate constant of 0.0012, precisely as expected for mung bean single copy DNA of this length (Murray et al. 1979), while at the permissive criterion the reaction appears to proceed about 2 times faster. Somewhat better fits can be obtained by analyzing the data in terms of two theoretical second order components (Table 1). At both standard and permissive criteria, a portion of the tracer reassociates about 10 times more rapidly than does mung bean single copy DNA. However, the reassociation rate of the slowest component observed at permissive criterion is now only about 50% faster than the predicted single copy rate and thus (within reasonable limits of experimental error) following true single copy kinetics.

The mung bean tracer kinetics at both criteria are

close enough to those expected for single copy DNA so that one need not postulate a substantial fraction of fossil repeats on the basis of kinetic data alone. However, the thermal stability profiles (Fig. 3C, Table 2) do show the existence of some diverged repeats in this tracer preparation. Mung bean single copy duplexes formed at standard criterion show length-corrected $\Delta$ mode and $\Delta T_{m,i}$ values of 2°C and 4°C, respectively, relative to native 300 NT fragments. At permissive criterion, the mode is unchanged but a low stability shoulder lowers the overall $T_{m,i}$ to 75°C. Thus while most (ca. 70%) of the mung bean tracer behaves as true single copy sequences, some fossil repeats are present.

*Pea cDNA*

A cDNA probe prepared from leaf polysomal poly(A) + mRNA was used to determine whether the coding sequences in the pea genome are true single copy or composed of diverged repeats (Fig. 1B, Table 1). Reassociation kinetics at both standard and permissive criteria may be described with a single theoretical second order component. The rate constant observed at standard criterion (0.00035) is very close to that predicted for pea single copy sequences of this length (Murray et al. 1978). Under permissive conditions the cDNA tracer reassociates about 40% faster than observed at standard criterion after taking into account the differences in the cDNA lengths in the two experiments. These data are consistent with numerous observations (e.g. Britten and Davidson 1976; Goldberg et al. 1973) that most mRNAs are transcribed from sequences present in only one or few copies per haploid genome. However, the reassociation kinetics at both standard and permissive criteria may also be fit with two theoretical second order components (Table 1). In this case about 10% of the reactable cDNA reassociates about ten times more rapidly than does single copy DNA at standard criterion and about 50 times more rapidly at permissive criterion. Thus we cannot exclude the possibility that a small fraction of the sequences represented in this cDNA preparation are repetitive.

Reassociated cDNA duplexes show essentially native thermal stability whether formed at standard or permissive criterion (Fig. 3B, Table 2). However, the denaturation profiles are slightly skewed to the low stability side, and this effect is most obvious at permissive criterion. These data are consistent with the notion that a small fraction of the cDNA sequences are forming mismatched duplexes, as suggested by the 2 component fits to the kinetic data.

The contrast in the behavior of the cDNA tracer and the single copy-enriched tracer is not simply attributable to the use of the ammonium sulfate-ethanol buffer with the former and the 1.0 M NaCl buffer with the latter. Virtually indentical results for cDNA have been obtained using the 1.0 M NaCl buffer system (Peters et al. 1979). It is also clear from the thermal denaturation profiles in Fig. 2A that mismatched duplexes form equally well in either buffer system.

## Discussion

*Pea Fossil Repeats*

A major conclusion from the work reported here is that *virtually all* of the sequences in the pea genome which reassociate with single copy kinetics at standard criterion are actually diverged repeats (fossil repeats). Difficulties encountered during attempts to purify a pea single copy fraction provided the initial evidence for this phenomenon. A single copy-enriched tracer (prepared from the sequences remaining single stranded after incubation to $C_0t$ 200 and $C_0t$ 1000) appeared to be contaminated with about 40% residual repetitive sequences as judged by its reassociation kinetics in the presence of excess total DNA (Murray et al. 1978). Reasoning that some of the repeat contamination might result from formation of poor quality repetitive duplexes which failed to bind to hydroxylapatite during the initial fractionations (Davidson et al. 1973), we carried out an additional reassociation to $C_0t$ 1000 in the presence of fresh unlabeled DNA. Contaminating repeats should have reassociated with the additional unlabeled DNA and been removed upon subsequent hydroxylapatite fractionation. About half of the tracer was removed as expected. However about 40% of the remaining tracer fragments still reassociated with repetitive kinetics as shown in Fig. 1A. Thus no further purification was obtained inspite of the removal of enough apparently repetitive material to account for all of the original contamination.

The unusual behavior of the single copy-enriched tracer is best explained if most of it is actually composed primarily of extensively diverged repeats. While a large fraction of this tracer does display single copy kinetics at standard criterion, it is clear that these "single copy" duplexes are extensively mismatched (about 9%) and therefore are repetitive. When the criterion is lowered another 10°C and more extensively mismatched duplexes are allowed to form, the average mismatch increases to about 14% and now all of the tracer reassociates some 19 times faster than should true pea single copy DNA.

To some extent the fact that diverged repetitive sequences can show single copy reassociation kinetics at standard criterion must reflect the well known tendency for mismatched duplexes to form more slowly than perfectly matched duplexes (reviewed in Wetmur 1976). However, the average $T_{m,i}$ reduction for the duplexes formed at permissive criterion is only 14°C. From the data of Bonner et al. (1973) and Marsh and McCarthy (1974) we calculate that a repetitive family with this amount of mismatch should reassociate only about two

times more slowly than if it were composed of identical sequences. Moreover, the same family should reassociate only 5–35% faster at $T_m$ -35°C than at $T_m$ -25°C. These effects are clearly insufficient to explain the 19-fold increase in the reassociation rate at the permissive criterion. A reasonable explanation for the differences is that the rate increase also reflects the inclusion of more members of each family in the pool of sequences able to form stable duplexes when the criterion is lowered, with a corresponding increase in the effective copy number of each family.

In genomes such as the pea, it is therefore futile to attempt the isolation of a pure single copy fraction by repeated cycles of kinetic fractionation at standard criterion. Since collisions occur randomly between different pairs during each cycle of reassociation a given family member will have a certain finite probability of forming a stable complex with another family member and thus of being scored as repetitive. In pea, our data suggest that the probability is about 40%. However upon each cycle a substantial fraction of the statistically possible sequence combinations simply cannot form stable duplexes at standard criterion.

*Coding DNA*

The slowly reassociating tracer discussed above would include the entire complement of single copy sequences (longer than about 250 NT) present in the pea genome. Thus the observation that the great majority of these tracer sequences are actually diverged repeats places a severe limitation on the fraction of pea DNA which can be composed of coding sequences. Although the data provide no compelling support for the presence of *any* measurable fraction, we can construct theoretical models in which up to 8% of the tracer reacts with single copy kinetics even at the permissive criterion (Table 1). Considering the fraction of the total DNA represented in the tracer as originally isolated, we can estimate that between 1 and 3% of the pea genome might behave as true single copy DNA. However, most of the cDNA reassociation can be modeled with single copy kinetics at both standard and permissive criteria, and in both cases most of the duplexes formed during reassociation are precisely paired. The small amount of more rapidly reassociating material and low stability duplexes at permissive criterion would be readily explained by the presence of some diverged multigene families (e.g. Jones et al. 1979; for a review see Hood et al. 1975).

If our cDNA probe accurately reflects the properties of coding sequences in general, it follows that coding regions do fit our operational definition of true single copy but they contribute only a tiny fraction to the total mass of the pea genome. This conclusion is similar to that reached from a study of repeat sequence interspersion distances (Murray et al. 1978), but in this case our interpretation does not depend on the assumption that any non-coding inserts within the structural genes are non-repetitive.

*Mung Bean DNA*

True single copy sequences do appear to constitute a substantial fraction of the mung bean genome. A majority (ca. 70%) of the sequences which behave as single copy at the standard criterion continue to do so at the permissive criterion. However, it is interesting to note that in absolute terms the bean genome contains only slightly more true single copy DNA than might be present in the pea genome. The mung bean genome contains about $4.7 \times 10^8$ NTP of DNA per haploid genome, making its genome about nine times smaller than that of pea. By conventional definition about 65% of the mung bean genome is composed of single copy sequences. Taking 70% of these (or about 46% of the total DNA) as true single copy, we estimate the mass of true single copy sequences in the mung bean genome as about $2.1 \times 10^7$ NTP. If as much as 3% of the pea genome were true single copy DNA it would comprise about $1.4 \times 10^7$ NTP.

*Evolution of Secondary DNA*

Plant genomes vary widely in size and tend to be larger, as a group, than most animal genomes. There is a surprisingly consistent relationship between plant genome size and the amount of "single copy" DNA estimated kinetically at standard experimental criterion (Thompson and Murray 1980; Thompson et al. 1980; Flavell 1980), with the fraction of the genome comprised by this class decreasing rapidly with increasing total DNA content up to about 1–2 pg and then remaining more or less constant at 20–30% in all of the larger genomes up to at least 12 pg. When single copy DNA mass is plotted against haploid genome size, points for the larger plant genomes fall close to a line with a slope of about 0.2 (Thompson and Murray 1980; Thompson et al. 1980), suggesting that the rate of increase in "single copy" DNA is rather consistently about 20% of that for total DNA. In contrast, a similar plot for animal species yields a single line with a slope of 0.7–0.8, reflecting the much higher single copy DNA content of animal genomes[3].

[3]While most animal genomes have less than 3 pg of DNA, a few (mostly urodele amphibians) have much more. The available data on these large genomes is usually not sufficient to determine whether or not they fit on the same line, since most investigators have used the hydroxylapatite techniques know to be sensitive to differences in interspersion patterns. However, the S-1 nuclease data of Bozzoni and Beccari (1978) suggest that *Triturus* (23 pg) may have as much as 70% single copy DNA and *Necturus* (52 pg) about 50%. Thus even these extremely large animal genomes probably have higher proportions of single copy DNA than typical plant genomes.

We have argued previously that the large differences in genome organization for peas and mung beans, both legumes of similar developmental and biological complexity, are not likely to reflect similarly large differences in basic coding and regulatory functions (Murray et al. 1979; Thompson and Murray 1980). Clearly, the coding sequences in pea DNA do not constitute a readily measurable fraction of the total genome. On a priori grounds, coding sequences are more likely to constitute a significant fraction of the mung bean genome, however, even the mung bean genome is considerably larger than those of some other plants and animals with seemingly similar biological complexities. *Drosophila* and *Arabidopsis*, for example, have haploid DNA contents approximately 25% and 40% of the mung bean haploid value, respectively (see Manning et al. 1979; Bennett and Smith 1976). We have as of yet no evidence which supports any sequence-dependent function for most of the additional DNA in mung beans and nearly the entire DNA complement in peas. Hinegardner (1976) has argued that much of the DNA in eukaryotes is secondary in that it serves no strongly sequence-dependent role in coding of gene regulation and is therefore relatively free to evolve by amplification, deletion and base substitution. If most of the mung bean and nearly all of the pea genomes are secondary then it is clear that stochastic evolutionary processes are the dominant force in determining genome organization in each. In the following section we illustrate how simple assumptions about relative rates of these processes can explain large variations in genome organization.

Our model is an extension of that first proposed for repeat sequence evolution by Britten and Kohne (1969) and Kohne (1970). Most of the repetitive sequences in eukaryotic DNA are considered to originate from periodic amplification events, with members of the resulting sequence families being subject to translocation, base substitution, and deletion events during their subsequent evolutionary history. Flavell (1980) has recently extended this model to include repeated cycles of amplification within a single family and "compound" amplification events in which adjoining portions of unrelated sequences are amplified together. When both amplifications and deletions occur during evolution, the portion of the genome subject to these processes can be viewed as *turning over* on a geological time scale. Genome size will remain constant when amplification and deletion are in balance, while if one or the other predominates it will either expand or shrink.

To account for our data on pea DNA we must postulate nearly all (97–99%) of this genome has been derived from amplification events, so at least in this case a very large fraction of the total DNA must be evolving by amplification or turnover. In contrast, the smaller fraction of fossil repeats and larger fraction of "true" single copy DNA in mung beans might be explained by assuming that a much smaller fraction of DNA in this genome is evolving by rapid amplification and deletion while the rest is conserved. While we cannot exclude this possibility, we believe that a simpler explanation for the difference in the pea and mung bean genomes is that a large fraction of both genomes is turning over but that the rate of turnover is slower in mung beans.

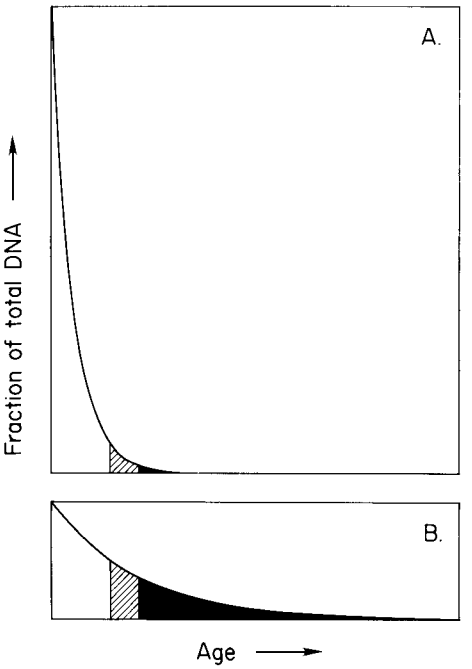


**Fig. 4 A and B.** Two hypothetical age distributions for secondary DNA in genomes characterized by different rates of amplification or turnover (amplification plus deletion). The area under each curve is the same, but the rate of amplification or turnover is 4-fold higher in **A** than in **B**. Sequence divergence from accumulated base substitutions will cause DNA from the older repeat families (■) to reassociate as though composed of single copy sequences, while younger families having less divergence may behave as repeats (□), or "fossil repeats" (▣) as described in the text. The exponential shape of the curves reflects the assumptions that both amplification and deletion events may affect any secondary DNA sequence with equal probability, and that the rates of these processes remain constant over long periods of time. Our model does not require strict conformance to either of these assumptions, and we expect actual age distributions to be more complex than those illustrated. The figure is intended merely to illustrate that differences in the long term average rates of amplification or turnover can have a profound effect on the fraction of secondary DNA which is perceived as repetitive at any arbitrary experimental criterion. For further discussion see text

Figure 4 illustrates two hypothetical age distributions for amplified DNA in genomes which have been characterized by different amplification or turnover rates for a long period of time. In both cases the older sequence families – those amplified earlier in evolution – constitute progressively smaller fractions of the total genome. In the absence of deletion events, this decline would simply reflect continuing increases in total DNA content from recurrent amplifications. When deletions are frequent, a similar decline would be produced by the progressive removal of individual sequences from

each family. Variations in the rates of these processes presumably occur during evolution so actual age distributions are probably more complex than those illustrated. The essential point as far as our model is concerned is only that the rate of amplification or turnover can have a pronounced effect on the age distribution of sequence families in the genome.

Since base substitutions accumulate with time, the fraction of previously amplified DNA which is still composed of sufficiently precise copies to be recognizably repetitive at a given criterion will depend on the rate of amplification or turnover which has characterized the evolution of the lineage. The crosshatched area under each curve in Fig. 4 represents the "fossil repeat" fraction as defined above – i.e., sequences capable of reassociating as repeats at $T_m$ -35°C. Older, more diverged sequences would continue to behave as "true" single copy DNA at the lower criterion even though they were derived from families whose members were at one time identical. It is readily apparent that there is no clear separation between any of these categories; they are defined simply by arbitrary points on smooth curves.

When a lineage has been characterized by rapid amplification or turnover the fraction of "true" single copy DNA will be small and "fossil repeats" will be prominent, as they are in pea DNA. On the other hand, slower rates will produce broader age distributions and thus yield larger fractions of "true" single copy DNA and smaller relative proportions of fossil repeats. This behavior would be similar to that of mung bean DNA.

According to this model, the small but relatively constant fraction of "single copy" DNA in most plant genomes can be explained by assuming that evolution of these genomes has been characterized by a high and relatively uniform turnover rate. Very small plant genomes such as that of the mung bean, as well as most animal genomes, would be viewed as turning over more slowly. The presence of "fossil repeats" in mouse (Ivanov and Markov 1978) and human (Deininger and Schmidt 1979) DNA is consistent with slow but significant turnover even in animals with relatively small fractions of repetitive DNA. In birds, where even smaller fractions of repetitive DNA are typical (e.g. Epplen et al. 1978) the turnover rate may be even lower. Burr and Schimke (1980) could detect no low stability duplexes in bird DNA reassociated at criteria comparable to those used here, although a substantial fraction of mismatched material was observed with even less stringent reassociation conditions.

A logical consequence of higher rates of amplification and turnover is an increase in the probability that secondary amplification events will affect one or more members of a previously amplified family. Secondary amplification has been shown to be quite common in large plant genomes (e.g. Flavell et al. 1977; Stein et al. 1979; Bedbrook et al. 1980). Extensive secondary amplification will result in repeat families which are heterogeneous in the sense of Bendich and Anderson (1977) – that is, composed of subfamilies amplified at different times and therefore showing different amounts of sequence divergence. Conversely, we expect that a lower rate of amplification will result in more homogeneous families of repeats where the range of divergence within a family is narrow. Indeed, detailed analysis of repeat sequence divergence reveal predominantly heterogeneous families in the pea genome and homogeneous families in the mung bean genome (Preisler and Thompson 1981).

We think it reasonable to suggest that patterns of sequence interspersion can in large part also be rationalized with this model. There is ample evidence that frequent transposition events have juxtaposed elements of old and new families of repeats (e.g. Rimpau et al. 1978; Smith and Flavell 1977). Flavell (1980) has argued that there may be some selection against excessive amounts of tandemly repeating DNA resulting from new amplification events, and thus in favor of frequent transposition to break up such arrays. We reported earlier that sequence organization in the pea genome is characterized by extensive interspersion at very short intervals with a modal length for "single copy" sequences (defined kinetically at $T_m$ -25°C) of only 300–400 NT (Murray et al. 1978). In view of the data shown here it is apparent that the foregoing interspersion studies actually revealed the interspersion of conventional repeats and extensively diverged repeats. In sharp contrast, most mung bean single copy sequences (defined the same way) are longer than 1000 NT and about half exceed 7000 NT (Murray et al. 1979). We visualize "repeat-single copy" interspersion as evolving from "repeat-repeat" interspersion as the older repeats diverge beyond the threshold whereby they are recognizably repetitive at standard criterion. When new repetitive families occupy a large fraction of the total DNA, as they will in genomes subject to rapid amplification or turnover, transposition of short elements of these families to regions of older DNA should produce a closer average spacing of the new elements than if a smaller amount of new DNA were interspersed in a larger amount of older material. The closer the spacing of the new sequences when originally inserted, the shorter will be the average length of "single copy" regions after further divergence of the older DNA. In addition, frequent additions of new DNA may decrease the average time period between interspersion events in a given chromosomal region. With a shorter time interval, only the oldest in a series of contiguous repeats will accumulate sufficient divergence to be considered single copy before a new transposition event occurs in that region.

The major difference between our hypothesis and previous models on which it is based (e.g., Kohne 1970) is that we visualize esentially the entire genome (with the exception of a few coding and regulatory sequences)

as participating in a largely stochastic turnover process. There is no need to postulate fundamentally different evolutionary histories for single copy and repetitive DNA, or to suppose that variations in the overall pattern of repetition and interspersion reflect selection for different patterns of gene regulation as proposed, for example, by Nagl (1978). The constancy of single copy DNA content among closely related species can be interpreted without supposing that DNA variation is somehow confined to particular combinations of sequences, as recently suggested by Hutchinson et al. (1980).

The turnover model of genome evolution is compatible with some aspects of the "selfish DNA" hypothesis advanced by Doolittle and Sapienza (1980) and Orgel and Crick (1980). In particular, it is quite possible to view turnover as being driven by intragenomic selection (Cavalier-Smith 1980) for DNA sequences particularly well adapted for amplification and/or transposition, but having little or no direct, sequence-dependent effect on phenotypic features above the DNA level. According to this view, diverged families of repeats, which make up most of the genome according to the turnover hypothesis, would be interpreted as degenerate descendants of sequences originally selected in this way. However, as noted by Cavalier-Smith (1980), "selfish DNA" or intragenomic selection cannot readily account for the wholesale variation of DNA content among eukaryotes unless it is accepted that nucleotypic selection, operating at the level of DNA mass without reference to particular sequences (Bennett 1972; Cavalier-Smith 1978) can favor both increases and decreases in total DNA content.

The dramatic and largely consistent difference in the proportion of repetitive DNA in plant and animal genomes (see above and Murray and Thompson 1980; Thompson et al. 1980) indicates that turnover rates have been higher during the evolution of most plant genomes than they have in a majority of animal species. Such a situation can, in principle, be readily explained by nucleotypic selection. Organisms unable to tolerate much variation in nuclear DNA content would be less likely to undergo the more frequent and/or extensive changes associated with high turnover, while lineages able to tolerate or exploit the nucleotypic effects of such changes would be much more likely to show a high turnover rate.

In general, then, we expect rapid amplification or high turnover to produce highly repetitive genomes with extensive short period interspersion, while slower turnover should result in a smaller fraction of recognizably repetitive DNA and an increase in the average interspersion difference. Differences between plants and animals can be interpreted as indicating differences in the extent to which nucleotypic flexibility has been an important evolutionary strategy. Acceptance of this view does not negate the possibility that *some* secondary DNA might participate in gene regulation or other cellular processes in a sequence-specific way, although we argue that the fraction involved must be quite small. What we do suggest is that variations in the overall structure of eukaryotic genomes are more likely to reflect variations in their evolutionary history than direct selection for specific sequences or organizational patterns.

*Acknowledgements.* 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). C.I.W. Department of Plant Biology publication No. 702. We thank Glenn Ford for his valuable assistance with computer programming.

## References

Angerer RC, Davidson EH, Britten RJ (1976) Chromosoma 56: 213–226
Beckman JS, Daniel V (1974) J Mol Biol 89:355–362
Bedbrook JR, O'Dell M, Flavell RB (1980) Nature 288:133–137
Bendich AJ, Anderson RS (1977) Biochemistry 16:4655–4663
Bendich AJ, McCarthy BJ (1970) Genetics 65:545–565
Bennett MD (1972) Proc Royal Soc Lond Ser B 181:109–135
Bennett MD, Smith JP (1976) Philos Trans R Soc Lond Ser B 274:227–273
Bonner TI, Brenner DJ, Neufeld BR, Britten RJ (1973) J Mol Biol 81:122–135
Bozzoni I, Beccari E (1978) Biochim Biophys Acta 520:245–252
Britten RJ, Davidson EH (1976) Fed Proc 35:2151–2157
Britten RJ, Kohne DE (1968) Science 161:529–540
Britten RJ, Cetta A, Davidson EH (1978) Cell 15:1175–1186
Britten RJ, Graham DE, Neufeld BR (1974) Methods Enzymol 29E:363–418
Britten RJ, Graham DE, Eden FC, Painchaud DM, Davidson EH (1976) J Mol Evol 9:1–23
Burr HE, Schimke RT (1980) J Mol Evol 15:291–307
Cavalier-Smith T (1980) Nature 285:617–618
Cavalier-Smith T (1978) J Cell Sci 34:247–278
Crain WR, Davidson EH, Britten RJ (1976) Chromosoma 59: 1–12
Cuellar RE, Ford GA, Briggs WR, Thompson WF (1978) Proc Natl Acad Sci USA 12:6026–6030
Davidson EH, Britten RJ (1979) Science 204:1052–1059
Davidson EH, Hough BR, Amen[illegible] CS, Britten RJ (1973) J Mol Biol 77:1–23
Deininger PL, Schmidt CW (1979) J Mol Biol 127:437–460
Doolittle WF, Sapienza C (1980) Nature 284:601–603
Egberts E, Hackett PB, Traub P (1977) Hoppe Seylers Z Physiol Chem 358:475–490
Epplen JT, Liepoldt M, Engel W, Schmidtke J (1978) Chromosoma 69:307–321
Flavell RB (1980) Ann Rev Plant Physiol 31:569–596
Flavell RB, Rimpau J, Smith DB (1977) Chromosoma 63: 205–222
Goldberg RB, Galau GA, Britten RJ, Davidson EH (1973) Proc Natl Acad Sci 12:3516–3520
Glisin W, Crkvenjakov R, Byus C (1974) Biochemistry 13: 2633–2637
Gray RE, Cashmore AR (1976) J Mol Biol 108:595–608
Hinegardner R (1976) In: Ayala FJ (ed) Molecular Evolution. Sinauer Assoc. Inc., Sunderland, p 179
Hood L, Campbell JH, Elgin SCR (1975) Ann Rev Genetics 9: 305–353
Hutchinson J, Narayan RKJ, Rees H (1980) Chromosoma 78: 137–145

Ihle JN, Lee KL, Kenney FT (1974) J Biol Chem 249:38–42
Ivanov IG, Markov GG (1978) Mol Cell Biochem 20:111–118
Ivanov I, Antonov P, Markova N, Markov G (1978) Mol Biol Rep 4:67–71
Jones CW, Rosenthal N, Rodakis GC, Kafatos FC (1979) Cell 18:1317–1332
Klein WH, Thomas TL, Lai C, Scheller RH, Britten RJ, Davidson EH (1978) Cell 14:889–900
Kohne DE (1970) Quart Rev Biophys 33:327–375
Manning JE , Schmid CW, Davidson N (1975) Cell 4:141–156
Marsh JL, McCarthy BJ (1974) Biochemistry 13:3382–3388
Martinson HG, Wagenaar EB (1977) Biochim Biophys Acta 474: 445–455
McCarthy BG, Farquhar MN (1972) In: Smith HH (ed) Evolution of Genetic Systems. Gordon and Breach, New York, p 1
McMaster GK, Carmichael GG (1977) Proc Natl Acad Sci 74: 4835–4838
Meyers JC, Spiegelman S (1978) Proc Natl Acad Sci 75:5329–5333
Murray MG, Palmer JD, Cuellar RE, Thompson WF (1980) Biochemistry 18:5259–5266
Murray MG Cuellar RE, Thompson WF (1978) Biochemistry 17:5781–5890
Nagl W (1978) Endopolyploidy and Polyteny in Differentiation and Evolution, North Holland Pub. Co., Amsterdam
Orgel LE, Crick FHC (1980) Nature 284:604–607
Pearson WR, Davidson EH, Britten RJ (1977) Nucleic Acids Research 4:1727–1735
Peters DL, Murray MG, Thompson WF (1979) Carnegie Inst Wash Year Book 78:208–212
Preisler RS, Thompson WF (1981) J Mol Evol (in press)
Rimpau J, Smith D, Flavell R (1978) J Mol Biol 123:327–359
Smith DB, Flavell RB (1977) Biochim Biophys Acta 474:82–97
Stein DB, Thompson WF, Belford HS (1979) J Mol Evol 13: 215–232
Thompson WF, Murray MG (1980) In: Davies DR, Hopwood DA (eds) Proceedings of the 4th John Innes Symposium. The Plant Genome and 2nd International Haploid Conference. John Innes Inst., Norwich, p 31
Thompson WF, Murray MG, Cuellar RE (1980) In: Leaver CJ (ed) Genome Organization and Expression in Plants. Plenum, New York, p 1
Wetmur JG (1976) Ann Rev Biophys Bioeng 5:337–361
Wilson AS, Carlson SS, White TJ (1977) Ann Rev Biochem 46:573–639

Received May 2, 1980/Revised September 9, 1980