PLANT MOLECULAR BIOLOGY REPORTER
Volume 1:1 (1983)
Pp. 9-16

REVIEW

Evaluation of Genomic Variability at the Nucleic Acid Level

CAROL J. RIVIN, ELIZABETH A. ZIMMER, CHRISTOPHER A. CULLIS, VIRGINIA WALBOT (Department of Biological Sciences, Stanford University, Stanford, CA 94305), THANH HUYNH and RONALD W. DAVIS (Department of Biochemistry, Stanford University, Stanford, CA 94305)

The study of variation is the basis of genetic analysis, evolutionary studies, and population biology, and of course, is fundamental in the improvement of agricultural plants. In this context, the uncovering of new sources of variation and the ability to describe novel types of variation is of great importance to plant biologists. In maize, as well as other plants, variation has been widely studied at several levels: for phenotypic characters such as morphology and plant products (Goodman and Bird, 1977), for protein level changes as are revealed by isozyme analysis (Newton and Schwartz, 1980), and at the level of the whole genome by analyzing the kinetics and stability of DNA:DNA reassociations in solution hybridization (Hake and Walbot, 1980).

Recently, it has become possible to examine genome variation in much greater detail through the use of specific cloned DNA probes. The use of cloned probes makes direct nucleic acid analysis a fairly simple and extremely sensitive method of detecting types of variability that are inaccessible in phenotype or protein analysis. We will show below that we can easily detect extensive variation between modern inbred lines of maize, as well as between maize and its close relatives, the teosintes. By using clones of repetitive DNA sequences, we have discovered considerable changes in the copy number of repeated DNA, and many restriction site polymorphisms. These types of variation cannot be detected at the phenotypic or protein level, although they may have important consequences in plant growth and adaptation. We also show that large numbers of restriction site polymorphisms can be found in the vicinity of genes defined by the use of cloned cDNA probes, and that these can be used to identify and map new genetic loci without having to find new phenotypes or new enzyme activities.

Copy Number Variation in Repetitive Sequences of Maize

The number of copies of highly repeated DNA sequences has been found to vary markedly between the genomes of related species (Flavell, 1982). We have asked if variation for copy number could also be found within the species *Zea mays* by comparing inbred lines. Additionally, we are interested to learn if there is any coordination in the copy number of

Offprint requests to: Carol J. Rivin

various repeated sequences within these specific genetic backgrounds, and whether the organization of the sequences affects that range of copy number variation they display (i.e., are dispersed repeats as variable in copy number as tandemly arrayed repeats?).

To answer these questions, we have measured the copy number of various highly repeated sequences in the genomes of several inbred maize lines. The data for four different sequences is shown in Figure 1. These cloned sequences were isolated in several ways.

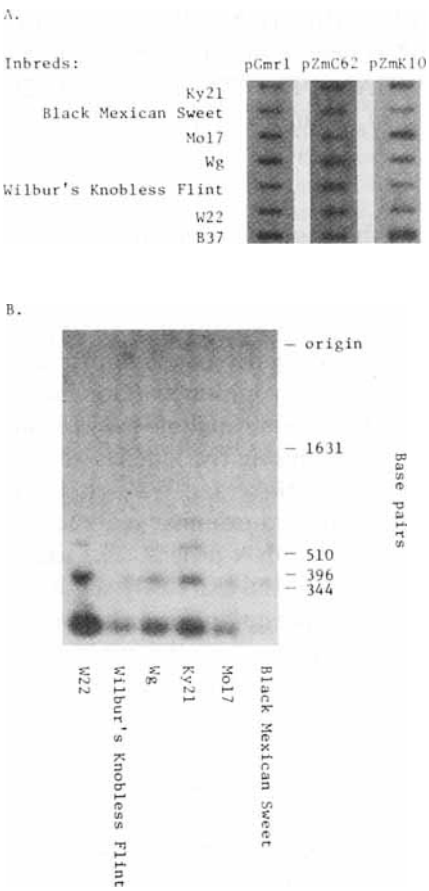

Figure 1. Copy number variation for repeated DNA sequences.

Figure 1A. Hybridization of cloned repeated DNA sequences to uncut maize DNA. 0.5 μg of unrestricted nuclear DNA isolated from 7 inbred lines (Rivin et al., 1982) was denatured and fixed to nitrocellulose paper through a slotted template. Identical "slot blots" were hybridized to nick translated plasmids carrying repeated DNA sequences. pGmr1 is a ribosomal DNA clone derived from soybean. pZmK10 and pZmC62 have short sequences of maize nuclear DNA. The films have been exposed to give a similar intensity of signal for each probe. Therefore, the relative copy number can be meaningfully compared within the DNA samples using a single probe, but relative copy numbers between probes can not be compared in these autoradiograms.

Figure 1B. Hybridization of cloned repeated DNA to restricted genomic DNA. 0.5 μg of nuclear DNA from 6 inbred lines was digested with *HaeIII*, electrophoresed in 3% agarose and blotted to nitrocellulose. The blot was hybridized with a nick translated plasmid, pZmK6, which carries a 180 base pair *HaeIII* fragment of maize DNA. Molecular weight markers are pBR322 digested with *Hinf I*.

pGmr1 is an rDNA probe from soybean that cross-reacts with corn rDNA. pZmK6 and pZmK10 are sequences that appear as bright, highly repeated bands on an *HaeIII* digest of maize DNA. pZm62 was selected from a pool of random *EcoR1-HmdIII* fragments.

To quickly measure the copy number of many sequences in a range of inbred lines, nick-translated recombinant plasmids were hybridized to equal amounts of uncut maize DNA immobilized on nitrocellulose filters. As shown in Figure 1A, the DNA was applied through a slotted template so that each type of DNA is represented as a discrete band. By making several exposures of the autoradiogram and quantifying the results with a densitometer, we have determined the extent of copy number variation between the inbred lines, and the pattern of hybridization of the set of probes within individual genetic backgrounds. In Figure 1B, the recombinant plasmid is hybridized to a blot of DNA digested with *HaeIII*. As in the "slot blots," copy number variation among the inbred lines is obvious and can be quantified. This type of experiment also provides information about the organization of the repeat units. The ladder pattern shown for pZmK6 is typical of tandemly arrayed repeat units; the occasional loss of restriction sites creating dimers and trimers of the basic repeat unit. pZmK10 and the rDNA are also tandemly repeated sequences.

Our data show that each probe has a distinctive pattern of hybridization intensity to the set of maize DNAs, and each displays some copy number variation. pZmK6 is the most variable, having a six-fold range of copy number in the lines tested. pZmC62 varies two-fold, as does the rDNA probe. pZmK10 has a four-fold range of copy number among the inbred lines. We obtained no evidence for coordinate control of copy number of the four probes; that is, they were not uniformly high in one line and low in another. Each inbred line was also different. No two had exactly the same quantities of each of the repetitive sequences. It is also apparent that the proportion of dimer and trimer relative to monomer of the pZmK6 sequence is variable between lines. We are investigating further how these polymorphic forms of the repeat may be varying independently in copy number. One of these sequences, the rDNA repeat, has been extensively characterized with respect to its restriction site polymorphism, which is described below.

Variation in Ribosomal DNA

The genes coding for ribosomal DNA (rDNA) exist in the genome as tandemly repeated units of transcribed interspersed with nontranscribed DNA (Figure 2). The overall pattern of rDNA organization is one of high conservation within species, with some divergence between lines or species. This pattern of within group homogeneity and between group heterogeneity is termed concerted evolution (Zimmer et al., 1980). Concerted evolution of rDNA genes previously has been described for yeast (Petes, 1980), *Drosophila* (Coen et al., 1982), and mammals (Arnheim et al., 1980).

We observe a similar pattern of rDNA change with restriction mapping studies of the rDNA genes of maize and teosintes. In addition to copy number variation between inbred lines, variation for restriction endonuclease cleavage sites and for the length of the nontranscribed spacer were observed among inbred lines. As expected, a greater number of changes are detected when the inbred lines of *Zea mays* are compared to the more distantly related teosintes.

For example, Figure 2 describes the results of using the restriction endonuclease *EcoR1* to examine the rDNA genes of *Zea* lines. Figure 2A is a schematic illustrating the organization of maize rDNA. In lines where *EcoR1* cuts just once per repeat unit (lines with the

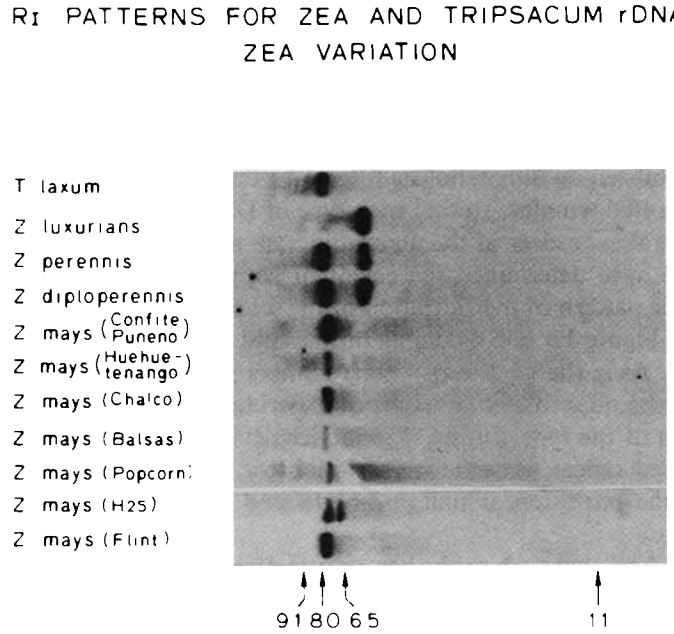

Figure 2. Maize rDNA genes

Figure 2A. A typical ribosomal repeat unit of *Zea*. Large filled-in boxes indicate those parts of the DNA that code for the mature 17S, 5.8S and 26S ribosomal RNAs. "E" indicates cleavage sites for the restrication endonuclease *EcoRI*. Asterisks mark those sites which are heterogeneous within the rDNA array of an individual.

Figure 2B. DNA fragments containing the rDNA genes of maize and teosintes. Nuclear DNA was digested with the enzyme *EcoRI*, subjected to electrophoresis in an agarose gel and hybridized with a nick-translated rDNA probe, pGmr1. Numbers represent the sizes of the fragments in Kb. "Z" refers to *Zea* and "T" to the genus *Tripsacum*. Flint, H25 and Popcorn are inbred lines of *Zea mays*, Confite Puneno is a South American *Zea mays*, and Balsas, Chalco and Huehuetenango are Mexican teosintes.

E site only such as *Z. mays*, Flint, Figure 2B), an rDNA fragment of 9Kb is generated. This conserved *EcoRI* site has been mapped to the 3' end of the 26 S gene. There are two additional *EcoRI* sites (E_1* and E_2*) noted in this figure. The presence of the E_1* site generates 8 Kb and 1 Kb fragments (see *Z. mays* H25 in Figure 2B). The E_1* site has been detected in a number of inbred lines of maize. The presence of the E_2* site results in 6.5 Kb and 2.6 Kb fragments (the 2.6 Kb fragment is only faintly detectable, due to use of the heterologous spacer probe). The E_2* site is found in *Zea diploperennis*, *Zea perennis*, and *Zea luxurians* rDNA arrays.

As can be seen in Figure 2B, both E and E + E_1* or E + E_2* type arrays can exist within an inbred line and even within individual plants. We are interested in the distribution of variants such as the E_1* site along the rDNA array, as the pattern of variation may reflect the mechanism(s) by which array homogeneity is maintained.

Restriction Fragment Length Polymorphisms Near Genes in the Maize Nuclear Genome

The two classes of maize DNA considered so far can both be described as highly repeated DNA sequences, present in thousands of copies per haploid genome. We have also

addressed the question of how much variability exists between different inbred lines of maize in the region of unique or nearly unique DNA sequences. This experiment required the use of two tools. First, it required probes for unique DNA sequences. We chose as a source of probes a library of cDNA clones made by reverse transcription of poly (A)$^+$ RNA isolated from etiolated seedlings. Second, it was necessary to have a means of detecting DNA sequence variability. This was provided by a number of restriction endonucleases, each of which cleaves double-stranded DNA at a specific nucleotide sequence. The design of the experiment was to determine the length of the restriction fragment of maize DNA to which a particular unique sequence DNA hybridized. Single-base changes and DNA rearrangements occurring during the evolution of the species cause the lengths of restriction fragments to change. Therefore, differences between inbred lines of maize in the DNA sequence near a particular gene would sometimes be detected as differences in the lengths of the restriction fragments containing that gene.

An example of the kind of results that are obtained when a low copy number DNA probe is hybridized to restriction fragments of maize DNA isolated from a number of different inbred lines is shown in Figure 3. DNA isolated from 14 randomly chosen inbred lines of maize was digested with the restriction endonuclease *EcoRI*. The resulting DNA fragments were fractionated by size in an agarose gel, transferred to paper, then hybridized with a radioactively labeled cDNA clone probe. As can be seen by examining any individual lane in Figure 3, the probe hybridized to several bands in *EcoRI*-digested maize DNA. The bands range in intensity from dark to light. The darkest band in each lane is of the intensity that would be expected of a single-copy gene. By comparing different lanes in Figure 3 to each other, it can be seen that while the probe hybridized to several bands in any one lane of maize DNA, the pattern of bands produced is different for each inbred line. Therefore the pattern of bands produced using a DNA probe can be used to identify a particular inbred line and hybrids between inbred lines.

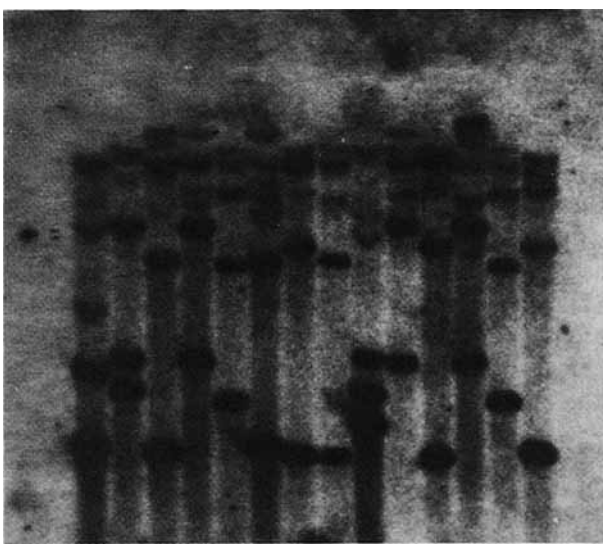

Figure 3. Restriction fragment length polymorphism near low copy number genes. DNAs isolated from 14 inbred lines of maize were digested with *EcoR1*, run in an agarose gel, transferred to diazotized paper, and hybridized with a ^{32}P-labeled cDNA clone probe.

We have used a classical genetic analysis to show that our probe is hybridizing to a small related gene family of four to six loci, and we can assign the hybridizing bands to particular loci within that group. We have analyzed the hybridization patterns of DNA from F1 and F2 individuals from a cross of inbred lines. F1 plants show all the bands present in the parents, and in F2 plants these bands segregate. By pairwise analysis of the bands in the F2 plants, we can show which bands segregate from each other as expected of polymorphisms at a single locus, and which segregate independently as expected of unlinked loci. In this way, we can use cDNA probes to detect many new genes and gene families, and because we find a high degree of restriction site polymorphisms surrounding all these genes, we can use classical genetic techniques to map them.

New Genetic Markers for the Maize Genome

An important application of the variation we have observed in the maize genome will be its contribution of a very large number of new genetic markers to the maize linkage map. Traditionally, variants for use in genetic mapping have been identified at the level of the phenotype of the organism or at the protein level. Now it is possible to define many new genetic loci with DNA probes, without the need to correlate them with new plant phenotypes or new enzyme activities. The loci can be defined because the DNA restriction fragments on which they reside are polymorphic. Their polymorphism allows them to be mapped by classical genetic techniques.

We believe that a very large number of new genetic loci can be identified in this way, perhaps many thousands. However, it is yet to be determined whether all regions of unique DNA sequence in the maize genome are as polymorphic. The new loci can be added to the existing linkage map using exactly the same methods that are used to map phenotypic and isozyme variants. Moreover, the three types of polymorphisms could be mapped with respect to one another simultaneously, since all could be scored in a single plant.

An example of an area in which a detailed genetic linkage map would answer a need is the area of plant breeding. For instance, when a linkage map of maize has been generated of sufficient detail, it may be possible to link specific restriction fragment length polymorphisms to important agronomic traits. If it is known that a desirable adult plant trait is closely linked to a particular restriction fragment length polymorphism, DNA from seedlings could be screened for that polymorphism to identify seedlings during a backcrossing program that are likely to contain the linked trait. By performing the screening procedure while the plants are at the seedling state, much time, space, and expense could be saved that would otherwise be spent screening a large number of adult plants for adult characteristics. Restriction fragment length polymorphisms are also likely to be found associated with unique DNA sequences of plant species other than maize. The advantages that we have described of using restriction fragment length polymorphisms as genetic markers are likely to apply to other important crop species.

Prospects

We have shown that direct examination of DNA is a very sensitive method for detection and analysis of variation within a species and between closely related species. Variation in both transcribed and nontranscribed sequences can be investigated, and changes can be detected in primary sequence, organization, and copy number. Using the polymorphisms we have discovered with various kinds of probes, we hope to address the question of how

flexible the maize genome may be: what kinds of sequences are subject to variation? of what kind? at what rate? by what mechanism? We want to know how sequence homogeneity is maintained in repetitive DNA, but also how variants can be propagated and amplified. We have the tools to ask whether sequence variation is more common in sequences adjacent to unique DNA than in tandem arrays of repeats that may be subject to homogenization mechanisms, and whether tandem and dispersed repeats are different in this regard. We can also ask if the ability to vary copy number and the range over which it may vary is dependent on the organization of the sequence and/or whether it is transcribed.

Although we can detect a great deal of variation at the nucleic acid level and infer something about the rates and mechanisms of this variation, we cannot yet relate these genomic changes to the organismal change that we measure. The rate of genomic change compared to organismal change is certainly not constant in different taxa of animals (Wilson et al., 1977), and a similar situation may exist for plants as well. Maize represents a case in which there have been recent radical alterations in organismal phenotype in both vegetative and reproductive characters. Both domestication and modern breeding have altered the phenotype (Walbot, in press). It is an open question whether or not these pressures have resulted in the kinds of variation we have observed or whether such changes are of agronomic value.

Phillips (1981) has demonstrated that maize rDNA number varies over approximately a two-fold range in different inbred lines of maize and that a similar range of variation existed for lines selected over the past 70 years for high or low protein content. Cullis (1981) has proposed that environmental stresses result in heritable changes in rDNA copy number in flax, and changes in other repetitive DNA sequence families may be correlated with organismal change. We can imagine ways in which nontranscribed sequences may also be important in selection. For example, chromosome knobs are variable features of the maize karyotype that are largely composed of a tandemly arrayed repeated sequence (Peacock et al., 1981). The knob structure participates in a meiotic drive mechanism that can lead to the preferential inheritance of knobbed chromosomes from heterozygous plants. Genes closely linked to knobs would also be preferentially inherited. Knob sequences could themselves accumulate through selection pressures on closely linked loci. Even without any structural or genic role, DNA sequences could be selected for copy number variation for their nonspecific effect on genome size.

While the significance of genomic variation to plant growth remains to be tested, we expect that the diversity we have found between inbred lines may have practical applications for plant breeding. Cloned probes could be used to test the purity of inbreds, to assess the relatedness of lines, and to add a vast number of new loci to the genetic linkage map.

Acknowledgments

Research was supported by grants from the USDA and NSF to V.W., by a NIH postdoctoral fellowship to E.A.Z., and by a NIH grant to R.W. Davis. C.A.C. was supported by a Civil Service (Nuffield and Leverhulme) Travelling Fellowship.

REFERENCES

Arnheim, N., M. Krystal, R. Schmickel, G. Wilson, O. Ryder, and E. Zimmer (1980) Molecular evidence for genetic exchanges among ribosomal genes on non-homologous chromosomes in man and apes. Proc. Nat. Acad. Sci. 77: 7323-7327

Coen, E., T. Strachan, and G. Dover (1982) Dynamics of concerted evolution of ribosomal DNA and histone gene families in the *melanogaster* species subgroup of *Drosophila*. J. Mol. Biol. 158: 17-35

Cullis, C.A (1981) Environmental induction of heritable changes in flax: defined environments inducing changes in rDNA and peroxidase isozyme band pattern Heredity 47: 87-94

Flavell, R (1982). Sequence amplification, deletion and rearrangement: major sources of variation during species divergence In. *Genome Evolution* (eds G A Dover and R B Flavell), Academic Press, New York. pp 301-323

Goodman, M M and R McK Bird (1977) The races of maize IV: tentative grouping of 219 Latin American Races. Economic Botany 31 204-221.

Hake, S and V Walbot (1980) The genome of *Zea mays*, its organization and homology to related grasses Chromosoma (Berl) 79 251-270

Newton, K J and D Schwartz (1980) Genetic basis of the major malate dehydrogenase isozymes in maize genetics 95 425-442

Peacock, W J , E S. Dennis, M M Rhoades, and A J Pryor (1981) Highly repeated DNA sequence limited to knob heterochromatin in maize Assoc Nat Acad. Sci 78 4490-4494

Petes, T D (1980). Unequal meiotic recombination with tandem arrays of yeast ribosomal DNA genes Cell 19: 767-774

Phillips, R L (1978) Molecular cytogenetics of the nucleolus organizer region In. *Maize Breeding and Genetics* (ed D B Walden), Wiley-Interscience, New York pp 711-741

Rivin, C J , E A Zimmer, and V Walbot (1982) Isolation of DNA and DNA recombinants from maize In: *Maize for Biological Research* (ed W F Sheridan), Plant Molecular Biology Association, Charlottesville, Va pp 161-164

Walbot, V (in press) Morphological and genomic variation in plants: *Zea mays* and its relatives In *British Society for Developmental Biology Symposium No. 6 Development and Evolution* (eds. B.C Goodwin, et al), Cambridge University Press, Cambridge, England

Wilson, A C , S S Carlson and T J White (1977) Biochemical Evolution Ann Rev Biochem 46. 573-639

Zimmer, E.A , S L Martin, S M Beverley, Y W Kan and A Wilson (1980) Rapid duplication and loss of genes coding for the chains of hemoglobin Proc Nat Acad Sci. U S A 77: 2158-2162.

Received December 7, 1982