Chromosoma (Berl.) 79, 251–270 (1980)

CHROMOSOMA

The Genome of *Zea mays*, Its Organization and Homology to Related Grasses

Sarah Hake and Virginia Walbot[1]

Department of Biology, Washington University, St. Louis, Missouri 63130, U.S.A.;
[1] addressee for reprint requests

Abstract. The pattern of genome organization of *Zea mays* has been analyzed, and the relationship of maize to possible progenitor species assessed by DNA:DNA hybridization. Reassociation of 470 and 1,350 bp fragments of maize DNA to various C_0t values demonstrates that the genome is composed of 3 major kinetic classes: highly repetitive, mid-repetitive, and unique. Mini-C_0t curves of the repetitive sequences at short fragment length indicate that the highly repetitive sequence class is 20% of the genome and is present at an average reiteration frequency of 800,000 copies; the mid-repetitive sequence class is 40% of the genome and is present at an average reiteration frequency of 1,000 copies. Thermal denaturation studies show that the highly repetitive sequences are 12% divergent and mid-repetitive sequences are 6% divergent. Most of the genome is organized in two interspersion patterns. One, approximately one-third of the genome, is composed of unique sequences of average length 2,100 bp interspersed with mid-repetitive sequences; the other, also one-third of the genome, is mid-repetitive sequences interspersed with highly repetitive sequences. The repetitive sequences are 500 to 1,000 bp by electron microscopic measurement. The remaining third of the genome is unique sequences farther than 5,000 bp from a palindromic or repetitive sequence. Hybridization of maize DNA from Midwestern Dent to popcorn and related grasses indicates that both the unique and repetitive sequence elements have diverged. Teosinte and popcorn are approximately equally divergent from Midwestern Dent whereas *Tripsacum* is much more divergent. The divergence times calculated from the depression of T_m in heterologous duplexes indicate that the divergence within *Zea mays* and between maize and near relatives is at least an order of magnitude greater than expected. This high degree of divergence may reflect the pressures of domestication of maize.

Introduction

Eukaryotes such as *Zea mays* L. have genomes much larger than would be expected to code for and regulate the expression of the genes required during

0009-5915/80/0079/0251/$04.00

development. Recent evidence indicates that eukaryotes express as many as 10^5 different gene size segments during the life cycle (Galau et al., 1976; Goldberg et al., 1978) whereas the genomes of many eukaryotes contain 5 or more times this sequence complexity in non-repetitive DNA. In all eukaryotes there is also repetitive DNA which functions either as gene sequences, rDNA and histones, for example, or in as yet undefined structrual roles in centromeres and other heterochromatic locations (see Pardue, 1975 for review). A portion of the repetitive DNA is interspersed with single copy DNA, the presumed Mendelian factors of the genome, in all higher eukaryotes thus far examined, and the pattern may contribute to the complexities of gene regulation in higher organisms (Davidson and Britten, 1973).

Studies on the organization of the genome have shown the presence of discrete patterns of sequence arrangements. The basic pattern found in higher eukaryotes is single copy DNA sequences, 1,000–2,000 bp long, interspersed with repetitive sequences, 200–800 bp long (Davidson et al., 1975). Individual genomes in many phyla have varying amounts of repetitive and single copy DNA organized in this pattern. The first exception to the pattern was *Drosophila*, which has single copy and repeat lengths approximately 10 fold longer (Manning et al., 1975). A number of plant genomes have been analyzed; basically, they have the short period pattern of organization as found in animals, but differ in the high proportion of repetitive DNA. Animal genomes contain an average of 10–30% repetitive DNA and plants contain an average of 50–80%. Part of the difference may be due to the fact plants tend to have larger genomes (Flavell et al., 1974).

A primary question in plant genome studies is the organization, divergence, and role of the repetitive DNA since repetitive DNA amount varies substantially within many genera and even within species (Walbot and Goldberg, 1979). Repetitive DNA in plants is found in three possible arrangements: interspersed with single copy DNA, interspersed with repetitive DNA of different reiteration frequency, or in long uninterrupted similar repeats. Repetitive DNA organized in different patterns may vary in the amount of sequence divergence present. In the soybean genome the short interspersed repetitive DNA has higher sequence divergence than longer, non-interspersed repeated sequences (Goldberg, 1978). In muskmelon the very highly repeated satellite DNA has no detectable sequence divergence even in the sequences that have interspersed into the main band DNA (Bendich and Taylor, 1977).

Our interest in defining the organization of the maize genome is several fold. Among plants, *Zea mays* has the best understood genetics; the 10 chromosomes are well mapped and complete stocks of monosomics and A–B translocations are available for genetic manipulation. The modification of gene activity by controlling elements in maize has been well documented by McClintock and others (see Fincham and Sastry, 1974 for review) and is one of the most complex and elegant examples of gene regulation known. Molecular studies in other species have demonstrated interesting facts about the organization of specific genes which were not expected from genetic studies. To maintain perspective it is important to understand the general features of genome organization in maize especially of the single copy and contiguous repetitive sequences for comparison to detailed studies of individual gene regions. Second, we are

interested in defining the origin of maize from suspected progenitors and the possible contributions such progenitors may have made to the current organization of the maize genome. Third, we want to understand the function and evolution of the highly heterochromatic regions of the maize genome such as knobs and B chromosomes.

Materials and Methods

Plant Material. Plants were grown from the following seeds: *Zea mays* W64A from J. Scandalios; *Zea mays* Mo17 from E. Coe; Ladyfinger Pop 7427-"40" from J. Beckett; *Zea mexicana*, race Chalco 331779, *Zea mexicana*, race Guatemala 306615, and *Tripsacum laxus* 314908 from the Southern Regional Plant Introduction Station. Wheat was taken from the Washington University greenhouse as cuttings of plants. Additional *Zea mexicana* seeds were obtained from G. Beadle and used in pilot experiments.

DNA Preparation. DNA was extracted from 8 day old etiolated *Zea mays* seedlings (line W64A). The method was essentially that of Goldberg (1978) modified from Britten et al. (1970). Seedlings, frozen in liquid N_2 and ground in a Waring blender, were first extracted with freshly distilled phenol:sevag (chloroform: isoamyl alcohol, 24:1), followed by two sevag extractions, one with butanol to 10%. The slurry was mixed with HAP[1] (BioRad), equilibrated with UP in large centrifuge bottles. The HAP was allowed to settle, and the supernatant poured off. When all the RNA was removed, the HAP was centrifuged at $200 \times g$, and the DNA eluted with 0.48M PB. The eluant was dialyzed against 0.2 M NaAcetate, precipitated with 2 volumes ethanol and resuspended in $1 \times$SSC. The resulting DNA was free of RNA and carbohydrate contaminants; the yield was 0.05 mg/gm fresh weight of tissue. The weight average molecular size was 5–10 Kbp with a range from 1–60 Kbp. Preparations enriched for long fragment lengths were obtained by fractionating the mixture of size classes on a 5–20% linear sucrose gradient in 1 M NaCl, 100 mM EDTA centrifuged for 6 hr at $100{,}000 \times g$ in a SW40 rotor and recovery of DNA of desired length. The optical properties of the DNA were as follows:

$$\frac{A_{260\,nm}}{A_{280\,nm}} = 1.9, \quad \frac{A_{260\,nm}}{A_{230\,nm}} = 2.3$$

as expected for a DNA of this base composition (Felsenfeld, 1971). The hyperchromicity of native DNA was 27% calculated as:

$$\frac{A_{260\,nm,\,98°C} - A_{260\,nm,\,50°C}}{A_{260\,nm,\,98°C}}.$$

Buoyant density was determined on native DNA in neutral CsCl in a Model E analytical ultracentrifuge relative to *Micrococcus luteus* ($\rho = 1.731$ gm cm^{-3}).

Shearing and Sizing DNA. DNA was sheared according to Britten et al. (1974). Shearing the DNA at 40,000 rpm in 66% glycerol in a Virtis 60 homogenizer routinely produced molecules of about 400 bp. Single or double stranded fragments were sized by electron microscopy assuming 2.07×10^6 daltons of DNA per μm (Cairns, 1963) and using φX174 single-stranded closed circular DNA as a standard: 5375 bp (Sanger et al., 1977).

Labeling and Removal of Zero-Time Binding DNA. DNA samples were labeled in vitro with ^{3}H-TTP using the nick translation procedure of Mackey et al. (1977). Zero-time binding DNA was removed

[1] *Abbreviations.* HAP hydroxylapatite; UP 8 M urea, 0.24 M sodiumphosphate buffer; SSC 0.15 M NaCl, 0.015 M NaCitrate; Kbp kilobase pairs; PB sodium phosphate buffer, equimolar mono- and dibasic; ds double-stranded molecules; ss single-stranded molecules.

by boiling the sample diluted with 0.12 M PB for 5 min, applying it to HAP at 60° C, and eluting it with 0.12 M PB. The zero-time binding DNA was eluted with 0.48 M PB and was 5–7% of the total. DNA free of zero-time binding DNA had a specific activity of $15–40 \times 10^6$ cpm/μg.

Reassociation Techniques. DNA was reassociated in 1 × SSC or 1.5 M Na^+ (7.7 × SSC) at $T_m - 25°$, except for preparation of repetitive fractions in which dilute DNA was reassociated in 0.12 M PB at $T_m - 25°$ C. Reassociations were terminated by freezing samples in ethanol and dry ice and storage at −4° C until analysis. The extent of reassociation was determined by separating ds and ss on HAP at 61° C and measuring the absorbance of each fraction. Recovery of DNA from HAP was routinely 95–105% of the column input.

DNA-DNA Hybridization and Melting Curves. Hybridizations were performed in 0.12 M PB or 0.48 M PB with 20–30,000 cpm of tracer and 100 μg 400 bp sheared driver DNA (ratio >1:10,000) at criterion temperatures. The duplex DNA molecules were bound to 1 ml HAP in 0.12 M PB at 60° C and washed with 5–10 ml 60° C buffer. The DNA was eluted with 2.5 ml 0.12 M PB in 4° C increments up to 100° C. DNA not eluted at 60° C but eluted at temperatures up to 100° C including a final 0.48 M PB wash, was taken as 100% of amount bound. The amount of unlabeled DNA released was determined by its $A_{260\ nm}$; labeled DNA was determined by precipitation with trichloroacetic acid (final concentration of 10%) at 4° C and collection on glass fiber filters for scintillation counting.

S1 Nuclease Treatment and Electron Microscopy of Reassociated Molecules. S1 nuclease experiments followed the procedure of Goldberg (1978) using conditions which digested 90–95% of denatured DNA. 20,000 units S1 nuclease (Miles Labs) per mg of DNA were routinely used. Following S1 nuclease treatment, molecules were spread directly for electron microscopy as previously described (Walbot and Dure, 1976). HAP chromatography was used to separate S1 nuclease resistant duplexes from digested DNA. The percent resistance of S1 nuclease was measured from the $A_{260\ nm}$.

Isolation of Kinetic Fractions. Kinetic fractions of maize DNA enriched in highly repetitive and mid-repetitive sequences were prepared from sheared DNA. Mid-repetitive DNA was bound to HAP at C_0t 100 but not bound at C_0t 0.1. The highly repetitive DNA was DNA that bound to HAP at C_0t 0.1.

Optical Melting. Thermal denaturation profiles of DNA were performed in a Gilford model 250 equipped with a thermal programmer 2527 and chart recorder 6051. The DNA samples were melted in 0.1 × SSC and 1 × SSC at concentrations of 10 μg/ml. Hyperchromicity was calculated as the fraction of denatured optical density between 50° and 98°. The hyperchromicity was also determined by base denaturation. The T_m was taken as that temperature which produced a 50% increase in hyperchromicity. The dispersion, at $\sigma_{2/3}$, was calculated by the method of Mahler and Dutton (1964).

Results and Discussion

Sequence Components in Maize DNA

Reassociation experiments were performed with fragments sheared to 1,350, 1,100 and 470 base pairs. The reassociation data were analyzed by a sum of squares fitting computer program (Pearson et al., 1977). At all fragment lengths used computer analysis shows three major kinetic classes: rapidly reassociating (low C_0t), fast reassociating (mid C_0t), and slow reassociating (high C_0t) DNA sequences; these three kinetic classes are general features of all eukaryotic DNAs so far examined (Walbot and Goldberg, 1979). The apparent percent contribution of each major kinetic classes can be determined from the fraction

Table 1. Kinetic Analysis of maize DNA at 1,350 and 470 Base Pair Fragments

Reassocation Curve	Component	Fraction of Fragments[a]	K[b]	$C_0t_{1/2\ pure}$[c]	Complexity in daltons[d]	Reiteration[e]
Whole genome 1,350 bp	highly rep.	0.415	27.66	0.015	1.96×10^7	79,333
	mid-rep.	0.167	0.159	1.050	1.37×10^9	1,133
	non rep.	0.302	2.54×10^{-4}	119,0	1.55×10^{12}	1
Whole genome 470 bp[f]	highly rep.	0.17	20.38	0.0083	6.43×10^6	296,000[g]
	mid-rep.	0.36	0.0427	8.431	6.50×10^9	280[g]

[a] The fraction of fragments which contain a duplex region

[b] K is expressed in $M^{-1}sec^{-1}$ and represents the 2nd order rate constant

[c] $C_0t_{1/2\ pure} = (1/K)$ (fraction of fragments)

[d] Complexity is calculated by comparison of the rate of reassociation of each component of maize DNA normalized to 300 bp to that of *E. coli* DNA at 300 bp: K = 0.22 (Goldberg, 1978), complexity = 2.8×10^9 daltons

[e] A value of 1 is assumed for single-copy components. The number of copies of repetitive DNA is ($C_0t_{1/2\ pure}$ single-copy/$C_0t_{1/2\ pure}$ component)

[f] C_0t analysis at this length stopped at C_0t 1000 and did not include unique sequence DNA

[g] The reiteration was derived from a theoretical $C_0t_{1/2\ pure}$ for the unique DNA at this length from the equation $(L_a/L_b)^{1/2}(K_b) = K_a$

bound to HAP as shown for 1,350 bp DNA (Table 1, Fig. 1). The low C_0t DNA appears to be a large portion of the total genome when whole genomic fragments of 1,350 bp are analyzed, due to the interspersion of this component with mid C_0t DNA. Mini-C_0t data, presented later, show that this low C_0t component is composed of several subclasses. When analyzed as a single component the overall $C_0t_{1/2\ pure}$ of this DNA fraction is 0.015, equivalent to a complexity of 2×10^7 daltons. The mid C_0t component of whole genomic DNA at 1,350 bp fragments is about 20% of the genome with a $C_0t_{1/2\ pure}$ of 1.05 and a complexity of 1.37×10^9. The slow reassociating DNA, high C_0t, is approximately 30% of the genome at all fragment lengths analyzed. The $C_0t_{1/2\ pure}$ of this component at 1,350 bp is 1,190 which gives a complexity of 1.55×10^{12}. Eighty-eight percent of the genome is accounted for as determined by HAP analysis of the reassociation. Three percent of the unaccounted for DNA is foldback DNA, which reassociates instantaneously and is not a consequence of bimolecular collision. The other 8% of the DNA does not reassociate even at high C_0t due to heat degradation after long incubations (Zimmerman and Goldberg, 1977). The genome size can be calculated from the complexity and proportion of the genome in the high C_0t class. The $C_0t_{1/2\ pure}$ of the high C_0t class is 2,530 when corrected to a fragment length of 300 bp, $\left(\frac{L_a}{L_b} \times K_b = K_a\right)$ (Britten et al., 1974); and the complexity is 1.5×10^{12} daltons. At this fragment length the high C_0t

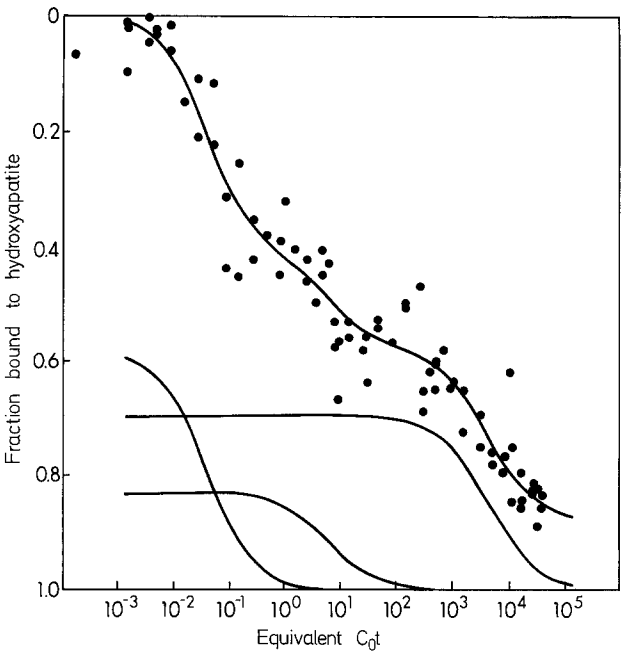

Fig. 1. Reassociation kinetics of unfractionated *Zea mays* DNA. Samples of 1,350 bp long fragments of *Zea mays* DNA were denatured and reassociated to the indicated equivalent C_0t values as described in Materials and Methods. The fraction of fragments containing a duplex region in each sample was determined by hydroxyapatite chromatography. The curve represents a least squares solution for three second order components. The computer-calculated second order solutions for each component are shown in the lower part of the figure

DNA is 35% of the genome. When the total genome is normalized to 100%, the high C_0t DNA becomes 40% (35/88 × 100). The haploid genome is 3.76×10^{12} daltons $\frac{(1.5 \times 10^{12})}{0.40}$ or 6.3 pg per haploid cell. Comparable values for the genome size of 5.5–7.0 pg DNA per cell have been determined by cytophotometric measurements (Ogur and Rosen, 1950) and reassociation experiments (Flavell et al., 1974).

Reassociation at Short Fragment Lengths

Maize DNA was also reassociated to low and mid C_0t values at shorter fragment lengths, 470 base pairs and 1,100 base pairs. The percent contribution from highly repetitive and mid-repetitive classes changes with fragment length (Table 1). There is an apparent shift of DNA from the mid-repetitive to the highly repetitive class at longer lengths. This presumably results from interspersion of highly repetitive and mid-repetitive DNA. These data also allow an estimate of the size of the middle repetitive sequences that are interspersed with highly repeated DNA. Between 470 and 1,100 bp the difference in amount of highly repetitive

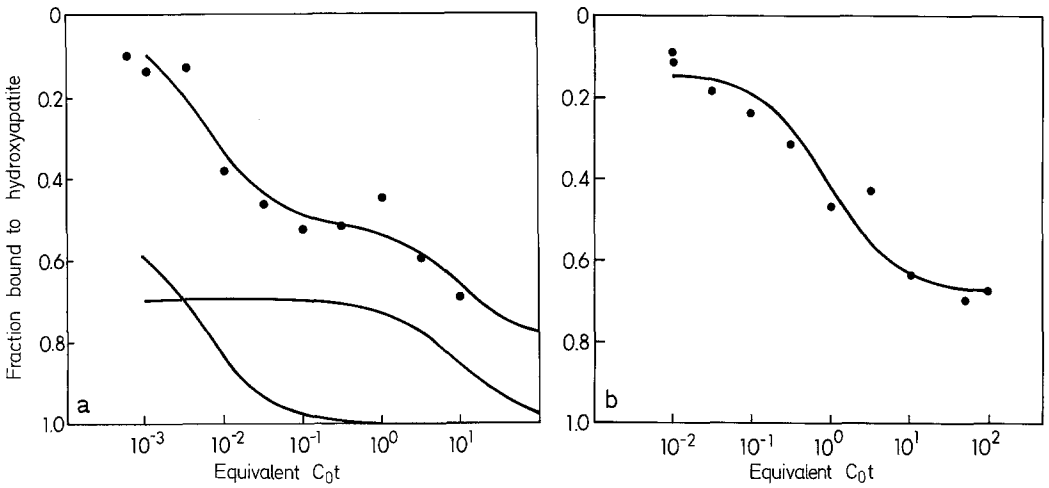

Fig. 2. a Reassociation of the highly repetitive fraction. A highly repetitive fraction was prepared from 400 bp fragments as described in Materials and Methods. The solid line represents a best least squares solution for two second order components. **b** Reassociation of the mid-repetitive fraction. A mid-repetitive fraction was prepared from 400 bp fragments as described in Materials and Methods. The solid line represents a best least squares solution for one second order component

DNA is minimal, 17% at 470 bp, 20% at 1,100 bp. Between 1,100 and 1,350 bp there is a large shift of mid-repeat DNA into the high repeat class, from 20% to 41%, suggesting that middle repetitive sequences of approximately 1,000 bp are interspersed with highly repetitive sequences. More exact values of the sequence lengths were determined by electron microscopy and are discussed later.

Reassociation of Enriched Kinetic Fractions

Mini-C_0t curves were generated to obtain a direct estimate of the C_0t pure values of individual sequence components. DNA fragments, sheared to 400 bp and highly enriched in each repetitive class, were prepared as described in Materials and Methods. The mid-repetitive DNA was defined as DNA that bound at C_0t 100 but did not bind at C_0t 0.1; highly repetitive DNA was that which bound at C_0t 0.1. The resulting DNA samples were reassociated to various C_0t values (Fig. 2 and Table 2). Computer analysis of the highly repetitive DNA data demonstrates 2 components with 1% error. The faster of the two components, two-thirds of the total, has a very low $C_0t_{1/2\ \text{pure}}$ value, 0.0014 (values normalized to 1,350 fragment length). It contains sequences of 1.57×10^6 dalton complexity. The slower component is typical of the mid-repetitive class in C_0t value, $C_0t_{1/2\ \text{pure}}$ equal to 1.57, complexity equal to 1.78×10^9 daltons. Computer analysis of the mid-repetitive class also shows 2 components with 1% error (as opposed to 1 component with 3% error). The faster component, approximately half of the total, has a $C_0t_{1/2\ \text{pure}}$ of

Table 2. Reassociation Kinetics of Enriched Kinetic Fractions

Reassocation curve	Component	Fraction of fragments[a]	K[b]	$C_0t_{1/2\ pure}$[c]	Complexity in daltons[d]	Reiteration[e]
$C_0t < 0.1$						
highly repetitive	highly rep.	0.48	188.3	0.00252	1.57×10^6	870,000
400 bp	mid-rep.	0.30	0.104	2.88	1.78×10^9	880
C_0t 0.1–100						
mid repetitive	highly rep.	0.31	12.39	0.025	1.54×10^7	87,000
400 bp	mid-rep.	0.36	0.153	2.35	1.45×10^9	930
	mid-rep.	0.54	1.09	0.495	3.05×10^5	4,400

[a] The fraction of fragments which contain a duplex region

[b] K is expressed in $M^{-1}sec^{-1}$ and represents the 2nd order rate constant

[c] $C_0t_{1/2\ pure} = (1/K)$ (fraction of fragments). Data is normalized to 1,350 bp fragment length for comparison to total DNA genomes analyzed

[d] Complexity is calculated by comparison of the rate of reassociation of each component maize DNA to that of *E. coli* DNA at 300 bp: $K = 0.22$ (Goldberg, 1978) complexity $= 2.8 \times 10^9$ daltons

[e] A value of 1 is assumed for single-copy components. The number of copies of repetitive DNA is ($C_0t_{1/2\ pure}$ single copy/$C_0t_{1/2\ pure}$ component). The $C_0t_{1/2\ pure}$ for single copy was taken from the whole genome reassociation kinetics, normalized to 400 bp fragment length $(La/Lb)^1/_2$ (Kb) = Ka

[f] The mid-repetitive mini-C_0t curve analyzed as 1 component instead of 2

0.014. The slower half has a $C_0t_{1/2\,pure}$ of 1.28. If the mid-repetitive class is treated as one component, the $C_0t_{1/2\ pure}$ is 0.495 and the complexity is 3.05×10^8 daltons. From analysis of the mini-C_0t curves it is evident that each component is contaminated to some extent with the other. This is due partly to the fact that highly repetitive and mid-repetitive DNA sequences are interspersed. The analysis also shows that the two classes of repetitive DNA are not completely distinct and a continuum of $C_0t_{1/2}$ values can be found between the two main values.

The number of copies of repetitive DNA can be calculated by comparison to single copy DNA. At the 1,350 bp fragment length there are approximately 80,000 copies of each very fast entity: if shorter DNA is analyzed, e.g. 400 bp fragments, the apparent reiteration of this component increases to 870,000. This fraction also contains the palindromic DNA. Estimates of the reiteration of the mid C_0t component range from 300 to 1,000 for each family. The variability in these estimates is due to the relationship between the length of molecules used in the experiment compared to the actual average length of members of each kinetic class. Small changes in fragment length can produce profound changes in the apparent reassociation kinetics especially since the mid-repetitive and highly repetitive sequences are interspersed. There is no evidence for the presence of a low reiteration repetitive components, i.e., less than 100 copies.

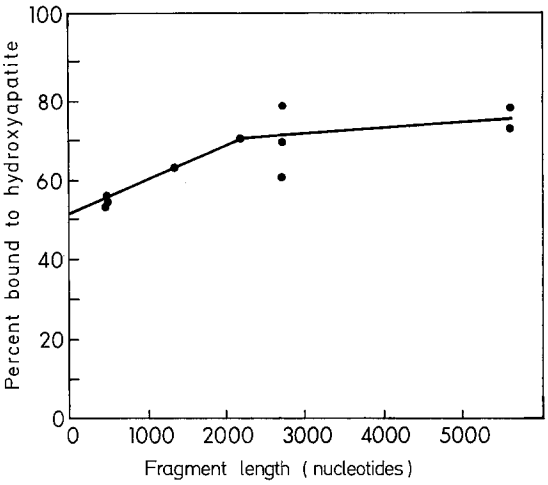

Fig. 3. The percentage hydroxyapatite binding versus fragment length for C_0t 50 reassociation. The DNA fragments were reassociated to C_0t 50, and the percentage binding to hydroxyapatite determined by monitoring the absorbance at $A_{260\ nm}$ of the fraction

Lengths of Repetitive and Unique Sequences

The length of the unique sequences was determined by reassociating DNA of different fragment lengths to C_0t 50. The percent DNA bound to HAP was plotted versus fragment length (Fig. 3). At this C_0t value only repetitive sequences should reassociate. As the fragment length increases, the percentage of DNA in duplex increases. This increase is due to single-strand tails of unique DNA contiguous to the reassociated repetitive DNA. An estimate of the length of the unique DNA is obtained from the change in slope in the graph of percent duplex versus fragment length. At the breakpoint in the graph (fragment length) a second repetitive sequence, which has presumably already reassociated, is at the end of the unique tail. The length of the interspersed unique sequences is estimated as the breakpoint, and in maize is 2,100 bp. The slight increase in percent DNA bound above this length is interpreted as evidence for a longer interspersion pattern of repetitive and unique copy sequences.

The average lenghts of repetitive sequences were determined by electron microscopy. At very low C_0t values, C_0t less than 10^{-3}, most of the duplexes were stem palindromes, resulting from inverted repeats, and stem and loop palindromes. The palindromes appeared to be clustered; molecules over 1,000 bp initial fragment length, either had 3 or 4 snapbacks or none. The lengths of the stems ranged from 50–100 bp weight average. The 4 tailed duplexes of the low C_0t DNA had a mass average length of 400 bp. The length of mid-repetitive duplexes was determined with samples reassociated to C_0t 50, and treated with S1 nuclease under conditions in which only long stretches of single strand DNA are digested. The resulting duplexes had a number average of 570 bp and a mass average of 970 bp. Reassociated duplexes were also measured

Table 3. Thermal stability of reassociated duplexes

DNA sample	T_m° C[a]	ΔT_m° C	Mis-match[b]	Hyper-chromicity[c]
Highly repetitive C_0t 0.015 bound	76.0°	11°	12.1%	12.0%
Highly and mid-repetitive, C_0t 300 bound	77.5°	9.5°	10.4%	14.2%
Mid-repetitive only, C_0t 10 bound minus C_0t 0.015 bound	81.5°	5.5°	6.0%	20.8%
Unique C_0t 5000 bound	85.0°	2.0°	2.2%	25.7%
Native DNA	87°	–	–	27.0%

All samples are in 1 × SSC. The fragments are 1,350 bp in length, except native which is 400 bp fragment length

[a] The T_m values were corrected for the expansion of water

[b] Mismatch was calculated by assuming that a 1° C reduction in native T_m equals 1.1% bp mismatch (Wetmur, 1976).

[c] Hyperchromicity was calculated as described in Materials and Methods

directly; the number average of fifty 4 tailed duplexes was 500 bp. This number is biased towards the shorter duplexes since long duplexes would not necessarily have tails. The lengths of the repetitive sequences were also determined from the hyperchromicity of reassociated duplexes. The mass average length of the repetitive sequences was calculated to be 780 bp, as discussed later. Thus, 3 independent lines of evidence, kinetic analysis at different fragment lengths, electron microscopy of Sl treated molecules, and calculations based on hyperchromicity, determine the length of mid-repetitive duplexes, excluding the palindromic sequences, to be 700–1,000 bp.

Thermal Characteristics

The thermal denaturation characteristics of DNA samples, native and reassociated, were determined by melting them as described in Materials and Methods. The T_m of 400 bp native DNA in 1 × SSC was 87° C and 71° C in 0.1 × SSC. The difference, 16° C, is close to the 15.4° C decrease expected from the effect of lower Na^+ concentration on T_m (Marmur and Doty, 1962). Longer fragments of DNA of 1,350 bp were also melted: in 0.1 × SSC the T_m was 75.5° C and in 1 × SSC the T_m was 88.4° C. From the relationship of duplex stability to fragment length: $T_n - T_m = B/L$, where $T_n = T_m$ of long DNA, $T_m = T_m$ of fragments, $B = 650$ at 0.18 M Na^+, L = fragment length, an increase of 1.5° C is expected for long fragments, which is close to what is found (Britten et al., 1974).

The GC content of DNA can be calculated from the T_m or the buoyant density (Marmur and Doty, 1962). Both measurements are affected by unusual

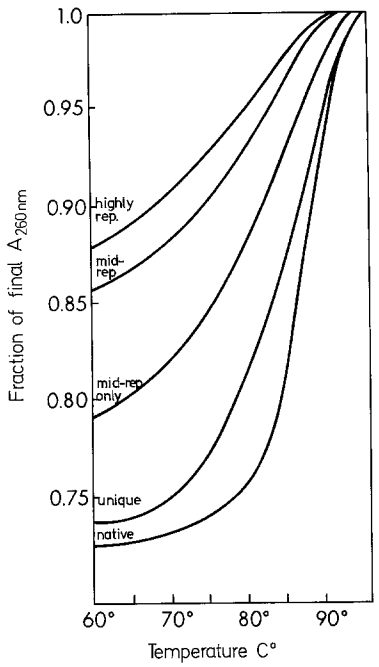

Fig. 4. Thermal denaturation profiles of native and reassociated kinetic fractions of DNA. Kinetic fractions of *Zea mays* DNA at 1,350 bp were prepared and melted in 1 × SSC as described in Materials and Methods. 400 bp fragments of native DNA are included for comparison. Unique DNA is bound at C_0t 5,000; mid-rep. only DNA is bound at C_0t 10 minus C_0t 0.015 bound; mid-rep. DNA is bound at C_0t 300; highly rep. DNA is bound at C_0t 0.015. The T_m and hyperchromicity values for each fraction are listed in Table 3

bases. The buoyant density is decreased by 1 $mg \cdot cm^{-3}$ for every percent methylation (Szybalski and Szybalski, 1971), and the T_m is increased in 0.1 × SSC 1.5° C with each 1% 5 MeC (Dawid et al., 1970). *Zea mays* is reported to have 6.2% methylated cytosines (Grierson, 1977). When this is taken into account the calculated buoyant density becomes 1.707 gcm^{-3} (Ingle et al., 1973). This would imply a GC content of 48.1%. The T_m in 0.1 × SSC was 75.5° C, adjusted for the unusual bases it is 74° C which implies a GC content of 50.9%. The average from both measurements is 49.5%.

The hyperchromicity of native DNA in 1 × SSC, at all lengths above 400 bp, is 27%. The optical dispersion, defined by Mahler and Dutton (1964), is a useful indicator of the amount of heterogeneity in the native and reassociated samples. In 1 × SSC the optical dispersion, $\sigma_{2/3}$, of native corn DNA is 7.6° C. Compared to a simple DNA like E. coli, $\sigma_{2/3}$ of 5.4° (Walbot and Dure, 1976), maize DNA is fairly heterogeneous.

DNA reassociated to various C_0t values was melted in 1 × SSC and 0.1 × SSC under the same conditions as native DNA. The hyperchromicity, T_m and $\sigma_{2/3}$ were calculated (Table 3 and Fig. 4). The divergence in T_m from native DNA, ΔT_m, is a quantitative estimate of the degree of mispairing. Each percent mispairing causes a decrease in T_m of about 1.1° C (Wetmur, 1976). The degree of apparent mismatch is partially determined by the criteria of stringency under which the DNA is reassociated since the stringency determines how well matched sequences must be to form a stable hybrid. Different degrees of mismatch are found when the DNA is reassociated at temperatures varying from T_m − 25° C (Murray and Thompson, 1978).

High C_0t DNA

Reassociated DNA to C_0t values of 3,000–5,000 includes most of the unique DNA ($C_0t_{1/2}=1{,}200$) and all of the repetitive fractions. The T_m of C_0t 5,000 DNA at 1,350 bp is 85° C, 2° C lower than native. The hyperchromicity is close to native values, 25.7° C. The difference from native is due to the steric inhibition to second nucleations on presumably single-stranded tails of duplexes (Britten and Davidson, 1976). The $\sigma_{2/3}$ of high C_0t DNA is 12° C, indicative of more heterogeneity than native DNA, resulting presumably from the mismatched repetitive fractions.

Repetitive DNA

DNA reassociated to a range of C_0t values, 0.01–300, was melted in $1\times$SSC. As the C_0t value increases, the ΔT_m decreases. At the $C_0t_{1/2}$ value of highly repetitive DNA, the ΔT_m is 11° C indicating 12.1% mismatch. At the $C_0t_{1/2}$ of mid-repetitive DNA, the ΔT_m is 9.5° C. If pure mid-repetitive DNA (bound at C_0t 10, unbound at C_0t 0.015) is melted, the ΔT_m is 5.5° C or 6.0% mismatch.

From the hyperchromicity of DNA samples reassociated to mid C_0t it is possible to calculate the proportion of mass involved in duplex and hence the length of the repetitive sequences. Fragments of 1,350 bp reassociated to C_0t 300 (Table 3, second line), have a hyperchromicity of 14.2% and a ΔT_m of 9.5° C; if such molecules are composed solely of duplex with 10.4% mismatching the expected hyperchromicity would be 24% [27–0.104(27)=24%]. Fourteen percent hyperchromicity indicates that only 58% of the mass is in duplex or an average of 780 pb for high and mid repeat DNA. This size estimate is in close agreement with the electron microscopic measurements of DNA duplex length of total repetitive DNA. The optical dispersion $\sigma_{2/3}$ of the repetitive DNA ranges from 6–11° C higher than native DNA indicating that there is a broad range of repetitive classes, some very well matched, others highly divergent.

DNA Sequence Arrangement in the Zea mays Genome

The major portion of the corn genome is organized in two short-period interspersion patterns. The first pattern contains unique copy plus mid-repetitive DNA. The length of the interspersed unique sequences, determined by reassociating DNA to mid-C_0t at different fragment lengths, is 2,100 bp. The amount of interspersed unique DNA can be determined from the same analysis. At the longest fragment length examined, 5,600 bp, 80% of the DNA has reassociated. We can assume that all the highly (17% of the genome) and mid-repetitive DNA (36%) have reassociated. Thus, the 27% remaining must be unique. Since the unique DNA is 35.5% of the total at this fragment length, 76% (27/35.5) of the unique DNA is interspersed with repetitive DNA at lengths less than 5,000 bp. Using an average of 2,100 bp for the interspersed unique sequence

length, there are a possible 8.35×10^5 interspersed, unique sequence elements

$$\frac{(0.76 \times 1.5 \times 10^{12} \text{ dalton complexity})}{660 \text{ daltons per bp} \times 2{,}100 \text{ bp length}}.$$

If the interspersed unique sequences are the structural genes, there is a large excess of unique sequences compared to estimates of active structural genes (Goldberg et al., 1978). The remaining 3.6×10^{11} daltons of unique sequence complexity, 24% of the unique, equivalent to 8% of the total DNA, is farther than 5,000 bp from a repeat. It is either in a long interspersion pattern or in blocks of unique. Estimating the length at 5,000 bp these non-interspersed unique sequences have 1×10^5 sequence elements $\frac{(3.6 \times 10^{11})}{660 \times 5{,}000}$, equivalent to the estimated number of structural genes active during the life of a higher plant. However, the unique sequence elements may not have a coding function. Under the criteria chosen, repetitive sequences with little homology will not form stable duplexes and consequently highly divergent repetitive sequences will appear in the unique sequence class.

There is more mid-repetitive DNA in the genome than required in a short-period interspersion pattern with the unique DNA. Given the average length of mid-repetitive sequences to be 1,000 bp (from electron microscopy studies), then 41% of the mid-repetitive DNA is in the short-period interspersion pattern with unique sequences

$$\frac{(41\% = 100 \times 8.35 \times 10^5 \text{ sequences} \times 660 \text{ daltons per bp} \times 1{,}000 \text{ bp})}{0.36 \times 3.76 \times 10^{12} \text{ daltons}}.$$

Half of the mid-repetitive DNA is interspersed with highly repetitive DNA as shown by analysis of the proportion of each repetitive sequence class at different fragment lengths. This is not a common feature of repetitive DNA in plants, although soybean is the only genome where this question was specifically addressed (Goldberg, 1978). The remaining repetitive DNA is probably in long blocks of non-interspersed repeats as found in tobacco (Zimmerman and Goldberg, 1977), soybean (Goldberg 1978), and cotton (Walbot and Dure, 1976).

A summary of maize genome organization is presented in Table 4. The major feature is two interspersion patterns: a unique repetitive pattern and a high and mid-repetitive interspersion pattern. The unique interspersion pattern includes 33% of the genome and consists of unique sequences averaging 2,100 bp long alternating with repetitive sequences an average of 1,000 bp long. The repetitive interspersion pattern is also approximately one third of the genome and consists of alternating high and mid-repetitive sequences, the length of the mid-repetitive sequences are inferred to be 1,000 bp long. The remaining third of the genome includes the unique sequences that are more than 5,000 bp from a repeat, the palindromic sequences which are 50–100 bp long, and the remaining mid-repetitive DNA that is not in an interspersion pattern.

Table 4. Description of the genome organization

Pattern	% total DNA	Length BP
A. Short period interspersion		
1) Unique	27	2,100
Mid-repetitive	15	1,000
2) Highly repetitive	17	short?
Mid-repetitive	18	short?
B. Long stretches of similar reiteration frequency sequences		
1) Pure single copy	8	> 5,000
2) Pure Mid-repetitive	3	unknown
C. Palindromic DNA	3	100

Table 5. DNA/DNA hybridization of *Zea mays* to related grasses

C_0t	Tracer[a]	ΔT_m° C[b]	% reassociation of tracer to driver	Number of hybridizations
100	*Zea mays* W64A	0.4	100	1
100	*Zea mays* Mo17	0.95 ± 1.06	100	2
100	teosinte Chalco	2.1 ± 0.53	100	2
100	teosinte Guatemala	1.9 ± 1.70	100	2
100	Ladyfinger Popcorn	1.7 ± 0.6	100	2
100	*Tripsacum*	7.0	56	1
100	wheat	> 20	10	1
10^4	*Zea mays* Mo17	0.2	100	2
10^4	teosinte Chalco	2.3 ± 1.0	100	3
10^4	teosinte Guatemala	4.2	100	2
10^4	Ladyfinger Popcorn	3.2 ± 0.45	100	3
10^4	*Tripsacum*	8.2 ± 0.64	51	2
10^4	wheat	> 20	10	1

[a] Driver in all cases is W64A maize DNA. Teosinte (*Z. mexicana*) races are identified by region of origin.

[b] The ΔT_m is the difference in T_m between driver, monitored by $A_{260\ nm}$, and tracer, monitored by cpm.

DNA Hybridization with Related Grasses

Maize DNA was hybridized to DNA of related grasses in order to determine the thermal stability of homologous sequences. Reassociations were carried out to C_0t 100 and C_0t 10,000 to assess the homology of repeated and unique sequences. Tracer DNAs were labeled by nick-translation and hybridized to a large excess of driver DNA. Duplex stability was monitored by thermal denaturation; driver DNA stability was assessed by $A_{260\ nm}$ and tracer stability by the release of radioactivity into single-stranded form. The least amount of mis-

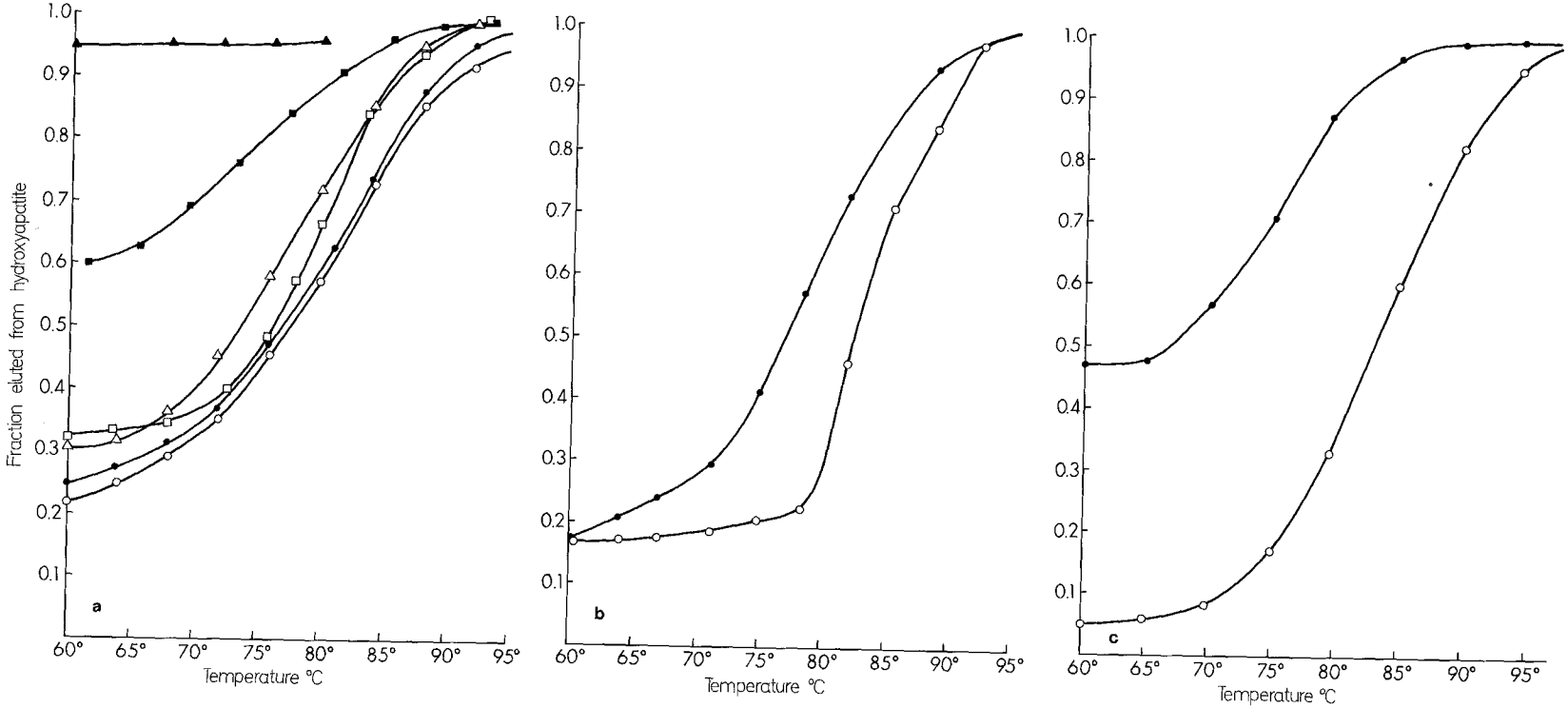

Fig. 5a–c. Hydroxyapatite thermal chromatography of *Zea mays* W64A DNA duplexes formed with tracer DNAs. **a** *Zea mays* DNA driver was reassociated with nick-translated heterologous DNAs to equivalent C_0t 100 in 0.12 M PB at 60° C and chromatographed as described in materials and methods. A melting profile for the homologous duplex was obtained by monitoring the $A_{260\ nm}$ for each reassociation reaction, only one is represented since they were essentially identical at this EC_0t. ○ homologous duplex; ● Mo17 tracer; □ Lady finger pop tracer DNA; △ teosinte Guatemala DNA; ■ *Tripsacum* tracer DNA; ▲ wheat. **b** *Zea mays* DNA driver was reassociated to tracer teosinte Guatemala DNA to equivalent C_0t 10,000. ○ Homologous duplex as monitored by $A_{260\ nm}$; ● heterologous duplex. **c** *Zea mays* driver was reassociated to tracer *Tripsacum* DNA to equivalent C_0t 10,000. ○ homologous duplex; ● heterologous duplex

match detectable by these methods is 0.4% since self-reassociation produces a ΔT_m of 0.4° C.

Maize DNA, W64A (midwestern Dent), was reassociated to the DNA of *Tripsacum* and teosinte *(Zea mexicana)*, purported progenitors; to DNA of popcorn, morphologically similar to an extinct purported progenitor; to Mo17, a midwestern Dent similar to W64A; and to wheat, an evolutionarily distant grass. The data are presented in Table 5. The repetitive sequences of popcorn and teosinte are equally divergent from maize. The $\Delta T_m s$ for teosinte, race Guatemala; teosinte, race Chalco (from Mexico); and popcorn are respectively 1.9° C, 2.1° C, and 1.7° C. The extent of hybridization was approximately the same for tracer and driver duplex, indicating that there are no gross quantitative differences. *Tripsacum*, however, is very divergent in repetitive DNA, ΔT_m equal to 7.0° C. The hybridization is approximately half of the driver DNA hybridization indicating that many repetitive sequences present in maize are either absent or less abundant in *Tripsacum*. The repetitive sequences of Mo17 are equivalent by these methods to those of W64A. The repetitive DNA of wheat shares no homologous sequences at these criteria, some homology may be detectable with less stringent conditions (Murray and Thompson, 1978).

At high C_0t both repetitive and unique sequences have reassociated, thus differences from the previous data reflect divergence in the unique sequences. The $\Delta T_m s$ at C_0t 10,000 for teosinte, race Guatemala; teosinte, race Chalco and popcorn are respectively 4.2° C, 2.3° C, and 3.2° C. *Tripsacum* unique sequences are highly divergent from maize, the ΔT_m is 8.2° C. As with the repetitive sequences, the extent of hybridization between *Tripsacum* and maize is approximately half that of the driver homologous duplex. There is no detectable decrease in T_m of hybrids of W64A and Mo17. The ΔT_m of wheat is greater than 20° C, indicating that, as with the repetitive DNA there are too few sequences in common to create a stable duplex under these conditions. To summarize, popcorn and teosinte DNAs are approximately equally divergent from corn. No doubt each has contributed, either ancestrally or due to introgression, to the present day maize genome. *Tripsacum* is less related, but clearly has many sequences in common. The stringency of these methods is appropriate for examining closely related species. They are not suitable for distantly related species since DNA sequences differing by more than 20% do not anneal with high specificity (Kohne, 1970).

Conclusion

The maize genome was analyzed to define its sequence organization and its relationship to purported progenitors. The genome is organized in two interspersion patterns; unique + repetitive and highly repetitive + mid-repetitive. No direct evidence exists for the function of either pattern, though data from analysis of animal genomes suggest that most mRNA is transcribed from unique copy DNA interspersed with moderately repetitive sequence (Davidson et al., 1975). The repetitive interspersion pattern may have a structural role or an undefined regulatory role; it may also play a role in evolutionary mechanisms or speciation

barriers. The origin of maize and its evolution have been a mystery despite extensive research on the morphological and genetic level. The relationship of maize to its possible progenitors has not been previously examined at the DNA sequence level. DNA hybridization experiments were performed to define the relationship of the maize genome to possible progenitors: teosinte, *Tripsacum*, and popcorn.

Tripsacum is a wild grass, in the same tribe as maize but in a different genus; teosinte, also a wild grass, is a separate species from maize. Ancient wild maize is a hypothetical corn-like pod-popcorn; support for its existence comes from archeological evidence of 7,000 year old corn cobs (Mangelsdorf et al., 1967). *Tripsacum* has been considered a progenitor for taxonomic reasons; genetic and morphological data are in disagreement (deWet et al., 1972). Teosinte has the adaptive, morphological and genetic characters to be the progenitor of maize (Beadle, 1972; Iltis, 1972), but fossil evidence and theories of natural selection do not necessarily support this hypothesis (Mangelsdorf, 1974). Archeological evidence, i.e. ancient cobs, support a theory with primitive maize as the progenitor and teosinte only adding variability (Mangelsdorf, 1974). Popcorn is a prototype on morphological grounds for the ancient pod corn, a now extinct wild maize. The separation between popcorn and other races of maize such as dents and flints is thought to have occurred very early in the domestication process.

It is evident from the hybridization data that maize DNA shares many sequences in common with popcorn and teosinte and that it is least similar to *Tripsacum*. Both Mexican and Guatemalan races of teosinte have contributed to the genome of maize, indicating a diverse and heterogeneous evolution, facts well supported by cytological, morphological and anthropological data (see Galinat, 1976 for discussion and further references). Though *Tripsacum* is at least two-fold more divergent and has significantly fewer sequences in common from modern maize than teosinte, it no doubt has contributed to the present day maize genome.

It is possible to calculate evolutionary distance between pairs of organisms form thermal denaturation data and an assumption of a fixed rate of base change with time (Wilson et al., 1977). If the changes in genomes occur purely randomly at a clock-like rate of 1% mismatch per 5 million years (Wilson et al., 1977), we calculate a divergence of 40 million years since the split between maize and *Tripsacum* and 10–20 million years since the divergence of teosinte and maize. These calculated divergence estimates are unlikely due to the recent advent of the grass family, perhaps 70 million years ago and of individual genera, 20–40 million years ago (Raven and Axelrod, 1974). Presumably the intense impact of human selection on the evolution of maize has contributed to the apparent distance of maize from close relatives.

A calculation of divergence using a molecular clock ignores phenomena such as DNA amplification, chromosome rearrangements, and changes in ploidy or base chromosome number. Using hybridization data alone it is difficult to differentiate between the contribution of a progenitor and introgression, i.e., to say whether teosinte is close to maize because it is a progenitor or because of extensive cross-breeding. It is also difficult to correlate base changes with differences at the chromosome organization level. Recent thinking in evolu-

tion maintains that speciation is not a consequence of base changes in genes but depends on changes in gene regulation and chromosomal organization (King and Wilson, 1975). In rye, another cultivated grass, there is evidence for rapid changes in the type and amount of highly repetitive sequence in chromosomes, and a possible role for the organization of these DNA sequences in evolution has been proposed (Bedbrook et al., 1980). *Secale* species will cross hybridize to differing extents; *S. cereale* and *S. silvestri* differ only in the arrangement and amount of highly repetitive and interrelated sequences in the telomeric portions of the chromosomes. The telomeric regions are important in hybrid viability since they are required for synapsis; thus, changes in these regions result in barriers to gene exchange and may aid in speciation.

The sequence divergence and organization of repetitive DNA provide a clue to the evolution of repetitive sequences. The processes involved are presumably amplification, divergence and transposition. Some families of repetitive sequences are exact repeats indicating recent amplification. The bulk of repetitive sequences are mismatched; divergence arises within a family from random mutations that accumulate over time and from recombination within individual repetitives sequences. For example, in maize the highly repetitive sequence class is twice as divergent as mid-repetitive and is therefore, the oldest class. However, this class is extensively interspersed with mid-repetitive sequences. Consequently, we must explain how the highly repetitive sequences became associated with less ancient mid-repetitive sequences. One possible explanation of this organization is as follows. A progenitor genome contained interspersed, moderately diverged repetitive and unique copy DNA; an amplification event involving most, if not all, of the interspersed DNA resulted in creation of a very highly repeated class contiguous to newly created precisely matched, moderately repetitive DNA (the former unique copy DNA). The results of subsequent divergence at equal rates in both classes of repeated DNA, are viewed today as highly divergent, very repetitive sequences interspersed with moderately repetitive sequences of less divergence. This model predicts that newly amplified repetitive sequences would contain a very highly repeated and moderately repeated portion and that these sequences would be initially organized in long blocks of similar repeated sequences. Specific probes of individual repetitive sequences are needed to be precise in describing the evolution of a repetitive sequence. A cloned repeat would allow one to ask how similar different families of repeats are, at what stage amplification took place, and how the organization of repeats evolved. Our study provides the general background on the typical organization patterns of the various sequences classes of maize for comparison to the more detailed studies in the future.

Acknowledgments. This project was supported in part by a BioMedical Research Support Grant awarded to V.W. S.H. received a fellowship from the Pioneer HiBred International Co. We thank Bob Goldberg for the computer analysis of our data.

References

Beadle, G.W.: The mystery of maize. Field Mus. Nat. Hist. Bull. **43**, 2–11 (1972)

Bedbrock, J.R., Jones, J., O'Dell, M., Thompson, R.D., Flavell, R.B.: A molecular description of telomeric heterochromatin in Secale species. Cell **19**, 545–560 (1980)

Bendich, A.J., Taylor, W.C.: Sequence arrangement in satellite DNA from the muskmelon. Plant Physiol. **59**, 604–609 (1977)

Britten, R.J., Davidson, E.H.: Studies on nucleic acid reassociation kinetics: Empirical equations describing DNA reassociation. Proc. nat. Acad. Sci. (Wash.) **73**, 415–419 (1976)

Britten, R.J., Pavich, M., Smith, J.: A new method for DNA purification. Carnegie Inst. Wash. Year Book **68**, 400–402 (1970)

Britten, R.J., Graham, D.E., Neufeld, B.R.: Analysis of repeating DNA sequences by reassociation. In: Methods in enzymology (L. Grossman and K. Moldave, eds.), vol. **29**, 363–418. New York: Academic Press 1974

Cairns, J.: The chromosome of Escherichia coli. Cold Spr. Harb. Symp. quant. Biol. **28**, 43–46 (1963)

Davidson, E.H., Britten, R.J.: Organization, transcription, and regulation in the animal genome. Quart. Rev. Biol. **48**, 565–613 (1973)

Davidson, E.H., Galau, G.A., Angerer, R.C., Britten, R.J.: Comparative aspects of DNA organization in metazoa. Chromosoma (Berl.) **51**, 253–259 (1975)

Dawid, I.B., Brown, D.D., Reeder, R.H.: Composition and structure of chromosomal and amplified ribosomal DNA's of Xenopus laevis. J. molec. Biol. **51**, 341–360 (1970)

deWet, J.M., Engle, L.M., Grant, C.A., Tanaka, S.T.: Cytology of maize-Tripsacum introgression. Amer. J. Bot. **59**, 1026–1029 (1972)

Felsenfeld, G.: Analysis of temperature-dependent absorption spectra of nucleic acids. In: Procedures in nucleic acid research (G.L. Cantoni and D.R. Davies, eds.), vol. **2**, 233–244. New York: Harper & Row 1971

Fincham, J.R.S., Sastry, G.R.K.: Controlling elements in maize. Ann. Rev. Genet. **8**, 15–50 (1974)

Flavell, R.B., Bennett, M.D., Smith, J.B., Smith, D.B.: Genome size and proportion of repeated nucleotide sequence DNA in plants. Biochem. Genet. **4**, 257–269 (1974)

Galau, G.A., Klein, W.H., Davis, M.M., Wold, B.J., Britten, R.J., Davidson, E.H.: Structural gene sets active in embryos and adult tissues of the sea urchin. Cell **7**, 487–505 (1976)

Galinat, W.C.: The origin of corn. In: Corn and corn improvement (G.F. Sprague, ed.), pp. 1–47. Madison: American Society of Agronomy 1976

Goldberg, R.B.: DNA sequence organization in the soybean genome. Biochem. Genet. **16**, 45–68 (1978)

Goldberg, R.B., Hoschek, G., Kamalay, J.C., Timberlake, W.E.: Sequence complexity of nuclear and polysomal RNA in leaves of the tobacco plant. Cell **14**, 123–131 (1978)

Grierson, D.: The nucleus and the organization and transcription of nuclear DNA. In: The molecular biology of plant cells (H. Smith, ed.), Botanical Monographs vol. 14, 213–255. Oxford: Blackwell Scientific Publications 1977

Iltis, H.H.: The taxonomy of Zea mays (Gramineae). Phytologia **23**, 248–249 (1972)

Ingle, J., Pearson, G.G., Sinclair, J.: Species distribution and properties of nuclear satellite DNA in higher plants. Nature (Lond.) New Biol. **242**, 193–197 (1973)

King, M.-C., Wilson, A.C.: Evolution at two levels in humans and chimpanzees. Science **188**, 107–116 (1975)

Kohne, D.E.: Evolution of higher-organism DNA. Quart. Rev. Biophys. **3**, 327–375 (1970)

Mackey, J.K., Brackman, K.H., Green, M.R., Green, M.: Preparation and characterization of highly radioactive in vitro labeled adenovirus DNA and DNA restriction fragments. Biochemistry **16**, 4478–4483 (1977)

Mahler, H.R., Dutton, G.: Nucleic acid interactions. V. Effects of cyclobuxine. J. molec Biol. **10**, 157–175 (1964)

Mangelsdorf, P.C.: Corn: its origin, evolution and improvement. Cambridge: Harvard University Press 1974

Mangelsdorf, P.C., MacNeish, R.S., Galinat, W.C.: Prehistoric wild and cultivated maize. In: The prehistory of the Tehuacan Valley (D.S. Byers, ed.) vol. **1**, 178–200. Austin: Univ. Texas Press 1967

Manning, J.E., Schmid, C.W., Davidson, N.: Interspersion of repetitive and nonrepetitive DNA sequences in the Drosophila melanogaster genome. Cell **4**, 141–155 (1975)

Marmur, J., Doty, P.: Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. J. molec Biol. **5**, 109–118 (1962)

Murray, M.G., Thompson, W.F.: Sequence repetition and interspersion in pea DNA. Carnegie Inst. Wash. Year Book **77**, 316–323 (1978)

Ogur, M., Rosen, G.: The nucleic acids of plant tissues. I. The extraction and estimation of desoxypentose nucleic acid and pentose nucleic acid. Arch. Biochem. **25**, 262–276 (1950)

Pardue, M.L.: Repeated DNA sequences in the chromosomes of higher organisms. Symposium on chromosome structure: XIII Int. Congr. Genet. **79**, 159–170 (1975)

Pearson, W.R., Davidson, E.H., Britten, R.J.: A program for least squares analysis of reassociation and hybridization data. Nucleic Acids Res. **4**, 1727–1737 (1977)

Raven, P.H., Axelrod, D.I.: Angiosperm biogeography and past continental movements. Ann. Missouri Bot. Garden **61**, 539–673 (1974)

Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchinson, C.A. III, Slocombe, P.M., Smith, M.: Nucleotide sequence of bacteriophage ϕX174 DNA. Nature (Lond.) **265**, 687–695 (1977)

Szybalski, W., Szybalski, E.H.: Equilibrium density gradient centrifugation. In: Procedures in nucleic acid research (G.L. Cantoni and D.R. Davies, eds.), vol. **2**, 311–354. New York: Harper & Row 1971

Walbot, V., Dure, L.S. III: Developmental biochemistry of cotton seed embryogenesis and germination. VIII. Characterization of the cotton genome. J. molec Biol. **101**, 503–536 (1976)

Walbot, V., Goldberg, R.: Plant genome organization and its relationship to classical plant genetics. In: Nucleic acids in plants (T.C. Hall and J. Davies, eds.), vol. **1**, p. 3–40. Boca Raton, Florida: CRC Press 1979

Wetmur, J.G.: Hybridization and renaturation kinetics of nucleic acids. Ann. Rev. Biophys. Bioeng. **5**, 337–361 (1976)

Wilson, A.C., Carlson, S.S., White, T.J.: Biochemical evolution. Ann. Rev. Biochem. **46**, 573–639 (1977)

Zimmerman, J.L., Goldberg, R.B.: DNA sequence organization in the genome of Nicotiana tabacum. Chromosoma (Berl.) **59**, 227–252 (1977)

Received December 7, 1979 / Accepted March 7, 1980 by M.L. Pardue
Ready for press March 30, 1980