Heterogeneity of Pumpkin Ribosomal DNA

Received for publication October 13, 1982 and in revised form January 14, 1983

ALBERT SIEGEL AND KATHRYN KOLACZ
Department of Biological Sciences, Wayne State University, Detroit, Michigan 48202

ABSTRACT

The ribosomal DNA (rDNA) of Cucurbita pepo L. has been found to consist of tandemly arrayed repeat units, most of which are 10 kilobases in length. Thirty-six repeat units, cloned into the HindIII site of pACYC 177, fall into seven classes which differ from each other in length and/or nucleotide sequence. Most of the heterogeneity occurs in noncoding portions of the repeat unit although there is some nucleotide sequence variation in the coding portion as well. Heterogeneity of base modification was observed in genomic rDNA of which two examples are: (a) all of the repeat units have three BamHI sites, one of which is unavailable for restriction in about half of the units and (b) all of the CCGG sites except one are methylated at the internal cytidine in many of the units; a second site is unmethylated in some of the units and in a very few units a third site remains unmethylated.

The genes for ribosomal RNA generally occur in multiple copies; the genomic number varies from as little as one in the micronucleus of Tetrahymena to several thousand in many plant species and the amphibian order Urodela (19). In higher eucaryotes, the ribosomal genes are organized into tandemly repeating units, each unit consisting of a transcription unit and a nontranscribed spacer. A precursor RNA is transcribed from the transcription unit and then processed into the functional 18S, 5.8S, and 25 to 28S ribosomal RNAs and sometimes into an additional 2S component (19). Although the repeat units of a species all are presumed to contain the same information, they frequently are not exactly alike. Heterogeneity has been observed in length, nucleotide sequence, and base modification. Two types of length heterogeneity have been described. One type that occurs in Droso-

The present study was undertaken to reassess rDNA heterogeneity in a plant species with a relatively large number of repeat units. The organism chosen for study is Cucurbita pepo L. whose rDNA is contained in an easily fractionated dense isopycnic satellite component (21). The genomic repeat unit number has been estimated to be 3400 (30) with some variation possibly resulting from differences between cultivars (22). We report here that the basic rDNA repeat length of this species is 10 kb and that there is considerable heterogeneity in length, nucleotide sequence, and base modification.

MATERIALS AND METHODS

Preparation of rDNA-Enriched Pumpkin Nuclear DNA. Nuclear DNA was isolated from young pumpkin (Cucurbita pepo L. cv Small Sugar) leaves (27) and fractionated by isopycnic centrifugation in CsCl. Preparations enriched in rDNA were obtained by recovering DNA from those fractions containing the dense satellite component (21).

Cloning the rDNA Repeat Unit. Most of the rDNA repeat units proved to have a single HindIII site with a minority having a second such site (Fig. 3). Thus, 400 ng of the rDNA-enriched fraction was cleaved with HindIII and incubated with 900 ng of HindIII-cleaved, bacterial alkaline phosphatase-treated bacterial plasmid pACYC 177 (4) under ligating conditions and the reaction product was used to transform competent cells of Escherichia coli strain HB 101. Ampicillin-resistant, kanamycin-sensitive colonies were tested for the presence of rDNA by colony hybridization (10) with 32P-labeled, random primed cDNA to unfraccionated turnip root RNA as a probe. Clones giving a positive signal in the colony hybridization test were subjected to a rapid plasmid analysis procedure (16) in order to determine insert size. Enough of each plasmid selected for more intensive analysis was prepared by scaling up the rapid plasmid analysis procedure.
Preparation of Turnip Root RNA. Ten gm diced turnip (*Brassica rapa* L.) were ground in liquid N$_2$ to a fine powder and extracted with a mixture of 10 ml extraction buffer (10 mM Tris, 10 mM NaCl, 1 mM EDTA, 5 mM 2',3'-AMP, 1% SDS, pH 8.2) and 10 ml phenol. The aqueous phase was reextracted with phenol and several times with chloroform-octanol (24:1). Nucleic acids were precipitated with ethanol and resuspended, and the high mol wt RNAs were precipitated with 2 M LiCl. When required, preparations of 25S and 18S RNAs were made by loading 100 µg of high mol wt turnip root RNA into a 1.5 x 80 mm well in a horizontal electrophoresis unit which contained 1% agarose in TAE buffer (40 mM Tris, 5 mM acetic acid, 1 mM EDTA, pH 8.2), 15 mM iodoacetic acid, and 0.1 µg/ml ethidium bromide. Electrophoresis was conducted until the bands were well separated (only two sharply defined bands representing the 25S and 18S RNAs were apparent). The 25S band was cut out and transferred to a trough in an identical gel (but lacking ethidium bromide). The 18S and 25S RNAs were then electroeluted into buffer (TAE with 5 µg/ml proteinase K) containing troughs cut just in front of the bands. The RNA solutions collected from the troughs were extracted with TAE-saturated phenol, two times with chloroform-octanol (24:1), precipitated with 1.5 volumes 2-propanol, and resuspended in H$_2$O. The quality of the RNA was monitored by electrophoresis in either urea-iodoacetate (18) or methylmercuric hydroxide containing agarose (2).

Restriction Endonuclease Analysis. Restriction site location and physical mapping were performed by described procedures (6).

The coding regions for 25S and 18S RNAs were located by digesting a cloned repeat unit with a number of restriction endonucleases, electrophoresing the fragments through 0.8% agarose in a horizontal gel unit using TAE buffer, transferring the separated fragments to nitrocellulose, and hybridizing these to random primed cDNA probes prepared to gel-purified turnip 18S and 25S RNAs.

RESULTS

Restriction Endonuclease Digestion of the Satellite Component. DNA extracted from nuclei of *C. pepo* leaves displays a prominent dense rDNA containing satellite component (ρ = 1.706) upon isopycnic centrifugation (21). Material consisting largely of the satellite component was prepared, digested with restriction endonucleases, and subjected to agarose gel electrophoresis. Figure 1 shows the results of digestion with *EcoRI*, *BamHI*, and a combination of the two. Prominent ethidium bromide-stained bands are visible, indicating considerable enrichment of a repetitive sequence, probably rDNA. That the repetitive sequence is actually rDNA is deduced from the observation that 32P-labeled probes prepared to 25S and 18S turnip root rRNA hybridize to many of the restriction fragments upon Southern blot analysis (Fig. 1).

It is possible to construct an uncertain physical map of the rDNA repeat unit as shown in Figure 2, but the indications are that not all of the units are identical. For instance, in order to construct the physical map, it is necessary to assume that about half of the repeat units have two *BamHI* sites and that the other half have three and also that some of the repeat units have a shortened version of the larger of the two *EcoRI* fragments.

Digestion with a number of other restriction endonucleases also indicates heterogeneity of the rDNA repeat units, with some of the endonucleases yielding rather complex patterns (Fig. 3). A restriction pattern of interest is that for *HindIII* (Fig. 3). A major band is seen which has the size of the complete repeat unit as deduced from the map shown in Figure 2. Thus, most of the repeat units contain a single recognition site for *HindIII*. Two minor bands are seen whose sizes add up to that of the major band, indicating that a fraction of the repeat units have a second *HindIII* site.

We conclude from the restriction endonuclease analysis of enriched genomic DNA that the rDNA repeat units of *C. pepo* are heterogeneous.

Cloning of rDNA Repeat Units into the Plasmid pACYC 177. To assess the extent and nature of pumpkin rDNA repeat unit heterogeneity, *HindIII*-cleaved fragments of enriched rDNA were inserted into the *HindIII* site of the bacterial plasmid pACYC 177. Of 916 transformants tested, 570 (62%) proved to be kanamycin sensitive and, thus, probably contain a chimeric plasmid. When these were tested for rDNA content by colony hybridization, 76 (13%) reacted positively to a probe prepared to high mol wt turnip root RNA. Twenty-five of these gave a weaker signal than the others and upon quick screening analysis proved to contain a fragment the same size as the smaller of the two *HindIII* fragments yielded by those repeat units presumed to contain two *HindIII* restriction sites (Fig. 3). Fifty-one clones gave strong signals upon colony hybridization and 33 of these were taken for further analysis.

Analysis of the Cloned rDNA Repeat Units. The 33 clones were subjected to restriction endonuclease analysis and physical maps...
were constructed. Three of these proved to contain chimeric plasmids with two repeat units so that the total number of repeat units analyzed was 36. The top line of Figure 4 shows the map for the cloned repeat unit 12A and the restriction site positions for each of the enzymes tested. The locations of the 18S and 25S coding sequences are indicated as deduced from Southern blot analysis using 32P-labeled cDNA probes prepared to turnip root 18S and 25S rRNAs and plasmid V-53 (group IIa, see below) digested with HindIII and either BamHI, Smal, Xhol, SacI, BgII, or HaeII.

Nine other cloned repeat units yielded the same pattern and physical map as did 12A and these were collectively given the class designation 1a. Six other classes were identified among the other cloned repeat units. These were given designations Ib, Ic, Id, IIa, IIb, and III, and were identified by analysis of 3, 1, 6, 10, 1, and 5 cloned repeat units, respectively. The details of how these classes differ from 1a is shown in Figure 4. The following points are noted:

The Classes Differ from Each Other in Nucleotide Sequence. (a) Ib and Ic have 28 restriction sites in common with 1a but differ from it by a single restriction site in a noncoding region at position 7.7 (except for the extra sites in Ic’s large insertion, see below). (b) Id is exactly like 1a except that it is only ¾ of a repeat unit. It is the same size as the larger of the two HindIII fragments generated from the genomic repeat unit that contains two HindIII sites rather than one (see Fig. 3). It is probable that the additional HindIII site results from a nucleotide sequence difference but it is also possible that Id has a small insertion that contains the extra HindIII site. (c) Class IIa consists of 10 cloned repeat units which have 26 restriction sites in common with those in 1a, but differs in having six extra sites lacking in Group 1a and in being deficient in two group 1a sites. The difference in nucleotide sequence between the classes does not appear to be uniformly distributed but to be concentrated in a specific noncoding region so that the region between map units 4 and 5 may be substantially different in groups I and II. It has not been determined whether the two changes observed at 1.6 to 2.0 map units fall in a coding unit or in internal transcribed spacer. (d) IIb is like IIa in nucleotide sequence except for an extra KpnI site present in its insertion (see below). (e) III consists of six cloned repeat units. It has 25 restriction sites in common with 1a but differs in having two additional sites and lacking one site (in addition to the two restriction sites lost because of its deletion, see below). At least two of the restriction site differences are in the 25S coding region.

The Classes Differ from Each Other in Length. The 20 10-kb repeat units of classes Ia and Ila are taken to have a standard size. Those which are longer or shorter than this are said to have insertions or deletions although some of the insertions may represent extra copies of an indigenous repetitious sequence. The approximate positions of the insertions and deletions as well as their size are indicated.

(a) Class Ib contains a 0.2-kb insertion between the 2 KpnI sites at 5.6 and 6.45. (b) Class Ic also contains what may be the same 0.2-kb insertion present in Ib and, in addition, has a 1.5-kb sequence inserted close by, between 6.45 and 7.7. The 1.5-kb
insertion has a number of restriction sites as indicated. (c) Class IIb has a 0.8-kb insertion close to or at the same place as the small class Ib, Ic insertion. It places a KpnI site just midway between the two bracketing KpnI sites at 5.65 and 6.45. (d) Class III is not as long as the other repeat unit classes because of a shorter noncoding region. It contains 1.3-kb deletion at the indicated position between 4.4 and 5.7.

**Heterogeneity in Base Modification of Genomic rDNA.** Another type of repeat unit heterogeneity, that resulting from variability in base modification, became evident upon comparison of genomic and cloned rDNA restriction patterns. An example of such variability is seen in the nature of the genomic BamHI sites. It was concluded from the genomic restriction pattern that there are two classes of repeat units, one with two BamHI sites and the other having the same two sites plus a third (Figs. 1 and 2). Examination of the cloned repeat units revealed that all 36 had the three BamHI sites (Fig. 4), and thus, it is concluded that the failure of cleavage at the specific BamHI site in about half the genomic repeat units results from a variable base modification at that site.

Another example of base modification heterogeneity is seen in the restriction patterns of genomic rDNA yielded by the endonucleases HpaII and MspI (Fig. 3). Both of these enzymes cleave at the restriction site 5'-CCGG-3' but they differ in that HpaII will not cleave at this site if the internal cytidine is methylated whereas MspI will (32). MspI digests genomic rDNA into small fragments (not clearly defined under the electrophoresis conditions employed) as might be expected for an enzyme with a four-nucleotide restriction site. HpaII, on the other hand, digests genomic rDNA primarily into 10-kb fragments, indicating the presence of but a single, specific, unmethylated HpaII site in most of the repeat units. Some of the units have a second such site as indicated by the presence of two fainter bands in the HpaII restriction pattern with sizes that add up to that of the repeat unit. The presence of two additional still fainter bands indicates a third very small class that contains a third unmethylated site.

The data indicate that in many rDNA repeat units every cytidine residue located 5' to a guanosine residue is methylated with one exception. There are some repeat units with a second specific exception and a very small class with still a third. The presence of discrete BamHI- and HpaII-generated fragments indicates that modification or escape from modification is not a random affair but occurs only at certain sites.

Is rDNA Repeat Unit Heterogeneity Unique to C. pepo? The rationale for undertaking analysis of the C. pepo rDNA repeat unit was that this species presents an example of an organism with thousands of such repeat units per haploid genome where, as a consequence, the rectification mechanism might not be able to keep up with diversity-generating phenomena and, indeed, it has been demonstrated that in this case there is considerable repeat unit heterogeneity. Are the rDNA repeat units of C. pepo a special case or is comparable heterogeneity a general feature of organisms with a high rDNA repeat unit number? We have examined the enriched genomic rDNA of tomato (Lycoperisum esculentum Mill cv Rutgers). As with pumpkin, a substantial enrichment of rDNA is possible in tomato by collecting a distinct dense satellite component (5). Such enriched tomato rDNA has been digested with a number of restriction enzymes and a tentative restriction map has been constructed. The data indicate a majority repeat unit length of 8.8 kb (not shown). As with pumpkin, difficulty is encountered in the attempt to construct a restriction map because of apparent rDNA repeat unit heterogeneity. Varying degrees of interpretive difficulty are presented dependent on the restriction enzyme used. Thus, like pumpkin, tomato rDNA contains heterogeneous repeat units. To what degree the repeat unit variability
results from heterogeneity in nucleotide sequence, length, and/or base modification awaits analysis of cloned tomato rDNA repeat units and comparison of these with genomic rDNA.

**DISCUSSION**

Restriction pattern analysis of genomic and cloned pumpkin rDNA reveals the existence of considerable heterogeneity. The 36 analyzed cloned repeat units fall into seven classes which differ from each other in length and/or nucleotide sequence. The seven classes appear to fall into three natural groups, the classes within a group being more like each other in nucleotide sequence than the classes between groups. Thus, not counting sites present in insertions, group I differs from group II in seven of 34 identified restriction sites and from group III in three of 31 sites. Group II differs from group III in eight of 36 sites. Within group I, classes Ib and Ic differ from Ia in having both an extra HaeII site and a 200-bp insertion. Ic has, in addition, a 1.5-kb insertion. Only ⅓ of the class Ia repeat unit has been compared with other repeat units and except for an extra HindIII site no difference from class Ia has been detected. Within group II, class IIb differs from IIa in having an 800-bp insertion. Classes Ia and IIa, which together account for 20 of the 36 cloned repeat units are taken to have a standard length of 10 kb. In comparison, group III, which only has one class, contains a 1.3-kb deletion in addition to differing from Groups I and II in several restriction sites. If the cloned repeat units analyzed are a fair representation of the cellular population, then it would appear that there are three major types of rDNA repeat represented by groups I, II, and III which are present in the approximate proportion of 19:11:6. In addition, there are a number of variants within each group. It remains to be determined whether repeat units of the different groups are organized into different tandem arrays.

The major groups and classes differ from each other primarily in the spacer region between the 18S and 25S coding regions. Most sequence differences between groups I and II map in the region 3.6 to 5.5, with group III lacking a portion of this region altogether. The identified insertions are further along the map, two mapping between 5.65 and 6.4 and the other some place between 6.45 and 7.7. Although classes differ from each other mostly in the nontranscribed spacer, there is variability in coding regions. Group III differs from groups I and II by two restriction sites in the 25S region. It is uncertain whether the two sites between 0.2 and 0.4, by which group II also differs from groups I and III, lies in a coding region or in internal transcribed spacer. Our results are in agreement with others (19) in indicating that heterogeneity of rDNA repeat units, where it exists, derives primarily from variability in the noncoding region but also, with lower frequency, in coding regions (14).

We have defined a ‘class’ of repeat units as having a unique nucleotide sequence. There are two reasons for believing that not all classes have been detected in our analysis. The first is that only a small proportion of the nucleotide sequence has been examined and, as more of it is determined, additional small differences may be found to distinguish between units that are now thought to be the same. For example, 36 restriction sites representing only 196 of the 10,000-base sequence have been examined. Despite this small window, five of the repeat units have been found to have a HindIII site which the other 31 lack and 15 repeat units differ from the other 21 at a HaeII site. Thus, for instance, it is not unreasonable to suppose that at least some of the 10 class Ia members may be found to differ from the rest as more of the nucleotide sequence is determined. The second reason for surmising that there are additional classes derives from the fact that two of the classes are represented by one cloned repeat unit each and it’s likely that other low frequency classes occur that were undetected in the present experiment.

Identified restriction sites are not evenly distributed along the repeat unit map (Fig. 4). There appear to be relatively more sites in the coding regions than in the noncoding region. The reason for this is not known but it can be speculated that, as in the case of other species (23, 26), the nontranscribed spacer may be partially composed of repetitive sequences which happen to lack sites for any of the enzymes so far tested.

A conclusion drawn from the present analysis is that, as in the case of organisms with a hundred to several hundred rDNA repeat units, a rectification mechanism is present in pumpkin, an organism with several thousand repeats, that operates efficiently to keep coding regions almost homogeneous. It is less efficient in homogenizing nontranscribed spacer. Nevertheless, intraspecific spacers have not diverged to the extent that they still do not all bear a striking resemblance to one another.

A high proportion of plant DNA cytosine residues have been found to be methylated, particularly in the configurations CpG and CpNpGp (15). It is not surprising, therefore, to find that the pumpkin genomic repeat unit is fairly resistant to digestion with HpaII. What is notable is that a specific HpaII site remains unmethylated in many repeat units and that in a few repeat units second and even third specific sites remain unmethylated. The preliminary observation has been made that different preparations of enriched genomic rDNA do not contain the same proportion of repeat units with a single unmethylated HpaII site. Although most of the repeat units in the preparation shown in Figure 4 are cleaved once, in other preparations only a minority are cleaved once, the majority being completely resistant to HpaII. Whether or not developmental stage or plant growth conditions affect the degree of specific methylation and thus, perhaps, the number of transcriptionally active ribosomal RNA genes remains to be determined. An observation similar to the one reported here has been made for X. laevis somatic rDNA; about half the repeat units contain a single unmethylated HhaI site. Evidence that methylation may serve as a control mechanism for rDNA expression comes from the observation that transcriptionally inactive rat rDNA is preferentially methylated (29).

Another type of base modification is reported here; one of three BamHI sites is resistant to digestion in about half of the genomic repeat units. Similar observations have been reported for the rDNA of other species. Among these are restriction site heterogeneity for a BamHI site in barley (11), a BglII site in soybean (31), and a HincII site in humans (9). The significance, if any, of this type of heterogeneity remains to be determined. Perhaps, the variably methylated sites are distinguished from the invariable sites by neighboring bases. For instance, it may be that the 3’-neighboring base of the variably methylated BamHI site, GGATCC, is guanosine, thus generating a CpNpG sequence, subject to methylation inheritance as discussed (15). The same consideration can be applied to the variable BglII and HincII sites but the question remains as to why the site in question is methylated in only some of the repeat units rather than being either unmethylated or methylated in all of them.

**Acknowledgment**—We thank Robert Leclerc for help with graphics.

**LITERATURE CITED**

5. Cherry M-D 1975 Ribosomal DNA in a nuclear satellite of tomato. Genetics 81: 469–483
6. Dann KJ 1980 Determination of fragment order through partial digest and
29. TANTRAVAHU U, R GUNTAKA, B ERLANGER, O MILLER 1981 Amplified ribosomal RNA genes in a rat hepatoma cell line are enriched in 5-methylcytosine. Proc Natl Acad Sci 78: 489–493
31. VERLAAN-DREINER A, JF GUELLA, C KEYS, D HOUSEMAN, D SULLIVAN, N BSSON, D VERMA 1979 The organization of a nuclear DNA sequence from a higher plant: molecular cloning and characterization of soybean ribosomal DNA. Gene 7: 317–334
32. WAALEN W, R FLAVELL 1978 MapL, an isochizomer of HpaII which cleaves both unmethylated and methylated HpaII sites. Nucleic Acids Res 5: 3231–3236