

Heterogeneity and organization of the ribosomal RNA genes of *Cucurbita maxima*

Robert J. Kelly,¹ Robert D. Johnson and Albert Siegel*

Department of Biological Sciences, Wayne State University, Detroit, MI 48202, USA (*author for correspondence); ¹ Present address: Howard Hughes Medical Institute, University of Michigan Medical Center, Ann Arbor, MI 48109, USA

Received 21 October 1989; accepted in revised form 8 February 1990

Key words: rDNA, intergenic spacer, IGS, tandem arrays, concerted evolution, *Cucurbita maxima*

Abstract

Thirty-six clones were recovered from *Cucurbita maxima* genomic DNA which had been enriched for rDNA and cleaved at the unique repeat unit *Hind* III site. Twenty-nine of these, which contain complete rDNA units, were compared to a standard whose intergenic spacer (IGS) nucleotide sequence has been determined. Twenty-one are identical in length and restriction site pattern. Eight which differ from the standard in length do so because of addition or deletion of varying numbers of IGS subrepetitive units of two different classes, with four of the length variants being different in both of these classes. Seven clones were isolated which contain incomplete repeat units, six of which are composites of rDNA and non-rDNA material. They have been cleaved at the unique rDNA *Hind* III site at one end and at a non-rDNA *Hind* III site at the other. We consider it most likely that these are derived from the termini of repeat unit tandem arrays, although other explanations are possible. Twelve individual plants of two different cultivars were examined for heterogeneity of IGS length distribution. They all appear to be identical in this regard.

Introduction

Organisms have multiple copies of the genes for ribosomal RNA (rRNA), presumably to ensure that adequate substrate is available for transcription of what amounts to more than 50% of the total cellular RNA [24]. These genes occur as tandem arrays in higher eucaryotes with transcription units for 18S, 5.8S and 25S rRNAs separated from each other by intergenic spacers (IGS) [14, 24]. The rRNA coding sequences are conservative in evolution, indicating a strong positive selection pressure. In contrast, the IGS evolves rapidly despite containing the rRNA

transcriptional control elements. Although there is usually a degree of IGS sequence variability within species, differences between species are much greater. The maintenance of such relative intragenic sequence homogeneity of repetitive sequence has been called concerted evolution, a phenomenon that is thought to be driven by unequal crossing over and/or gene (or sequence) conversion [4, 5]. Most animal species have tens to hundreds of rDNA repeat units per haploid genome, whereas most plant species and the urodele amphibians have an order of magnitude more [14] and in these the problem of IGS sequence homogenization is magnified. Although IGSs of

disparate species have little or no sequence homology, they do have a feature in common; they all contain one to several families of repetitive DNA [7, 19]. Unequal crossing over within these repetitive families probably generates the IGS length heterogeneity that has been observed in a number of species [1, 6, 8, 19, 21, 22, 27] and, at the same time, is at least partially responsible for maintaining the considerable intraspecific sequence homogeneity [2, 3, 26]. We present here an analysis of IGS length variability in *Cucurbita maxima*, a plant species that has been estimated to have ca. 5000 rDNA repeat units, constituting 7% of the haploid genome [9, 10]. A comparison has been made of restriction maps of cloned IGS length variants with that of a standard whose sequence has been determined. The data serve to confirm and extend observations made for other species that rDNA length variants result from changes in the number of units in any of several internal IGS repetitive families. Data are also presented that can be interpreted to indicate that there may be a number of rDNA tandem arrays per haploid genome, each of limited length. This information provides support for the hypotheses that unequal crossing over is operative in the rDNA tandem arrays and may play an important role in maintaining intraspecific IGS sequence homogeneity.

Materials and methods

The methods for preparation of *C. maxima* rDNA clones, the restriction mapping of the clones and the determination of the nucleotide sequence of the IGS of one of these clones have been described [12]. The *Escherichia coli* strains JM109 and DH5 α , both *recA1*, were used as plasmid hosts to lessen the chance that IGS variants would be generated during cloning manipulations. DNA was isolated from ca. 3 g of leaves from individual plants as described [18]. Southern blots were performed as described [25]. DNA fragments were subcloned into the *Hind* III site of pBS(+)(Stratagene) for nucleotide sequence determination by the dideoxy chain termination

method using double-stranded DNA as template [28].

Results and discussion

Isolation and restriction map comparison of full-length rDNA clones

Information concerning the structure of *C. maxima* rDNA was obtained by digesting samples enriched for rDNA [15] with several restriction enzymes. Digests obtained with either *Eco* RV or *Hind* III resulted in a single ca. 10.8 kb fragment, indicating that the rDNA repeat units contained a single site for these enzymes. Thirty-six clones containing rDNA coding sequence, as indicated by colony hybridization [22], were obtained by ligating a *Hind* III digest of rDNA enriched genomic DNA into *Hind* III cleaved, dephosphorylated pUC8. These clones were subjected to restriction analysis for construction of physical maps. They fall into two general size classes. Twenty-nine are approximately the same size as the insert in pMAX23 (taken as a reference standard), contain a complete rDNA repeat unit and are considered 'full length'. The remainder are considerably shorter, do not contain a full-length rDNA repeat unit and are called 'short'. Twenty-one of the 29 full-length clones fall into the most common class (class I) and have a restriction pattern the same as pMAX23 (Fig. 1). This clone was selected as the standard and subjected to more detailed analysis, including determination of its IGS nucleotide sequence [12]. In addition to the enzyme sites shown on the map, the following enzymes were tested and found not to cleave pMAX23; *Apa* I, *Bcl* I, *Bst* EII, *Pst* I, *Pvu* II and *Sca* I. The major *C. maxima* IGS, as represented by pMAX23, contains a highly ordered structure consisting of 5 sets of repeat units (A–E, E is embedded in the D repeat) and 3 unique regions (Fig. 1). The other full-length clones fall into 5 distinct classes (classes II–VI) which differ from the standard and each other, primarily in IGS length. Classes II and III contain three and two of the clones, respectively,

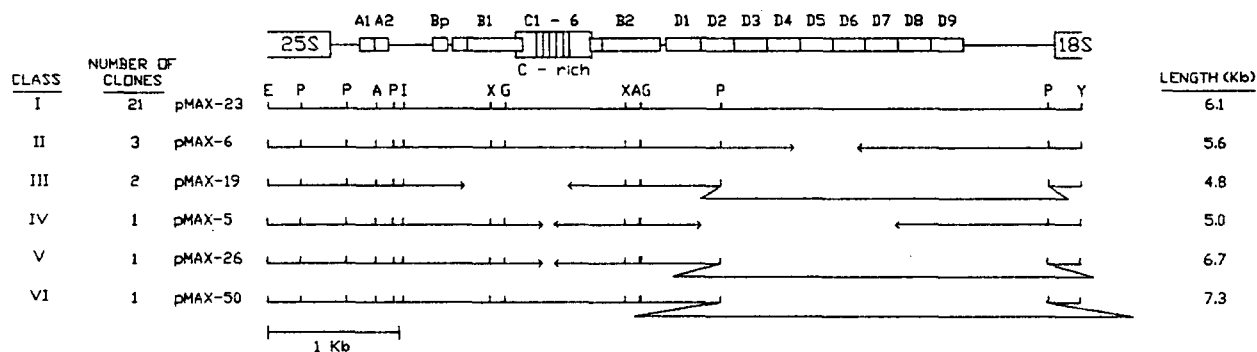


Fig. 1. Restriction map alignment of IGSs of full-length *C. maxima* rDNA clones. The top figure shows the organization of repetitive sequences in the *Eco* RI/*Xba* I fragment of a full-length rDNA repeat unit (pMAX23). The fragment contains the IGS bordered by small portions of the 25S and 18S rRNA coding sequences. The blocks labelled A–D show the position, size and number of different repetitive units in the IGS [15]. The line below the figure is a restriction map of the pMAX23 fragment, taken as a standard. The other lines are restriction maps of other cloned full-length repeat unit fragments. Gaps indicate a shorter distance between restriction sites than that in the standard and below the line inserts indicate a longer distance. The number of recovered clones with *Eco* RI/*Xba* I fragments identical to the ones shown is indicated on the left of each restriction map.

Restriction sites: A, *Hae* II; E, *Eco* RI; G, *Bgl* I; I, *Hpa* I; P, *Sph* I; X, *Xho* I; Y, *Xba* I.

and classes IV–VI are each comprised of single clone. The nature of IGS length heterogeneity was assessed by comparing the variant restriction maps with that of the standard (Fig. 1). The length of diagnostic restriction fragments permitted the correct positioning of expanded and reduced internal regions in the several IGSs. The results show that the length variants can be accounted for by differences in the number of units in one or more of the IGS repetitive families. Classes II–VI differ from class I (the standard) as follows.

– Class II is shorter than the standard by ca. 500 bp in the region of the 259 nt, D repetitive family and, thus, appears to lack 2 D repeats units.

– Class III apparently has 10 instead of 9 D repeats and is probably missing B1 or part of the C region.

– Class IV appears to lack 4 D repeats including D2 (the *Sph* I site characteristic of D2 is not present) and, in addition, 2 C repeats are probably missing because the sites which flank the C-rich region (*Bgl* I and *Xho* I) are detectably closer together by an estimated 80–100 nt than they are in Classes I, II and VI.

– Class V has an estimated 3 additional D re-

peats and, like Class IV, appears to lack 2 C repeats.

– Class VI appears to have 5 additional D repeats.

Three of the five variant IGS classes differ from the standard in more than one region (Fig. 1; pMAX19, pMAX5 and pMAX26). Two of the several possible explanations to account for this phenomenon are: 1) multiple close unequal crossovers in the same cell (negative interference) when the subrepetitive sequences are misaligned, and 2) a single unequal crossover when there is misalignment resulting from the pairing of a standard repeat unit with a length variant.

The observed variant IGS classes differ from the standard primarily in number of D and C units (Fig. 1) and there is no reason to believe that differences in numbers of the other subrepetitive series would not also be observed if additional clones were examined. The nature of the variant clones indicates that they were generated by unequal crossing when chromosomal misalignment had occurred in one or another of the IGS subrepetitive series. Unequal crossing over can occur when there is either ‘macro’-misalignment of tandem arrays by a complete rDNA unit (or multiple units) or ‘micro’-misalignment of subrepetitive se-

quences. There is an increase or decrease in number of rDNA units in an array when macro-misalignment occurs, with a tendency towards fixation of a single length variant [23]. Micro misalignment, on the other hand, leads to change in the number of units of a subrepetitive series and, as a consequence, heterogeneity of IGS length. Thus, unequal crossing over drives IGS length heterogeneity in opposite directions as a consequence of the two types of misalignment while, at the same time, promoting maintenance of relative intraspecific IGS sequence homogeneity. It may be that the several repetitive series in the IGS may foster concerted sequence evolution with, however, length heterogeneity as a consequence.

Characterization of short clones

Seven of the isolated rDNA clones were considerably shorter than the others and ranged in size from 2.9 to 8.2 kb. In order to help deduce their nature and possible origin, their restriction maps and terminal sequences were determined and these were matched to that of the pMAX23 standard. The results (Fig. 2) show that they fall into two groups. One consists of the single clone, pMAX34, whose insert ends are the same as those in pMAX23 and which has a large deletion encompassing the entire IGS and parts of the coding regions. The other six have inserts with only one terminus in common with the standard and have the same restriction sites as pMAX23 for different lengths before they diverge in sequence at an undetermined distance from their other end. The inserts in two of the clones (pMAX36 and pMAX37) have a terminus identi-

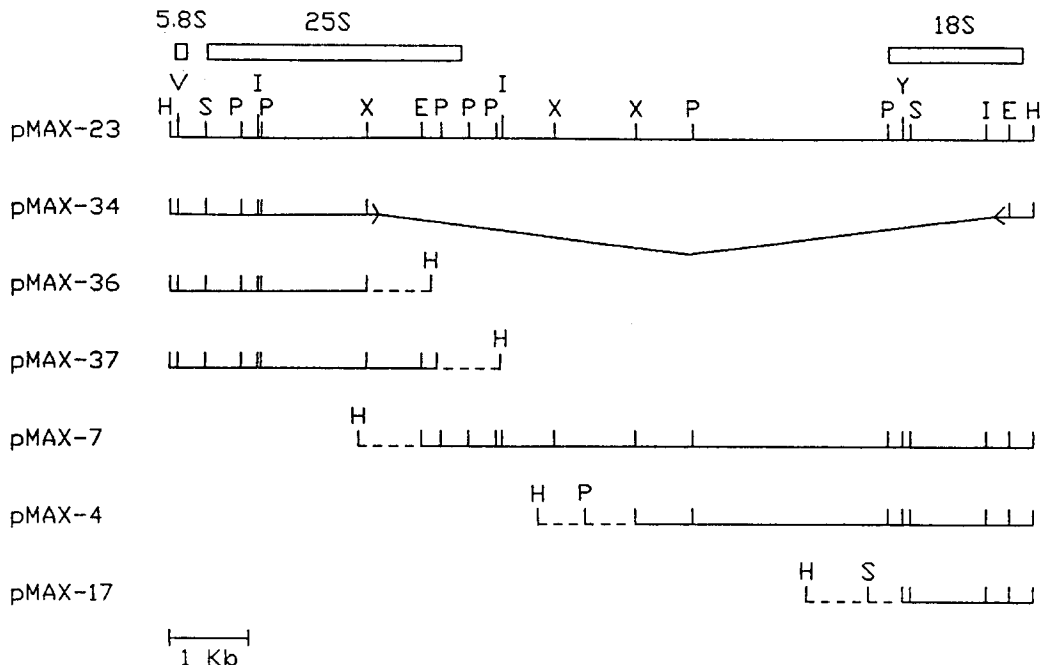


Fig. 2. Restriction map alignment of short *C. maxima* rDNA clones. The top line is a restriction map of the rDNA repeat unit in pMAX23. The blocks above it show the position of the coding sequences for the 5.8S, 25S and 18S rRNAs. The other lines represent restriction maps for the short clones. Regions with a restriction pattern comparable to that of pMAX23 are represented by solid horizontal lines. Ends with a unique nucleotide sequence are shown by a broken line. The bowed line in pMAX34 indicates the deletion in this clone. Restriction sites: H, *Hind* III; V, *Eco* RV; S, *Sal* I. Other sites are as in Fig. 1.

cal to that of the 5.8/25S end of pMAX23 (left end in Fig. 1); the other 4 (pMAX4, pMAX7, pMAX15 and pMAX17) have a terminus identical to the other end. Two of the clones, pMAX4 and pMAX15, appear to be the same; otherwise, the divergent ends of these clones bear no resemblance to each other, to any region of the standard pMAX23 or to any sequence in release 55 of GenBank.

We consider three possibilities for the origin of the six clones that have a single rDNA terminus; they may consist of the beginning and ending of tandem arrays, they may have come from orphan rDNA repeat units or they may have arisen from *Hind* III site containing mobile elements adventitiously present in IGS. We consider the last possibility unlikely because we have not observed the presence of a mobile element that lacks a *Hind* III site in any of the 29 analyzed long clones and because this explanation would require the existence of at least three completely different mobile elements each containing at least one *Hind* III site. We also consider it unlikely that they come from orphan units because, except for the unrelated nucleotide sequences, the short inserts are exactly like the comparable segments of the standard pMAX23 insert both in restriction sites and in sequence at the rDNA *Hind* III terminus. Without the homogenization mechanisms generally associated with tandemness, one would, perhaps, not expect the rigid sequence maintenance for orphan repeat units such as we have observed. Thus, we consider it likely that the short clones contain material taken from the beginning or ending of tandem arrays. If this is indeed the case, we have identified at least 3 and possibly 5 different arrays in our non-exhaustive survey involving a total of 36 clones. The analysis of Ohta [16] indicates that the generation time for fixation of a variant in a tandem array as a consequence of unequal crossing over is a function of the square of the number of units in the array. The number of tandem rDNA units in a plant array is unknown and in a plant such as *C. maxima*, with ca. 5000 units per haploid genome, it would be inordinately large if all of the repeat units were in a single array. However, the array lengths might

not be exceptionally long if our interpretation of the nature of the 'short' clones is correct.

Length heterogeneity of IGS in genomic DNA

Genomic DNA was analyzed to determine whether the IGS length heterogeneity observed in the cloned material mirrored the situation in native rDNA or, perhaps, was generated during cloning and culture manipulations despite the use of *recA1* bacterial strains for plasmid replication. DNA isolated from nuclei was enriched for the dense, discrete rDNA containing satellite and restricted with *Eco* RI and *Xba* I to yield 3 fragments of 6.1, 3.5 and 1.4 kb (Fig. 3a). The 6.1 kb fragment contains the IGS and flanking short segments of 25 and 18S rDNA, whereas the other two are composed mostly of rRNA coding sequence (Fig. 1). Figure 3a shows that the staining intensity of the 6.1 kb fragment is less than that of the other two components, whereas if it had been present in equimolar amount to the other two, it would have had a greater staining intensity. The reason for this apparent deficiency is seen in the densitometric scan of the electropherogram (Fig. 3c) which shows that the IGS is probably present in several fragments of different sizes, rather than in a single defined component. In addition to the major 6.1 kb band, there are 3 bumps indicating fragments that are both smaller and larger than the major one, with estimated sizes of 4.8, 5.6 and 6.7 kb, similar to the sizes of the cloned IGS classes II–V (Fig. 1). That these represent IGS size variants is seen in a Southern blot of an immobilized *Eco* RI/*Xba* I digest of rDNA enriched genomic DNA probed with an IGS transcript (Fig. 3b). In agreement with the observation of Ganai and Hemleben [8], a second component of 4.8 kb is apparent in addition to the dominant one of 6.1 kb. Other minor components are also present, both larger than 6.1 kb and intermediate in size between 6.1 and 4.8 kb which are more easily visualized on original autoradiograms than on the reproduction seen in Fig. 3b. We conclude that the IGS size heterogeneity observed in the cloned rDNA repeat units reflects the situation in genomic rDNA.

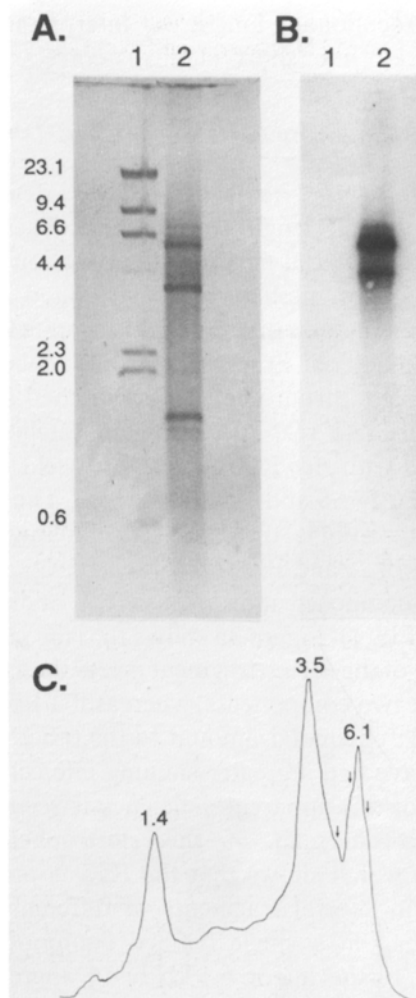


Fig. 3. Analysis of IGS heterogeneity in *C. maxima* genomic DNA. A. Ethidium bromide-stained electropherogram of satellite-enriched genomic DNA cleaved with *Eco* RI and *Xba* I (lane 2). Lane 1 contains *Hind* III cleaved lambda DNA as a size marker. The fragment sizes are indicated on the left. B. Autoradiogram of A after transfer to nylon membrane and probed with a transcript of pMAX23 IGS. C. Densitometric tracing of the electropherogram in A. Peaks associated with band sizes are indicated. Bumps which probably represent IGSs which differ from the major size group are indicated by arrows.

Comparison of rDNA length heterogeneity in individual plants and cultivars

A limited survey was made to determine whether individual plants were all alike in IGS length distribution or whether they might differ in this re-

gard. Electropherograms of *Eco* RI/*Xba* I digested DNA extracted from 6 individual *C. maxima* plants of the Golden Hubbard cultivar and 6 of the Buttercup cultivar were probed with an IGS transcript. Results indistinguishable from those shown in Fig. 3 were obtained in all cases. Thus, it appears, from the limited sample examined, that most plants of these two different cultivars do not differ appreciably in IGS length distribution.

Species differ in the extent to which IGS length distribution varies from plant to plant, accession to accession and cultivar to cultivar. In some, such as broad bean [20] and *Clematis* [13], the distribution differs among individual plants. In others, such as barley [21], wheat [1], and pea [11, 17], accessions and cultivars are internally homogeneous but differ from each other. Our limited survey indicates that *Cucurbita maxima* may have the same length distribution throughout the species although additional data may prove otherwise. What controls this type of variability is unknown and we speculate that it may be a function of the amount and balance of micro- and macro-misalignment during unequal crossing over and the organization of the rDNA into arrays of different lengths.

Acknowledgements

This work was supported in part by NSF grant PCM 8104996, NIH grant GM 32608 and the Wayne State University Center for Molecular Biology. We thank D.K. Davis and J. D'Errico for technical assistance.

References

1. Appels R, Dvorak J: The wheat ribosomal DNA spacer region: its structure and variation in populations and among species. *Theor Appl Genet* 63: 337-348 (1982).
2. Arnheim, N: Concerted evolution of multigene families. In: Nei M, Koehn RK (eds) *Evolution of Genes and Proteins*, pp. 38-61. Sinauer Associates, Sunderland, MA, USA (1983).
3. Coen ES, Dover G: Unequal exchanges and the coevolution of X and Y rDNA arrays in *Drosophila melanogaster*. *Cell* 33: 849-855 (1983).

4. Dover GA: Molecular drive: a cohesive mode of species evolution. *Nature* 299: 111–117 (1982).
5. Dover GA: Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. *Trends Genet* 2: 159–165 (1986).
6. Ellis THN, Davies DR, Castleton JA, Bedford ID: The organization and genetics of rDNA length variants in peas. *Chromosoma* 91: 74–81 (1984).
7. Flavell R: Ribosomal RNA genes and control of their expression. In: Miflin B (ed) *Oxford Surveys of Plant Molecular and Cell Biology*, vol 3, pp. 251–274. Oxford University Press, Oxford (1986).
8. Ganai M, Hemleben V: Comparison of the ribosomal RNA genes in four closely related Cucurbitaceae. *Plant Sys Evol* 154: 63–77 (1986).
9. Goldberg RB, Bemis WP, Siegel A: Nucleic Acid hybridization studies within the genus *Cucurbita*. *Genetics* 72: 253–260 (1972).
10. Ingle J, Timmis JN, Sinclair J: The relationship between satellite deoxyribonucleic acid, ribosomal ribonucleic acid gene redundancy and genome size in plants. *Plant Physiol* 55: 496–501 (1975).
11. Jorgensen RA, Cuellar RE, Thompson WF, Kavanagh TA: Structure and variation in ribosomal RNA genes of pea. *Plant Mol Biol* 8: 3–12 (1987).
12. Kelly RJ, Siegel A: The *Cucurbita maxima* ribosomal DNA intergenic spacer has a complex structure. *Gene* 80: 239–248 (1989).
13. Learn Jr GH, Schaal BA: Population subdivision for ribosomal DNA repeat variants in *Clematis fremontii*. *Evolution* 41: 433–438 (1987).
14. Long EO, Dawid IB: Repeated genes in eukaryotes. *Ann Rev Biochem* 49: 727–764 (1980).
15. Matsuda K, Siegel A: Hybridization of plant ribosomal RNA to DNA; the isolation of a DNA component rich in ribosomal RNA cistrons. *Proc Natl Acad Sci USA* 58: 673–680 (1967).
16. Ohta T: Simple model for treating evolution of multigene families. *Nature* 263: 74–76 (1976).
17. Polans NO, Weeden NF, Thompson WF: Distribution, inheritance and linkage relationship of ribosomal DNA spacer length variants in pea. *Theor Appl Genet* 72: 289–295 (1986).
18. Rogers SO, Bendich AJ: Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. *Plant Mol Biol* 5: 69–76 (1985).
19. Rogers SO, Bendich AJ: Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer. *Plant Mol Biol* 9: 509–520 (1987).
20. Rogers SO, Honda S, Bendich AJ: Variation in the ribosomal RNA genes among individuals of *Vicia faba*. *Plant Mol Biol* 6: 339–345 (1986).
21. Saghai-Marooof MA, Soliman KM, Jorgensen RA, Allard RW: Ribosomal DNA spacer-length polymorphisms in barley; Mendelian inheritance, chromosomal location and population dynamics. *Proc Natl Acad Sci USA* 81: 8014–8018 (1984).
22. Siegel A, Kolacz K: Heterogeneity of pumpkin ribosomal DNA. *Plant Physiol* 72: 166–171 (1983).
23. Smith GP: Evolution of repeated sequences by unequal crossover. *Science* 191: 528–535 (1976).
24. Sollner-Webb B, Tower T: Transcription of cloned eukaryotic ribosomal RNA genes. *Ann Rev Biochem* 55: 801–830 (1986).
25. Southern EM: Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 98: 503–517 (1975).
26. Szostack J, Wu R: Unequal crossing over in the ribosomal DNA of *Saccharomyces cerevisiae*. *Nature* 284: 426–430 (1980).
27. Yakura K, Kato A, Tanifuji S: Length heterogeneity of the large spacer of *Vicia faba* rDNA is due to the differing number of a 325 bp repetitive sequence elements. *Mol Gen Genet* 193: 400–405 (1984).
28. Zagursky R, Baumeister K, Lomax N, Berman M: Rapid and easy sequencing of large linear double-stranded DNA and supercoiled plasmid DNA. *Gene Analysis Techniques* 2: 89–94 (1985).