The structure and evolution of angiosperm nuclear genomes

Jeffrey L Bennetzen

Despite several decades of investigation, the organization of angiosperm genomes remained largely unknown until very recently. Data describing the sequence composition of large segments of genomes, covering hundreds of kilobases of contiguous sequence, have only become available in the past two years. Recent results indicate commonalities in the characteristics of many plant genomes, including in the structure of chromosomal components like telomeres and centromeres, and in the order and content of genes. Major differences between angiosperms have been associated mainly with repetitive DNAs, both gene families and mobile elements. Intriguing new studies have begun to characterize the dynamic three-dimensional structures of chromosomes and chromatin, and the relationship between genome structure and co-ordinated gene function.

Addresses
Department of Biological Sciences, Purdue University, West Lafayette, IN 47907-1392, USA; e-mail: maize@bilbo.bio.purdue.edu

Current Opinion in Plant Biology 1998, 1:103–108

http://biomednet.com/elecref/1369526600100103

Abbreviations

LTR	long terminal repeat
MAR	matrix attachment region
MITEs	miniature inverted repeat transposable elements

Introduction

The genomes of flowering plants are less well characterized than are those of model fungi or animals. Many hundreds of plant genes have been cloned, sequenced and functionally characterized, however, as isolated segments of DNA. As in other eukaryotes, plant genes are relatively small and discrete entities consisting mainly of coding regions (introns) and a few hundred basepairs of regulatory sequences. Although regulation of a gene by sequences at a great distance may be profoundly important in plants, as recently demonstrated by such phenomena as transgene-mediated silencing [1,2,3•], the minimal components necessary for fairly specific gene expression and regulation are usually present on only a few kilobases of contiguous sequence containing the gene. Hence, an understanding of this most basic level of gene structure and regulation has not required any significant understanding of genome structure. The importance of genome structure to gene regulation, however, is indicated by the routine observation that the same transgene integrated into different locations will display different levels of expression [1,2,3•].

Comprehensive studies of the structure of the nongenic DNA of plants, and its relationship to gene function and chromosomal mechanics, have begun only recently. Five years ago we knew that plant genomes vary greatly in size, caused partly by variation in ploidy but largely by differences in the amount of repetitive DNA [4]. We also knew that some of these repetitive DNAs were tandemly arrayed genic or nongenic sequences, and many were mobile DNAs somehow interspersed with genes and other repeats [5]. Until the advent of technologies that allowed the cloning and analysis of large segments of contiguous plant DNA, this vague description could not be significantly augmented.

Recent discoveries indicate that plant genomes have very specific patterns of organization for their nongenic sequences. One advantage of flowering plant genome studies is that, although few studies have been very deep or comprehensive, a wealth of different species has been investigated. This allows an evolutionary perspective often lacking in other eukaryotic systems.

Perhaps the most exciting developments have occurred over the past five years or so, which have produced studies directly investigating the relationship between DNA sequence arrangement, chromosome structure, and genetic activity in plants. These studies have utilized a whole array of molecular and cytogenetic tools to uncover how chromosome structure changes over time, and how the expression and three dimensional structure of an individual gene may be affected by its chromosomal location. This review begins by describing the common structural features of angiosperm genomes, and goes on to indicate how these structures are conserved or vary over time and between species. The review continues with a discussion of groundbreaking insights into the relationship between genome structure and co-ordinated gene/chromosome function and concludes with a brief glimpse at the many questions of genome organization and behavior that remain to be answered.

The components of any angiosperm genome

Like all other eukaryotes, standard plant chromosomes contain genes, mobile repetitive DNAs (otherwise known as transposable elements), and various classes of tandemly repeated sequences. Many of the tandem sequences, including telomeric repeats and ribosomal DNA repeats, are essential for the survival of the organism. Other tandem or interspersed repeat classes, like minisatellites, microsatellites and transposable elements, may be parasitic or selfish DNAs. Although repetitive sequences are present in all plants, their relative representation can be quite variable. For instance, tandem satellite repeats make up a major share of the total repetitive DNA in *Arabidopsis* but are a minor component of the maize genome where

the vast majority of repetitive DNA is composed of retrotransposons [6••,7].

Genes and gene families

Genes are present on most plant chromosomes, with the possible exception of the heterochromatic B chromosomes found in some isolates of many plant species. Genes tend to be unevenly dispersed along a chromosome: they are rare or totally absent from centromeres, telomeres and other heterochromatic sites, and may be largely missing from the majority of the chromosome, particularly in big chromosomes with large and heterochromatic centromeres, like those of wheat [8].

In small-genome species, genes may be very close to each other [9,10]. For example, in a 190 Mb region of *Arabidopsis* chromosome 4, genes are found at a density of about one per 4.8 kb [10]. Often, only a few hundred basepairs separate individual genes. In large-genome species, genes are often separated by large blocks of repetitive DNA. For instance, in the *Adh1*-F region of maize, about 80% of the DNA is repetitive and 4–8 genes are found in a contiguous segment of 240 kb [6••,11]. Even in large-genome species, however, tandemly repeated genes may often be found very close together, as with the ribosomal DNA. This presumably reflects either a recent origin of the repeated genes (without time for subsequent intervening mobile DNA insertions) and/or a concerted process to remove most subsequent sequence changes by unequal recombination.

Centromeres

Plant centromeres are defined genetically and cytologically as chromosomal constrictions that are required for chromosomal segregation in mitosis and meiosis. Centromeric chromatin is highly condensed and, in large chromosomes, can comprise over half of the entire chromosome. At the DNA level, a plant centromeric region can be as small as 1 Mb (e.g. *Arabidopsis*) or over 100 Mb (e.g. wheat) [8]. Many of the sequences in these heterochromatic regions can also be found elsewhere in the genome [12,13]. In fact, it is expected that most centromere-associated sequences are not involved in centromere function, but have accumulated within the associated heterochromatin, perhaps due to the lack of recombinational activity of the DNA in these large centromeric regions. The required functional components of a centromere have not been well defined in plants, although rearrangement of B chromosomes in maize has shown that a single centromere can sometimes be divided into more than one functional centromere, with a size of less than a megabase [14]. In many cases, plant chromosomes contain noncentromeric sites that can exhibit centromeric behavior under some circumstances. These 'neocentromeres' are generally condensed heterochromatic structures, and have been associated with tandemly repeated minisatellites, such as the knob satellite of maize [15]. Recent cloning and *in situ* hybridization studies have identified sequences that are tandemly repeated at all cereal centromeres [16,17]. Although not a direct proof of function, the conservation of centromere-associated repeats suggests that these repeat sequences are necessary for centromeric function. Moreover, the fact that neocentromeres, B chromosome centromeres, and standard chromosomal centromeres all share several features, including tandem repeats of about 180 bp and a highly heterochromatic state, suggests that this structure is required for centromeric activity. Recent studies in maize have shown that a mutation in *suppressor of meiotic drive 1* affects neocentromeric function [18], an observation that may eventually provide access to a protein that will link centromere DNA/chromatin structure to chromosome segregation.

Telomeres

As in most other eukaryotes, the termini of the linear plant chromosomes in the nucleus contain short tandem repeats of a sequence added by the enzyme telomerase [19]. Telomerase activity compensates for the inability of semi-conservative DNA replication to replicate linear chromosome ends, and also prevents the various activities (chromosome fusion, terminal degradation, etc.) that McClintock first noted for broken chromosome ends [20]. Recent cloning studies have directly investigated the healing of broken chromosome ends in wheat by the addition of telomere repeats [21]. Telomere repeat sequences are fairly similar across all eukaryotes, indicating the conserved nature of the RNA template within the telomerase reverse transcriptase. Other repeats are found preferentially associated with telomeres [22]. It is not known whether these sequences might be parasitic or selfish sequences that accumulate at telomeres or whether they play some role in telomere function. Plants often have some telomere repeats near their centromeres [23,24], or centromere repeats near their telomeres [17], suggesting either that full arm inversions or translocations are common in evolution or that many chromosomes have arisen by centromere fusions, or both. Like centromeres, telomeres may play an important role in the process of chromosomal synapsis and/or segregation, as suggested by their clustering near the nuclear periphery before meiotic prophase [25••].

Satellite repeats

Tandemly repeated sequences are scattered throughout all eukaryotic genomes. Short tandem repeats of a few base pairs, so-called microsatellites or simple sequence repeats, are usually hypervariable and have become useful markers for genetic mapping. Although they are found on all chromosomes in large numbers, their small size dictates that they will not represent more than a tiny percentage of a total plant genome. The overall genomic organization of microsatellites has not been extensively investigated, although they do appear to be somewhat over-represented in introns [9,10] and under-represented in mobile DNAs (P Miguel, A Tikhonov, personal communication). The larger minisatellite repeats are also ubiquitous in plants,

although they rarely make up more than a few percent of the total genome size. Monomer minisatellite repeats are often 180 bp–220 bp in size, compatible with a single repeat per nucleosome. These minisatellites commonly are present in many thousands of tandem copies, where they will form a large and fairly homogeneous knob of heterochromatin.

Mobile DNAs

Mobile DNAs fall into two major categories, those that transpose as DNA molecules and those that transpose through an RNA intermediate. The better-known elements, like *Ac* and *En/Spm*, fall into the former class. Most of these elements contain short inverted repeat termini, and preferentially insert into gene-associated regions [26]. The copy numbers of these elements are usually fairly low, often ≤4 per genome for the active element (e.g. *Ac*) that encodes the mobilizing transposase, and usually less than a few hundred defective elements (e.g. *Ds*) that respond to the transposase. Hence, these elements comprise at most a few percent of any plant genome. One exception to this general low copy number for inverted repeat (DNA) elements is provided by the miniature inverted repeat transposable elements (MITEs). These elements can be found in thousands of copies per genome, and fall into several families [27,28]. The mechanism of transposition for these elements is not known, although they appear in many ways similar to small defective elements like *Ds1*. Like other inverted repeat transposable elements in plants, MITEs are found mainly in or near genes. Most recently, MITEs have been found to be preferentially associated with putative matrix attachment regions [29•]. Because of their general small size, around 200 bp or so, MITEs tend to represent only a small part of most plant genomes despite their often high copy number.

Mobile elements that move through an RNA intermediate are called retroelements, and fall into several classes that are found in all eukaryotes, including plants [28,30,31•]. In large-genome plants like barley, lily and maize, retrotransposons that contain long terminal repeats (LTRs) make up the majority of the DNA in the nuclear genome. These elements tend to be fairly large, from a few kilobases up to more than 15 kb, and can have copy numbers ranging anywhere from one copy to hundreds of thousands of copies per genome [6••,31•]. In maize the highly repetitive LTR-retrotransposons are primarily found in nested blocks between genes, indicating a preference for insertion and/or retention within inactive and methylated regions [6••]. Lower copy number retrotransposons do sometimes insert within or near genes, and these gene-associated insertions can serve as the raw material for the evolution of new *cis*-regulators of adjacent genes [25••,32].

The evolution of genome organization: colinearity

Sequence changes associated with genes have been extensively studied by evolutionary biologists, who have utilized the process of genetic drift within genes as a basis for phylogenetic characterization. Cytogeneticists have also studied the nature and rate of changes in chromosome structure and number in plants. Because of technical limitations, little has been known about the nature of genome change between the levels of gene sequence and cytogenetic structure. Recombinational mapping using DNA markers, artificial chromosome clones containing large DNA inserts, and improved sequencing and gel technologies now allow scientists to fill that gap.

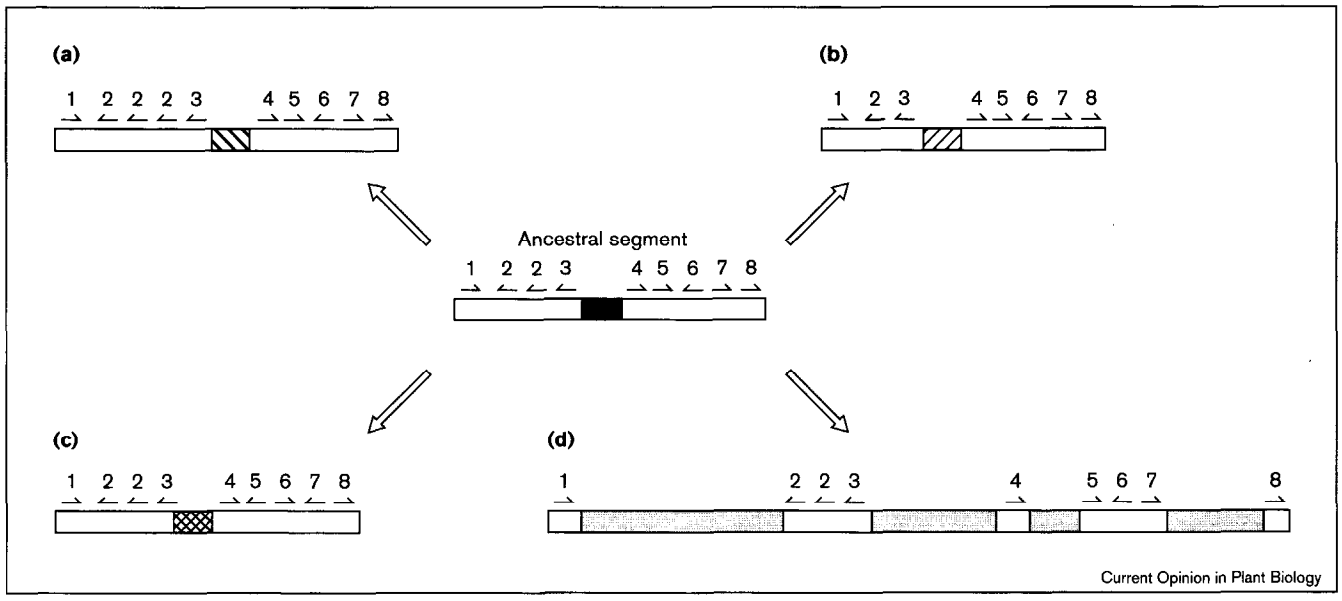
Recombinational maps using the same DNA probes in different species have indicated that different plant genomes have similar overall gene content and extensive regions of colinear gene order [33–35]. This is particularly true among the grasses, where colinearity (or synteny) can be a basis for the merging of knowledge and tools from all grass species into a single model system [36]. Many exceptions to this colinearity have been noted, but even distantly related angiosperms (like sorghum and *Arabidopsis* [37]) may have a useful frequency of colinear segments. Among the more frequent rearrangements observed are chromosome fissions/fusions; whole chromosome arm duplications, translocations or inversions; large segmental duplications; and variation in the copy numbers of gene families. None of these events occurs at a scale that significantly limits the usefulness of colinearity.

Relatively few studies have tested for small rearrangements within otherwise colinear genomes. In most cases, these studies have uncovered extensive similarity in local gene order, at the level of adjacent loci, but with some exceptions [9,11,38–40]. Disease resistance genes, which may evolve by different and more rapid mechanisms than most genes [41], often may not show colinearity [42]. As shown in Figure 1, the various local rearrangements that may be common in plants would usually interfere with microcolinearity, but would not be detected by standard recombinational mapping. For instance, none of the changes depicted in Figure 1 would qualitatively affect the linkage of markers 1 and 8, and most recombinational maps have a probe density and population size too small to score markers this close together.

The evolution of genome size

Plants vary tremendously in genome size, from the 110 Mb of *Arabidopsis* to some lily genomes that are about 1000-fold larger. Some of this variation is caused by differences in ploidy, but most can be attributed to higher amounts of mobile repetitive DNAs. All plants appear to have many different families of these elements, of all types, but larger genome plants appear to accumulate some small subsets of these elements (a few families of LTR-retrotransposons) at very high copy numbers [7,30,31•]. Hence, it is these elements that appear to be responsible for the C-value paradox, at least in the grasses [7,43•]. It is not clear whether small genome plants have less of these elements because they are better

Figure 1



Possible local variations in genome structure that might be both frequent and undetected by recombinational mapping. The central bar depicts a small chromosomal segment containing nine genes whose placement, relative transcript size, and transcript orientation are shown by the small arrows. Gene 2 is already duplicated in this ancestral genome. **(a)** Tandem duplication of gene 2. **(b)** Deletion of a copy of gene 2. **(c)** Small inversion, involving genes 5, 6 and 7. **(d)** Amplification of genome size by insertion of nongenic DNA between genes. The filled blocks indicate mobile and/or repetitive DNAs, and the differences in the fill indicate that these sequences usually do not cross-hybridize between species outside of the same genus (because they are different elements and/or because they have diverged rapidly).

able to inhibit their amplification, because they have somehow avoided contact with sources of most LTR-retrotransposons, or because they have some unknown mechanism for removal of these repeats [43•]. Unlike genes, retrotransposons do not usually cross-hybridize between even closely related plant species, suggesting that they often may have been derived from recent horizontal transfers (for instance, by viral infection or via a wide cross). A retrovirus-like gene transduction by some of these elements [31•,44,45] also suggests an infectious transfer, as does their activation by stress and their ability to function across wide species boundaries [30,31•,46•]. These LTR-retrotransposons, however, apparently evolve more rapidly than adjacent genes, at least partly because of their methylated and presumably heterochromatic status, and this may be responsible for their lack of detected conservation in related plants.

Genes in a region: structure and function

In plants with small genomes, genes are often very close together. In larger-genome plants, genes are often separated by long stretches of apparently nongenic sequences. For instance, in sequenced regions of *Arabidopsis*, rice, sorghum and maize, gene distances are about one per 5 kb, 10 kb, 10 kb, and 40 kb, respectively, in accord with their respective genome sizes of 110 kb, 440 kb, 750 kb, and 2400 Mb [9–11]. The genes themselves, in comparable regions, are of similar size. Moreover, because recombination is largely limited to genes, five genes scattered across 200 kb in maize may cover the same number of centiMorgans as those five genes do when present in 50 kb of rice DNA. In many large plant genomes, most of the repetitive DNA may be associated with centromeric heterochromatin, leaving the gene-containing regions to have a fairly high gene density [8] which is comparable to that of smaller-genome species.

Plant genomes have found mechanisms, probably the various epigenetic phenomena associated with transgene-mediated suppression and position effects [1,2,3•], that keep mobile and other repetitive DNAs in a usually inactive form. Somehow, plants have managed to differentiate between desirable repeats, like gene families, and potentially damaging repeats, like transposable elements. In many ways, genes that exhibit paramutation [47] behave as though they are caught somewhere between these two poles. Studies in animals, and some in plants [3•,48•], suggest that flanking matrix-attachment regions (MARs) can decrease the frequency of transgene inactivation. In cases where MARs have been mapped across large contiguous segments of plant DNA, they were often found to be adjacent to promoters, in many cases separating those promoters from repetitive/heterochromatic DNA blocks [49,50•]. Most recently, a comparison of the *Sh2/A1*-homologous regions of rice and sorghum has shown conservation of MAR location between the two species,

even though the primary DNA sequences that provide MAR function were not detectably conserved [29•].

Conclusions

We are now beginning to delineate the components of plant nuclear genomes, and to determine both their organization and their primary structural contributions. LTR-retrotransposons appear to be the major class of DNA in large-genome plants, but their contribution to a plant's physiology is minimized by their low levels of expression and their separation from genes. Mobile DNAs with low copy numbers appear to play a disproportionate role in gene mutation and the provision of raw material for the evolution of new functions because of their common insertion near genes. Rapid variation in the copy number of some gene families is another major route for the evolution of new gene specificities. The contributions of particular repeat sequences to chromosomal dynamics, including recombination, segregation, mutation, and replication, are now being investigated with molecular, genetic and cytogenetic tools.

Many questions remain to be answered. We still do not understand why some genomes have accumulated amplified retrotransposons and others have not. These questions can be approached with transgenic retrotransposon experiments and by more comprehensive phylogenetic characterizations. The fine details of local genome rearrangements, specifically their relative frequencies and sizes, remain to be determined. These studies will require extensive chromosome sequencing, and in more than just one or two plant species. To date, there is no case in plants where we understand the *trans*-acting factors (i.e. enzymes) that drive such outcomes of genome organization as telomere aggregation before prophase, centromere-driven chromosome synapsis or segregation, gene-specific recombination or transposon insertion, higher-dimensional chromosome folding, position effect, paramutation, or transgene-mediated suppression. Even with all these tasks before us, understanding these questions will provide only our first entry into understanding the relationship between genome organization and function. Subsequent generations of studies will investigate three-dimensional genome structure in the nucleus, with respect to both the dynamics of the cell cycle and the changes associated with environmental and developmental responses. The tools for these analyses are just now coming into existence, and the future holds enormous promise for the study of plant genome organization.

Acknowledgements

The writing of this manuscript, and much of the research described herein, was supported by grants from the United States Department of Agriculture (#94-37300-0299 and 94-37310-0661). The author thanks Sheri Frank for producing the drawing in Figure 1.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Meyer P, Saedler H: **Homology-dependent gene silencing in plants.** *Annu Rev Plant Physiol Plant Mol Biol* 1996, **47**:23-48.

2. Stam M, Mol J, Kooter J: **The silence of genes in transgenic plants.** *Ann Bot* 1997, **79**:3-12.

3. • Iglesias V, Moscone E, Papp I, Neuhuber F, Michalowski S, Phelan T, Spiker S, Matzke M, Matzke A: **Molecular and cytogenetic analysis of stably and unstably expressed transgene loci in tobacco.** *Plant Cell* 1997, **9**:1251-1264.

The first comprehensive study of the possible correlations between transgene expression and the chromosomal location, various genome features (matrix attachment regions [MARs], repeat types), and inserted structure of that transgene. Centromeric location and adjacent binary vector sequences were associated with unstable transgene expression, while telomeric location and nearby MARs were associated with stable expression. A model for future studies.

4. Flavell RB, Bennett MD, Smith JB, Smith DB: **Genome size and proportion of repeated nucleotide sequence in plants.** *Biochem Genet* 1974, **12**:257-269.

5. Lapitan NLV: **Organization and evolution of higher plant nuclear genomes.** *Genome* 1992, **35**:171-181.

6. •• SanMiguel P, Tikhonov A, Jin Y-K, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL: **Nested retrotransposons in the intergenic regions of the maize genome.** *Science* 1996, **274**:765-768.

This paper describes an unexpectedly simple organizational pattern and composition of the repetitive DNA blocks found between genes in maize. Fairly intact long terminal repeat (LTR)-retrotransposons are inserted within each other, exhibiting a preference for accumulation in these methylated, heterochromatic regions. The LTR-retrotransposons discovered are present in copy numbers ranging from a few to a few tens of thousands, and overall comprise over half of the maize genome.

7. SanMiguel P, Bennetzen JL: **Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons.** *Ann Bot* 1998, in press.

8. Gill KS, Gill BS, Endo TR, Taylor T: **Identification and high-density mapping of gene-rich regions in chromosome group 1 of wheat.** *Genetics* 1996, **144**:1883-1891.

9. Chen M, SanMiguel P, Bennetzen JL: **Sequence organization and conservation in *sh2/a1*-homologous regions of sorghum and rice.** *Genetics* 1998, **148**:435-443.

10. Bevan M, Bancroft I, Bent E, Love K, Goodman H, Dean C, Bergkamp R, Dirkse W, Van Steveran M, Stikema W *et al.*: **Analysis of 1.9 Mb of contiguous DNA sequence from chromosome 4 of *Arabidopsis thaliana*.** *Nature* 1998,**391**:485-488.

11. Avramova Z, Tikhonov A, SanMiguel P, Jin Y-K, Liu C, Woo S-S, Wing RA, Bennetzen JL: **Gene identification in a complex chromosomal continuum by local genomic cross-referencing.** *Plant J* 1996, **10**:1163-1168.

12. Harrison GE, Heslop-Harrison JS: **Centromeric repetitive DNA in the genus *Brassica*.** *Theor Appl Genet* 1995, **90**:157-165.

13. Edwards KJ, Veuskens J, Rawles H, Daly A, Bennetzen JL: **Characterization of four dispersed repetitive DNA sequences in *Zea mays* and their use in constructing contiguous DNA fragments using YAC clones.** *Genome* 1996, **39**:811-817.

14. Kaszas E, Birchler J: **Misdivision analysis of centromere structure in maize.** *EMBO J* 1996, **15**:5246-5255.

15. Yu H-G, Hiatt EN, Chan A, Sweeney M, Dawe RK: **Neocentromere-mediated chromosome movement in maize.** *J Cell Biol* 1997, **139**:831-840.

16. Jiang J, Nasuda S, Dong F, Scherrer CW, Woo S-S, Wing RA, Gill BS, Ward DC: **A conserved repetitive DNA element located in the centromeres of cereal chromosomes.** *Proc Natl Acad Sci USA* 1996, **93**:14210-14213.

17. Aragon-Alcaide L, Miller T, Schwarzacher T, Reader S, Moore G: **A cereal centromeric sequence.** *Chromosoma* 1996, **105**:261-268.

18. Dawe RK, Cande WZ: **Induction of centromeric activity in maize by *suppressor of meiotic drive 1*.** *Proc Natl Acad Sci USA* 1996, **93**:8512-8517.

19. Blackburn EH: **Structure and function of telomeres.** *Nature* 1991, **350**:569-573.

20. McClintock B: **The stability of broken ends of chromosomes in *Zea mays*.** *Genetics* 1941, **26**:279-282.

21. Tsujimoto H, Yamada T, Sasakuma T: **Molecular structure of a wheat chromosome end healed after gametocidal gene-induced breakage.** *Proc Natl Acad Sci USA* 1997, **94**:3140-3144.

22. Vershinin A, Schwarzacher T, Heslop-Harrison JS: **The large scale genomic organization of repetitive DNA families at the telomeres of rye chromosomes.** *Plant Cell* 1995, **7**:1823-1833.

23. Richards EJ, Goodman HM, Ausubel FM: **The centromere region of *Arabidopsis thaliana* chromosome 1 contains telomere-similar sequences.** *Nucleic Acids Res* 1991, **19**:3351-3357.

24. Presting GG, Frary A, Pillen K, Tanksley SD: **Telomere-homologous sequences occur near the centromeres of many tomato chromosomes.** *Mol Gen Genet* 1996, **251**:526-531.

25. •• Bass H, Marshall W, Sedat J, Agard D, Cande W: **Telomeres cluster de novo before the initiation of synapsis: a three-dimensional spatial analysis of telomere positions before and during meiotic prophase.** *J Cell Biol* 1997, **137**:5-18.

Advanced cytogenetics were used to characterize co-ordinated telomere movements prior to chromosome condensation in meiosis. This trail-breaking study begins scientific investigation into the three-dimensional dynamics of intranuclear components.

26. Cresse AD, Hulbert SH, Brown WE, Lucas JR, Bennetzen JL: ***Mu1*-related transposable elements of maize preferentially insert into low copy number DNA.** *Genetics* 1995, **140**:315-324.

27. Bureau TE, Wessler SR: **Mobile inverted-repeat elements of the *Tourist* family are associated with the genes of many cereal genomes.** *Proc Natl Acad Sci USA* 1994, **91**:1411-1415.

28. Wessler SR, Bureau TE, White SE: **LTR-retrotransposons and MITEs: important players in the evolution of plant genomes.** *Curr Opin Genet Dev* 1995, **5**:814-821.

29. • Avramova Z, Tikhonov A, Chen M, Bennetzen JL: **Matrix attachment regions and structural colinearity in the genomes of two grass species.** *Nucleic Acids Res* 1998,**26**:761-767.

Comparison of >30 kb of DNA sequence from orthologous regions of the rice and sorghum genomes indicates that only the three genes in the regions are conserved at a sequence level. MARs are mapped in the region, and found to place each gene in a separate putative chromatin loop. MARs map to homologous locations in each species, indicating conservation of a structural function that is not detected by primary sequence comparisons. MITEs are found to be preferentially associated with MARs.

30. Grandbastien M-A: **Retroelements in higher plants.** *Trends Genet* 1992, **8**:103-108.

31. • Bennetzen JL: **The contributions of retroelements to plant genome organization, function and evolution.** *Trends Microbiol* 1996, **4**:347-353.

Presents an overview of the status of retrotransposon studies in plants. A special section discusses the important discovery that some plant retrotransposons, like retroviruses of animals but unlike all other known retrotransposons, have the capacity to acquire and maintain a portion of a normal cellular gene (see [44,45]).

32. White SE, Habera LF, Wessler SR: **Retrotransposons in the flanking regions of normal plant genes: a role of copia-like elements in the evolution of gene structure and expression.** *Proc Natl Acad Sci USA* 1994, **91**:11792-11796.

33. Bonierbale MW, Plaisted RL, Tanksley SD: **RFLP maps based on a common set of clones reveals modes of chromosomal evolution in potato and tomato.** *Genetics* 1988, **120**:1287-1292.

34. Hulbert SH, Richter TE, Axtell JD, Bennetzen JL: **Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes.** *Proc Natl Acad Sci USA* 1990, **87**:4251-4255.

35. Moore G, Devos KM, Wang Z, Gale MD: **Grasses, line up and form a circle.** *Curr Biol* 1995, **5**:737-739.

36. Bennetzen JL, Freeling M: **The unified grass genome: synergy in synteny.** *Genome Res* 1997, **7**:301-306.

37. Paterson AH, Lan T-H, Reischmann KP, Chang C, Lin Y-R, Liu S-C, Burow MD, Kowalski SP, Katsar CS, DelMonte TA *et al.*: **Toward a unified genetic map of higher plants, transcending the monocot-dicot divergence.** *Nat Genet* 1996, **14**:380-382.

38. Dunford RP, Kurata N, Laurie DA, Money TA, Minobe Y, Moore G: **Conservation of fine scale DNA marker order in the genomes of rice and the Triticeae.** *Nucleic Acids Res* 1995, **23**:2724-2728.

39. Kilian A, Kudrna DA, Kleinhofs A, Yano M, Kurata N, Steffenson B, Sasaki T: **Rice-barley synteny and its application to saturation mapping of the barley *Rpg1* region.** *Nucleic Acids Res* 1995, **23**:2729-2733.

40. Chen M, SanMiguel P, Oliviera AC, Woo S-S, Zhang H, Wing RA, Bennetzen JL: **Microcolinearity in *sh2*-homologous regions of the maize, rice and sorghum genomes.** *Proc Natl Acad Sci USA* 1997, **94**:3431-3435.

41. Richter TE, Pryor AJ, Bennetzen JL, Hulbert SH: **New rust specificities associated with recombination in the *Rp1* complex in maize.** *Genetics* 1995, **141**:373-381.

42. Leister D, Kurth J, Laurie DA, Yano M, Sasaki T, Devos K, Graner A, Schulze-Lefert P: **Rapid reorganization of resistance gene homologues in cereal chromosomes.** *Proc Natl Acad Sci USA* 1998, **95**:370-375.

43. • Bennetzen JL, Kellogg EA: **Do plants have a one-way ticket to genomic obesity?** *Plant Cell* 1997, **9**:1509-1514.

Presents recent evidence regarding the nature and rate of evolutionary change in plant genome size. Raises a controversial possibility that changes may be unavoidably directed towards larger genomes.

44. Jin Y-K, Bennetzen JL: **Integration and nonrandom mutation of a plasma membrane proton ATPase gene fragment within the *Bs1* retroelement of maize.** *Plant Cell* 1994, **6**:1177-1186.

45. Bureau T, White SE, Wessler SR: **Transduction of a cellular gene by a plant retroelement.** *Cell* 1994, **77**:479-480.

46. • Hirochika H, Otsuki H, Yoshikawa M, Otsuki Y, Sugimoto K, Takeda S: **Autonomous transposition of the tobacco retrotransposon *Tto1* in rice.** *Plant Cell* 1996, **8**:725-734.

First demonstration that a retrotransposon from a dicot can function in a monocot, suggesting high levels of conservation of any necessary host factors.

47. Hollick J, Dorweiler J, Chandler V: **Paramutation and related allelic interactions.** *Trends Genet* 1997, **13**:302-307.

48. • Allen GC, Hall G, Michalowski S, Newman W, Spiker S, Weissinger AK, Thompson WF: **High-level transgene expression in plant cells: effects of a strong scaffold attachment region from tobacco.** *Plant Cell* 1996, **8**:899-913.

Expression of a transgene flanked by a strong scaffold attachment region (an SAR, similar to a matrix attachment region) was increased nearly 140-fold in stably transformed cells, but only threefold in transient expression assays, suggesting that the SAR fragments were not acting primarily as transcriptional enhancers. The SAR effect was greatest in cells that had a low transgene copy number. The authors propose that the major effect of SARs is to reduce the severity and/or likelihood of homology-dependent silencing in cells with low transgene copy numbers.

49. Avramova Z, SanMiguel P, Georgieva E, Bennetzen JL: **Matrix attachment regions and transcribed sequences within a long chromosomal continuum containing maize *Adh1*.** *Plant Cell* 1995, **7**:1667-1680.

50. • Van Drunen CM, Oosterling RW, Keultjes GM, Weisbeek PJ, van Driel R, Smeekens SCM: **Analysis of the chromatin domain organisation around the plastocyanin gene reveals an MAR-specific sequence element in *Arabidopsis thaliana*.** *Nucleic Acids Res* 1997, **25**:3904-3911.

Three MARs were found in a 16 kb region around the plastocyanin gene that contained seven total genes, as determined by complete DNA sequencing of the region. This high gene and MAR density predicted that there would be two genes per loop. The sequence uncovered a 21 bp element in all of the MARs, and in four other MARs characterized in *Arabidopsis*. This sequence, however, is not found in MARs from most plant species (see [29•]).