Comparative studies have provided the basis for some of the most important discoveries in biology. The study of differences, whether at the level of gene alleles or living kingdoms, has shown the critical features and function of most biological structures. The framework for comparative studies of organisms was perhaps laid out by the earliest taxonomists and medics: Dioscorides (80) and others used chemosystematic properties and morphology to group plants with similar medicinal properties. In the first half of the twentieth century, cytologists such as Darlington (1931; Darlington and LaCour, 1942), Kihara (1924), and Sears (1941) studied plant chromosomes, making key discoveries such as the importance of polyploidy in plant evolution and showing that whole chromosomes carried similar groups of genes in different species. Soon thereafter, an understanding of evolutionary changes at the cytological level was established that included an appreciation for chromosome translocation, fusion, fission, and the correlation between chromosomal particularities and phylogeny. Today, molecular markers show that gene orders are conserved over substantial evolutionary distances (Gebhardt et al., 1991; Ahn et al., 1993; Devos and Gale, 1993, 1997; see also Devos and Gale, 2000, in this issue), with the number of chromosomal and genetic differences between species generally increasing with evolutionary distance (Bennetzen et al., 1998; Tikhonov et al., 1999).

Comparative studies are useful in elucidating the function of biological structures and in providing markers for evolutionary investigation, whether in the context of plant breeding, ecology, or biodiversity. Evolutionary comparisons, moreover, encompassing the very origin of life, complement other genomic studies. As pointed out by Gesteland and Atkins (1993), many relics of the “RNA world,” existing 3.5 billion years ago, have been discovered in modern organisms, demonstrating the extraordinary conservation of nucleic acid sequence and function. Examples of such relics include ribosomes (where the RNA aggregate in the absence of proteins is able to synthesize peptide bonds; Nitta et al., 1998), ribozymes, and features of codon usage (see Heslop-Harrison, 2000). Comparative studies enable the application of data from one species to investigations of taxonomically disparate species, as exemplified in the use of Escherichia coli, human, Arabidopsis, and rice in understanding the wheat genome. The genome sequencing projects are in fact based on the premise that knowledge of the whole sequence of Arabidopsis, for instance, will aid in the isolation from crop plants of agronomically important genes that bear homology to Arabidopsis genes. Work from one kingdom often suggests models to test in another, although techniques easily applicable in one species may be impossible or inappropriate in another.

Many features of plant, animal, fungal, and even prokaryotic genomes are remarkably similar, but there are some elements that are not conserved (e.g., centromeres; see below). A notable feature of angiosperms is the widespread occurrence of polyploidy—even over experimentally observable time frames—involving either the doubling of chromosome numbers within a species or the interspecific union of chromosome sets. In contrast, recent and evolutionarily significant polyploidy is unusual in gymnosperms, most vertebrates, and well-studied species such as Caenorhabditis or Drosophila. One might suggest that the genes enabling regular meiosis with strict bivalent formation between homologous but not homeologous pairs of chromosomes are an early feature of angiosperm evolution. Such genes might conceivably control the stringency and timing of meiotic chromosome pairing as well as influence the disparate organization of repetitive DNA sequences in plants and animals.

THE LINEAR DNA SEQUENCE

With the first sequences of complete plant chromosomes now published (Lin et al., 1999; Mayer et al., 1999), it is...
appropriate to consider the relationship of linear sequence information to the organization and function of chromosomes within the context of the genome. DNA sequence data have averred the general model of the structure of the DNA component of the chromosome. Sequencing provides direct evidence that the double-helical DNA molecule is continuous from telomere to telomere, and there is no evidence for bases other than A, T, G, and C. These conclusions, although long anticipated, are not trivial, because the resolution from sequencing is magnitudes greater than can be observed by microscopy, and biochemical assays are hampered when handling DNA molecules of several million kilodals. It is also worth noting that the prokaryotic vectors and enzymes used in sequencing can process the angiosperm DNA in its entirety, confirming the universal nature of DNA.

Arabidopsis was chosen as the first plant target for complete sequencing because its genome size (130 to 140 Mbp) is rather small, ~200 times smaller than other plant genomes (see http://www.rbgkew.org.uk/cval/database1.html); the Arabidopsis genome is diploid, with five pairs of chromosomes (2n = 2x = 10). There is significant correlation between genome size and plant niche (e.g., Bennett et al., 1998), although this is not immediately obvious: rye (9000 Mbp genome size; 2n = 2x = 14) and Arabidopsis are short-lived, annual plants, whereas oak (200 Mbp) and pines (23,000 Mbp) are temperate trees. Admittedly, the angiosperms represent an unusually broad taxonomic spectrum, but among birds, a taxon that is also quite large, genome sizes vary by only a factor of two (from 2000 to 3800 Mbp; Tiersch and Wachtel, 1991).

The division of genomic DNA into independent chromosomes is a fundamental feature of genome architecture. Like genome size, chromosome number varies widely among plant species, such that 2n ranges in value from 4 to more than 1000, although the number within any given species, with the exception of supernumerary or B chromosomes, is usually constant. Some taxa, such as the family Cruciferae, have highly variable chromosome numbers, whereas the number is conserved in others. Polyploids tend to have higher chromosome numbers, and species in which n is a multiple of 6 or 7 are frequent. Nevertheless, except for polyploidy, there are few plant characteristics that correlate clearly with chromosome number. Of course, chromosome number has a genetic consequence in that genes are reassorted at meiosis on different chromosomes.

The placement of genes and their introns within the broader genomic context is an important area of research, involving detailed annotation of the genome (see www.tigr.org for the current status in Arabidopsis). Most of the genes found in species with much larger genomes are present in Arabidopsis, which is estimated to contain ~25,000 genes. But the smallness of the Arabidopsis genome means that other characteristic sequence regions, such as those that are highly repeated, are less abundant than in other species. DNA motifs, ranging in length from a single base to thousands of bases, repeated many hundreds or thousands of times, are a characteristic of all eukaryotic genomes and represent between 50 and 90% or more of all DNA. In Arabidopsis, many duplications of gene sequences have been found both within (e.g., ~250 tandem duplications each ~10 kb on chromosome 2) and between chromosomes (e.g., regions ~4 Mb long between chromosomes 2 and 4, or 700 kb long between chromosomes 1 and 2; Lin et al., 1999). Furthermore, a large part of the mitochondrial genome, ~270 kb, is inserted into chromosome 2. The transfer of genes from organelles to nucleus over evolutionary time is now well established in plants and animals (Martin and Herrmann, 1998; Vaughan et al., 1999).

To fully exploit sequence information from Arabidopsis, Caenorhabditis (C. elegans Sequencing Consortium, 1998), yeast (Goffea et al., 1996), and human chromosomes (Dunham et al., 1999), we must also appreciate what sequence data alone cannot tell us. Modified bases, and particularly 5-methylcytosine, are not distinguished from their unmodified equivalents, and there is no information about chromatin packaging and three-dimensional organization, topics that are essential to a complete understanding of the genome. There are, moreover, a few gaps at which DNA sequence is not available. In Arabidopsis, such regions occur at the centromeres and at nucleolar organizer regions in the two chromosomes sequenced to date, and there are more extensive gaps at 11 sites in human chromosome 22 (Dunham et al., 1999). Below, I discuss the localization of key sequence motifs along the plant chromosome, using Arabidopsis, the triticaceae cereals, and pine as examples to support a general model of sequence distribution along plant chromosomes (Schmidt and Heslop-Harrison, 1998).

**REPEITIVE DNA SEQUENCES AND THE LARGE-SCALE ORGANIZATION OF THE CHROMOSOME**

The regions that have remained inaccessible from the otherwise fully sequenced chromosomes consist, at least in the Arabidopsis chromosomes and most of human chromosome 22, of long and relatively homogeneous stretches of repetitive DNA motifs. Such stretches are not tractable by current technologies that read sequences of only a few hundred base pairs that must then be ordered so that they represent the complete chromosome. It is important to know about the length of the sequence gaps, the homogeneity of repeat motifs, and the level of variation within the motifs before one can begin to hypothesize how they evolve and function in the context of the genome. In Arabidopsis, multiple random fragments of individual bacterial artificial chromosomes, averaging 80 kb long, show the repetitive sequence to be homogeneous, with no interspersion of low-copy sequence among the tandem 180-bp repeats (Lin et al., 1999); however, some genes exist amid the repeated sequences, found in both forward and reverse orientations, in
the centromeric regions (Figure 1). The extent to which repeat motif variants relate to chromosomal function as opposed to satisfying some “bulk” requirement remains to be determined.

Of sequence motifs that are highly repeated, some are highly conserved from one species to another: the rRNA genes are present as hundreds of copies with only a small percentage of variation in all eukaryotes. Other sequence motifs are extremely variable, even between accessions of a species, providing tools to assess potential functions of particular aspects of genome architecture and for studying interorganismal relationships. Hence, the study of repetitive DNA sequence motifs and their chromosomal distribution in a comparative context—comparing, for instance, sequences from Arabidopsis and wheat (Figure 2) versus conifers and Crocus—has considerable potential for understanding genome evolution and sequence components. Specifically, individual sequences of a particular repeated motif may vary in both their copy number and exact sequence, giving rise to the concept of sequence families (see Figure 1). Various classes of repeated sequences (see below) are easily recognized: (1) tandemly repeated sequences, in which one copy follows another in an array of many tens or even thousands of copies; (2) retroelements, in which amplification occurs through an RNA intermediate (acting as a template for protein translation as well as DNA transcription) before reinsertion into the genome; and (3) those that are special classes such as telomeric sequences or rDNA units.

Cytogenetic methods offer a powerful system for looking at the organization of DNA repeat motifs along a chromosome using in situ hybridization of labeled probe sequences to the denatured DNA of chromosomes spread on microscope slides. The techniques are robust and reliable, with chromosomal target regions containing a few kilobases (even if dispersed over much longer chromosome segments) of sequence homologous to the probe (Schmidt and Heslop-Harrison, 1998; Schwarzacher and Heslop-Harrison, 2000). Sequences suspected to be abundant because of their frequent occurrence in a library or strength of membrane hybridization may prove difficult to interpret in terms of chromosome localization because hybridization probes to size-separated restriction fragments often give multiple dense bands or smears. Where sequence information is available, results are consistent with those from in situ hybridization. For example, the completed sequence of Arabidopsis chromosome 2 (Lin et al., 1999) confirms in situ hybridization data that had unexpectedly shown copia-like retroelements to be dispersed along the chromosome arms and clustered at centromeric regions (Figure 2; Brandes et al., 1997a). In polyploid species such as bread wheat, some sequences are much more abundant in one chromosome set than in another (Figure 2).

In situ hybridization methods also offer advantages in comparing different accessions or species. In addition, viral and mitochondrial sequences within the nucleus of various accessions can be located by in situ hybridization without knowledge about the nuclear flanking DNA. Analytical difficulties by in situ hybridization are encountered in relation to neither genome size nor repetition of a sequence motif.

**Figure 1.** Tandem Repetition of a 180-bp Motif in Arabidopsis Chromosome 2.

The dot plot shows the tandem repetition of a 180-bp motif within 14 kb of sequence around the centromeric region (sequence data from Lin et al., 1999; see www.tigr.org). A dot is printed whenever 90% of bases within a window of 30 as numbered along the abscissa match those numbered on the ordinate. The continuous diagonal shows that the sequences along the two axes are identical, whereas the many other diagonal lines show that particular motifs, mostly the 180-bp motif, are repeated in tandem at multiple sites. Diagonals that represent obtuse angles relative to the abscissa indicate forward repeats; diagonal hits representing acute angles represent reversed complementary sequences. The repeat motifs are uniform in length (parallel lines have equal spacing), but repeat blocks vary in length.

**rDNA**

The 45S rDNA loci consist of tandem arrays of repeating units of the 18S, 5.8S, and 26S rRNA genes and the transcribed and nontranscribed spacers, each unit being typically 10 kb long in plants. Hundreds or thousands of copies of the repeat units may be present, together representing up to ~10% of the genome (8% in Arabidopsis; Pruitt and Meyerowitz, 1986). The units, along with the SS rRNA genes (occurring as tandem repeats independent of the 45S rDNA), are localized at one or more sites per chromosome set, and their characteristic positions along chromosomes provide useful markers for chromosome identification (see,
Figure 2. In Situ Hybridization of Metaphase Chromosomes and Interphase Nuclei from Various Species.
Telomeres

Telomeres are specialized structures that stabilize chromosome ends and enable replication (see Zakian, 1995). The telomeric region is highly conserved and consists of a short repeat, the sequence of which is similar to TTAGGG, in tandem arrays many hundreds of units long at the physical ends of chromosomes in most eukaryotes (Drosophila is a notable exception; see Fuchs et al., 1995). Unlike most chromosomal DNA, terminal sequences cannot be fully replicated by a semiconservative mechanism but rather require the enzyme telomerase to supply an RNA template at the DNA terminus. The number of telomeric repeats is a species-specific characteristic, equivalent to 2 to 5 kb in Arabidopsis (Richards and Ausubel, 1988), 12 to 15 kb in cereals (Figure 2; see also Schwarzacher and Heslop-Harrison, 1990), and up to 60 to 160 kb in tobacco (Fajkus et al., 1995). The number of copies of the repeat also differs among the chromosome arms of the karyotype (Figure 2; Schwarzacher and Heslop-Harrison, 1990) and possibly varies from cell to cell and tissue to tissue (Kilian et al., 1995). Through its ability to attach the telomeric sequences to new chromosomal ends, telomerase also provides a mechanism to stabilize and repair broken chromosomes (Wang et al., 1992). In a sugar beet (Beta vulgaris) hybrid line incorporating an alien chromosome fragment from Beta vulgaris, telomeric sequences were detectable by in situ hybridization on all chromosome ends except one terminus of the alien fragment. Perhaps the particular conformation of the DNA at this end precludes the action of telomerase and thereby leads to lack of stability of the line (Schmidt et al., 1997).

Subtelomeric repetitive sequences have often been revealed by staining patterns of chromosomes. Analysis of these sequences on rye chromosomes shows that they are able to evolve in copy number rapidly (Alkhimova et al., 1999) and may be part of a complex chromosome end structure (Vershinin et al., 1995; Figure 2). Zhong et al. (1998) have used in situ hybridization to show that each chromosome end in tomato has a unique organization of the telomeric and a particular subtelomeric repeat, with large differences in lengths of each array. On chromosome 4 of Arabidopsis, the tandem repeats of the rDNA abut the telomeric repeats with <500 bp intervening (Copenhaver and Pikaard, 1996). Using the telomeric sequence to probe discrete chromosomal fragments resolved by pulsed-field gel electrophoresis, Ganai et al. (1992) were able to determine the genetic ends of chromosomes and hence show the complete map of tomato in terms of centimorgans.

Centromeres

During mitosis and meiosis, chromosomal segregation depends on the attachment of microtubules, however indirectly,
The sites differ in location, size, and order in the six genomes. The variations do not always reflect those found by genetic mapping of other molecular markers that may show greater conserved synteny among the species. No inversions have been detected among wheat, rye, and barley, although the rDNA genes show both orders on group 1 chromosomes. The rDNA sites provide useful markers for following the evolution of cereal chromosomes and also assist in the identification of individual chromosomes.

Figure 3. Diagrammatic representation of 45S and 5S rDNA on chromosomes from Groups 1 and 5 in various Triticeae.

Transposable Elements and Retroelements

Retroelements (class I transposable elements) are discrete components of the plant nuclear genome that replicate and reinsert at multiple sites in a complex process that involves activation of excision, DNA-dependent RNA transcription, translation of the RNA into functional proteins, RNA-dependent DNA synthesis (reverse transcription), and reintegration of newly generated retroelement copies into the genome (reviewed in Kumar and Bennetzen, 1999). Major classes of retroelements include LINEs, SINEs, copia- and gypsy-like elements, and retroviruses (Hull and Covey, 1996; Kumar,
of tandemly repeated DNA elements, show a lower proportion frequently, the rDNA and centromeric regions, consisting of associate with particular genomic regions (Figure 2). Most somes, consistent with their mode of amplification, but may associate with particular genomic regions (Figure 2). Most frequently, the rDNA and centromeric regions, consisting of tandemly repeated DNA elements, show a lower proportion of gypsy- and copia-like retroelements than do other regions (Kamm et al., 1996; Heslop-Harrison et al., 1997; Kubis et al., 1998a; Schmidt, 1999). It is hypothesized that retroelements are more abundant around the centromeres of Arabidopsis chromosomes so as to limit the disruption of genes (Figure 2; Brandes et al., 1997a). Relatively little is known about the chromosomal organization of LINEs (Kubis et al., 1998b).

As they insert themselves into the genome, retroelements act as mutagenic agents, thereby providing a putative source of biodiversity (Hirochika et al., 1996; Heslop-Harrison et al., 1997; Ellis et al., 1998; Flavell et al., 1998) and serving as markers of diversity. Regulatory mechanisms may act to protect genomes from insertional mutagenesis (Lucas et al., 1995), and it has been suggested that transgene-induced gene silencing reflects mechanisms aiming to prevent genome invasion by retroelements. Plant retrotransposon activity can be regulated at any step of the replication cycle, including transcription, translation, reverse transcription, nuclear import, and integration. Along with DNA (class II) transposable elements and other elements such as miniature inverted tandem elements (MITEs; Wessler et al., 1995; Casacuberta et al., 1998), insertion of retrotransposon elements can inactivate or alter gene function (Wessler et al., 1995). Indeed, transposition is estimated to account for 80% of the mutations detected in Drosophila (Capy, 1998). Transposons can excise, partially or completely restoring gene function, and can also lead to chromosome rearrangements such as inversions or translocations. Transposable elements can also act to move elements such as exons and promoters into existing sequences so as to create new gene functions and contribute to evolution (Plasterck, 1998; Moran et al., 1999). Indeed, retroelements are activated under stress conditions (Wessler, 1996; Grandbastien, 1998; Kumar and Bennetzen, 1999; Walbot, 1999). Alternative splicing of genes caused by transposable elements has been shown in maize (Bureau and Wessler, 1994a, 1994b). Methylation of retroelements can also affect adjacent sequences and lead to transcriptional repression (Yoder et al., 1997; Goubely et al., 1999).

The sequences of degenerate and potentially active retroelements give valuable data about genome evolution and phylogenetic relationships (Figure 4). In three species in the Vicia genus, copia retroelement copy number varies from 1000 to 1,000,000, with more sequence heterogeneity being present in species with higher copy number (Pearce et al., 1996). Although in part due to random mutation of the high number of copies present in most plant genomes, sequence variability is often nonuniformly distributed along the retroelement: regulatory regions (including the long terminal repeats of copia elements) can evolve faster than coding regions, perhaps enabling elements to coexist with their host genomes without detriment (Vernhettes et al., 1998). Although retroelement amplification leads to large genomes (Bennetzen and Kellogg, 1997), it is probable that retroelement turnover and loss can occur in a directed manner (Tatout et al., 1998), leading to different retroelement compositions between species. For example, chromosome sets in the cultivated hexaploid oat, Avena sativa, can be discriminated by the presence of retroelement families (Katsiotis et al., 1996).

Simple Sequence Repeats (Microsatellites)

Runs of single nucleotides or motifs of up to ~5 bp, described as microsatellites or simple sequence repeats (SSRs), are ubiquitous elements of eukaryotic genomes (Tautz and Renz, 1984). Genetic mapping using microsatellites as markers involves amplification of repeat arrays by

Figure 4. A Phylogenetic Tree (Clustal Method) According to Repeat Sequences.

Fourteen plant species and Drosophila are arranged in accordance with genomic representation of copia group retroelements (Genbank EMBL database). The copia elements are dispersed along the chromosomes (see Brandes et al., 1997a; see also Figure 2B), consistent with their mode of amplification through an RNA intermediate. Units (bottom) indicate the number of substitution events over ~260 bp.
the polymerase chain reaction with primers flanking the arrays. SSRs also provide highly informative and polymorphic markers for plant, fungal, and animal fingerprinting (Weising et al., 1991). Synthetic oligonucleotide SSRs have been used for in situ hybridization to chromosomes, revealing that microsatellite sequences vary widely with regard to genomic organization, raising implications for amplification and dispersion mechanisms and hence evolution. In some cases, synthetic SSRs have been used to detect sites within previously characterized repeat motifs. For example, a tandemly repeated motif near the centromeres of all 16 pairs of sugar beet chromosomes includes an (AC)$_n$ motif (Schmidt and Heslop-Harrison, 1996), and the polypurine motif (GAA)$_n$ has been correlated with the positions of C-bands in barley (Pedersen and Linde-Laursen, 1994). Notably, although conventional staining systems give very different chromosome banding patterns in wheat and rye, the hybridization pattern of the motif GACA with some 40 amplified sites is very similar in the two species, suggesting that the pattern was established before their evolutionary separation (Cuadrado and Schwarzer, 1998). In the human genome, changes in copy number of different microsatellite classes may occur through interallelic replication slippage of AT-rich sequences or complex, conversion-like events of GC-rich regions, with recombination in DNA flanking the repeat array (Bois and Jeffreys, 1999).

**Tandem Arrays of Repetitive DNA**

Many repetitive sequence motifs occur as tandem repeats at a number of discrete sites—typically between one and 30—in the genome. Using in situ hybridization, these tandem repeats can provide useful markers for chromosome identification, and their presence and distribution can reveal evolutionary changes (Figure 2; Kubis et al., 1997). Both the site distribution and sequence of tandemly repeated sequences may show polymorphism between species and accessions of a species. However, the evolution of tandem repeats does not show characteristics of a “molecular clock” with a constant mutation rate. All evidence points to its occurrence in bursts or evolutionary waves, perhaps occurring during periods of rapid speciation or stress.

In many species, the distribution of different repetitive DNA sequences closely follows their taxonomic relationships: eight different sequences isolated from Beta spp can be used to elucidate the relationships between the four related sections of the genus (Schmidt and Heslop-Harrison, 1994). In contrast, taxonomy within the genus Crocus shows little correlation with the distribution of repetitive sequence, reflecting not only a disparity between taxonomy and actual phylogeny but also the explosive speciation occurring at one evolutionary period (Frello and Heslop-Harrison, 2000).

A family of repetitive sequences originally isolated from rye, named pSc119.2 (Bedbrook et al., 1980), is abundant in all species of the tribe Triticeae, and even in related tribes such as Avenae, but absent from cultivated barley and close relatives. Because it is likely that the sequence was present in the common ancestor of the Triticeae tribe, its absence from barley implies that high-copy sequences may be superfluous to the genome and again suggests there is no molecular clock to gauge evolution. In rye itself, more distal subtelomeric sequences, pSc200 and pSc250, are relatively species specific (Vershinin et al., 1995) and have presumably evolved more recently.

Tandem repeats are normally regarded as transcriptionally silent (Radic et al., 1987), although a significant proportion of RNA in rice has been shown to represent a particular subtelomeric tandem repeat (Wu et al., 1994). It is possible that such RNA is due to read-through transcription in which a stop codon is ignored, which might occur more frequently under stressful conditions.

Frequently, unequal crossover and recombination of chromosome strands within the tandem arrays are considered to be involved in the evolution and amplification of repeat units (Dover, 1982; Charlesworth et al., 1994). McAllister and Werren (1999) have presented experimental evidence for the unequal crossover model and also suggest how turnover of repeats allows migration of retroelements toward the ends of arrays. In yeast, Paques et al. (1998) conclude that the expansion and contraction mechanisms for tandem arrays have their origin in DNA repair rather than genome replication mechanisms. It is also evident that genome-scanning mechanisms can homogenize different units of a tandem repeat, making all sequences identical. Much work showing homogenization has been performed on rDNA repeat units, and it is possible that similar mechanisms may act in other repeats. Schlotterer and Tautz (1992) have shown that intrachromosomal homogenization occurs rapidly in Drosophila rDNA, whereas interchromosomal homogenization occurs at a slower rate. In cotton, a tetraploid species, Wendel et al. (1995) have shown that the rDNA has become homogenized to resemble the variant found in only one of the ancestral Gossypium spp.

**DNA SEQUENCE IN THE CHROMOSOME**

Within the nucleus, DNA is modified by the addition of methyl groups, and most DNA is wrapped around histone proteins, forming nucleosomes and the 30-nm fiber as the fundamental structural subunit of chromosomes (Manuelidis and Chen, 1990; Wolffe, 1995). Higher levels of packaging, often very dynamic, result in chromatin fibers such that varying chromatin density is seen within the three dimensions of the interphase nuclei (Figure 5) as metaphase chromosomes appear. The packing of the genomic DNA can directly affect aspects of RNA transcription, DNA replication, recombination, DNA repair, and chromosome segregation (Cremer et al., 1993; Heslop-Harrison et al., 1993).
In plants, as well as in most prokaryotes and animals (except for Drosophila), modification of DNA by cytosine methylation is extensive (Finnegan et al., 1996): ~80% of cytosines in CG dinucleotides are modified (Gruenbaum et al., 1981). Plants, like animals, may contain unmethylated CG-rich regions (CpG islands) related to transcriptionally active genes (Antequera and Bird, 1993), and extensive evidence suggests that methylation is a mechanism for regulating gene expression. Numerous reports have correlated hypermethylation near genes, or in gene promoters, with reduced levels of gene expression (Barlow, 1993; Razin and Cedar, 1993; Sardana et al., 1993; Neves et al., 1995; Finnegan et al., 1996). Repression occurs at the level of transcription initiation (Tate and Bird, 1993), although methylation does not seem to repress the activity of all genes, including those borne by transposons (Martienssen, 1998). Many DNA methylation patterns are established during ontogeny and may remain stable through later development (Jahnner and Jaenisch, 1984; Razin and Cedar, 1993; Neves et al., 1997). Studies of floral homeotic mutants (Finnegan et al., 1996; Ronemus et al., 1996) suggest a direct correlation between DNA methylation and normal regulation of developmentally important genes (Jactobsen and Meyerowitz, 1997).

In animals, most methylation seems to occur at symmetrical sites in the DNA molecule, where the nucleotide combinations CG or CNG (N is any nucleotide) occur on both DNA strands. After DNA replication, methylation patterns are copied by maintenance methylases that respond to the methylation status of diagonally opposite Cs in the newly replicated DNA strand. In plants, it appears that methylation does not always occur at symmetrical positions (Fulnecek et al., 1998; Goubely et al., 1999); methylation sites must be established de novo after each replication cycle, perhaps by a DNA–DNA (Matzke et al., 1994) or RNA–DNA (Pelissier et al., 1999) pairing process. Wasseneeger et al. (1994) indicate that overexpressed mRNAs might direct sequence-specific de novo methylation of the DNA template and thus regulate gene activity. Such mechanisms may be involved in gene-silencing phenomena.

DNA methylation usually represents a terminal stage of differentiation but may be modulated, as is apparent by the activation in tissue culture of previously inactive retroelements (Grandbastien, 1998). Some methylation patterns change during plant development, particularly through meiosis (Silva et al., 1995) and embryogenesis (Castilho et al., 1999). Progressive reduction in methylation levels can occur upon DNA replication so as to result in hemimethylated and subsequently unmethylated DNA in daughter nuclei (Matzke et al., 1989; Kilby et al., 1992; J eddeloh et al., 1998). For the experimental reduction of DNA methylation, the cytosine analog 5-azacytidine, with a nitrogen atom rather than carbon atom at the 5-position of the pyrimidine ring, has revealed that reduced methylation of tandem DNA repeats in tobacco is maintained during protoplasting and plant regeneration (Bezdek et al., 1991; Koukalova et al., 1994).

Henikoff and Comai (1998) have found that Arabidopsis, like mouse and pea, has multiple methyltransferase specificities, probably resulting from multiple genes, and certain specificities may be tissue specific. In pea, methylase activities that recognize CG and CWG (where W is A or T) probably arise from the post-translational modification of a single gene product (Pradhan and Adams, 1995; Pradhan et al., 1995). Different enzymes are most likely to be involved in methylation of asymmetrical sites as opposed to maintenance of methylation of symmetrical sites (Goubely et al., 1999). In mammals, a CG demethylase has been identified (Bhattacharya et al., 1999), revealing a new mechanism of gene regulation presumably also present in plants.

Smith (1998) has suggested that the function of DNA methyltransferases and DNA methylation is in maintenance of eukaryotic chromosome stability. DNA methyltransferases participate in DNA repair complexes and also stabilize nucleoprotein assemblies required in the inactivation and imprinting of chromosomes. Methyltransferases may incorporate a
chromodomain, a protein module that mediates interactions between key chromatin proteins (Henikoff and Comai, 1998). Antibodies to methylcytosine have shown that different regions of chromosomes have different levels of methylation both in humans (De Capoa et al., 1995) and in plants (Figure 6; Frediani et al., 1996; Oakeley et al., 1997; Siroky et al., 1998; Castilho et al., 1999).

Structure and Packaging of Linear DNA into Chromosomes

The DNA double helix is wrapped around histone core particles, with ~146 bp of DNA forming the two turns around each nucleosome. Nucleosomes are connected by linker DNA, typically 20 to 35 bp long. Using micrococcal nuclease to cleave DNA in the linker region between nucleosomal core particles (Figure 7), it has become clear that chromatin higher-order structures and nucleosomal organization are not homogeneous along chromosomes (Fischer et al., 1994; Wolffe and Pruss, 1996) and that the dynamic chromatin structure found in animal systems applies also to plants. For example, Vershinin and Heslop-Harrison (1998) have shown small but significant variation in the nucleosomal organization and linker DNA length between telomeric DNA and various repetitive DNA sequence motifs in the bulk chromatin of rye, wheat, and their relatives. Furthermore, differences in linker DNA length and the sensitivity of cereal chromatin to micrococcal nuclease were observed in rye and wheat despite their relatively close taxonomic relationships.

Repetitive sequences, in particular tandem arrays, probably play a key role in stabilizing DNA packaging and higher-order chromatin condensation. Repetitive DNA motifs usually show a strictly defined arrangement (phasing) around nucleosomes. Gazdova et al. (1995) determined the position of the nucleosomal core next to a 10- to 11-bp AT track in a monomer of a tobacco tandem repeat, and Vershinin and Heslop-Harrison (1998) showed the defined phasing of tandem repeat motifs of 120, 360, and 550 bp. Nucleotide base stacking and twisting angles have been derived by Calladine et al. (1988) and provide the basis for predicting natural curvature of DNA molecules. Radic et al. (1987) hypothesized that bends in satellite DNA represent an essential structural signal for complete heterochromatin condensation. Repeated tracts of four to six adenines in phase with the helix produce bends, and bent DNA preferentially assembles into nucleosomes. After the packing of the repeats, the small proportion of single-copy DNA, regardless of its natural curvature preferences, can be fitted.

The frequent occurrence of sequence motifs ~180 bp long, or multiples of this length, indicates that the natural fit of the DNA molecule to the nucleosome core may be an important feature with respect to selection of lengths of repetitive DNA motifs. Breakage is observed in a small percentage of metaphase chromosomes and is often enhanced in divisions in interspecific hybrids: one might speculate that the poor fit of linkers between nucleosomes increases the breakage frequency, and repair mechanisms may be less efficient in the hybrid background.

Chromatin Remodeling and Histone Acetylation

Along with DNA methylation, chromatin remodeling and histone acetylation have been implicated in the modification of gene transcription (Martienssen and Henikoff, 1999). Histone acetylation per se may both change the relative positions of nucleosomes and influence the structure of chromatin (Turner, 1991). Chromatin remodeling involves specific enzymes affecting nucleosome structure and posi-
tioning along the DNA molecule (Cairns, 1998). Tazi and Bird (1990) have suggested that DNA methylation silences transcription through assembly of a repressive nucleosomal array. Wade et al. (1997) have suggested that nucleosome positioning may be critical in regulating the rate of transcription by modulating access and procession rate of the polymerase complex. Furthermore, methylation could suppress gene expression through an indirect mechanism affecting chromatin structure (Kass et al., 1997; Bergman and Mostoslavsky, 1998). Methylation may also mediate interaction between transposon sequences and chromatin factors, which conceal the sequences from the rest of the genome (Kass et al., 1997). The DDM1 gene in Arabidopsis causes rapid hypomethylation of repetitive DNA (Vongs et al., 1993; Kakutani et al., 1995, 1996), followed by hypomethylation of genes over many plant generations (J eddeloh et al., 1998). Sequence analysis indicates that the DDM1 protein does not have a direct role as a methyltransferase but rather modifies the accessibility of chromatin to methylation (J eddeloh et al., 1999). The DDM1 gene is similar to sequences from animals and fungi, which act to modify or disrupt protein-DNA interactions of multiprotein complexes that include the DDM1-like component (J eddeloh et al., 1999). Thus, the DDM1 protein probably functions in the DNA methylation system by affecting chromatin structure, perhaps by directing certain sequences to the methylation machinery or by modulating nucleosome remodeling.

Chromatin remodeling might have an ancient origin in the modulation of genome organization and may be a general requirement for replication of condensed, inactive regions of the genome. In vertebrates, it is well known that methylation of CG dinucleotides correlates with alterations in chromatin structure and gene silencing (Antequera et al., 1990; Antequera and Bird, 1993). Prymakowska-Bosak et al. (1996) have argued that genes involved in basal cellular functions are probably influenced relatively little by alterations in chromatin structure, whereas genes involved in specific developmental programs are likely to be regulated by factors related to chromatin constitution. The classic phenomenon of position-effect variegation in Drosophila occurs as chromosomes become heterochromatic (see Henikoff et al., 1993), and related phenomena have been associated with pea (Kass and Adams, 1993; S abl and Henikoff, 1996). Johnson et al. (1995) have established interrelationships between gene transcription and methylation and DNA packaging.

Local chromatin structure and its modification in early meiosis are important in the positioning and frequency of meiotic double-strand breaks in DNA that enable recombination in yeast (Ohta et al., 1994; Wu and Lichten, 1994). Earlier studies (Chandley and McBeath, 1987; Raman and Nanda, 1986) had also discussed that the regions of the human genome where the chromatin undergoes conformational changes from mitosis to meiosis could encompass recombinational hot spots. The lack of condensation of early replicating chromosomal segments during premeiotic interphase could be a prerequisite for crossover at pachytene.

**THE THREE-DIMENSIONAL NUCLEUS**

**Genome Architecture**

Genome architecture refers to the structural organization of the plant genome in the three-dimensional nucleus and can be extended to describe its dynamics and the relationship between structure and function. Cockell and Gasser (1999) concur with the emerging view that gene regulation cannot be fully explained by linear, two-dimensional models involving merely the binding of factors to regulatory elements. It has been widely suggested that nuclear architecture is related directly to the control of gene expression and that the multiple levels of organization of the chromatin provide functional regulation of DNA behavior. DNA packing and unpacking, replication, repair, mutation, and transcription are all regarded as cell-type specific aspects of a dynamic architecture. The scientific literature is now full of direct and

---

**Figure 7. Nucleosomal Structure of Rye Chromatin.**

Micrococcal nuclease digestion of extracted chromatin followed by size separation by agarose-gel electrophoresis and probing with a telomeric sequence results in a ladder of discrete bands that changes with the course of the digestion reaction. The bands differ in increments of ~170 bp. See Vershinin and Heslop-Harrison (1998) for more details.
indirect acknowledgment of the importance of nuclear architecture, including unpredictable positional effects and rearrangements. However, much of the literature about nuclear architecture and chromatin structure is based on mammalian, insect, or yeast models, often using cultured or model cell types such as fibroblasts or Drosophila polytene nuclei.

Electron micrographs show that DNA is largely condensed in plant interphase nuclei and that this condensed interphase chromatin is similar in appearance to chromosomes (Figure 5); the chromatin of cereals is largely condensed even in interphase nuclei (Muller et al., 1980). Measurement of chromosome volume, although inaccurate because of edge effects, indicates that volumes are similar in G2 interphase nuclei compared with mitotic chromosomes (Heslop-Harrison et al., 1988). Thus, little nuclear DNA is truly “decondensed.”

Packaging of Nuclear DNA

The traditional twentieth-century view of the nucleus as an unstructured jumble of spaghetti-like chromatin fibers is largely discounted, and most researchers agree that there are intranuclear frameworks that provide the dynamic genome with functional organization. Various levels of intranuclear compartmentalization can be regarded: individual chromosomes (Figure 2), euchromatic and heterochromatic regions, the nucleolus (Figure 5), and regions of active RNA synthesis and processing. Furthermore, telomeres and centromeres may be attached to or closely adjacent to the nuclear envelope (Schwarzacher and Heslop-Harrison, 1990; Rawlins et al., 1991) and occupy defined parts (poles) of the nucleus in many species (Rabl, 1885; Cremer et al., 1982; Anamthawat-Jónsson and Heslop-Harrison, 1990).

Cook (1997) has argued that each chromosome in a haploid set has a unique array of transcription units strung along its length and that chromatin fibers will therefore be folded into unique arrays of loops, with homologs sharing similar arrays. At meiosis, homologous chromosomes come together; this occurs when they are transcriptionally active, so that pairing may be an inevitable consequence of the transcription of partially condensed chromosomes (Cook, 1997). Similarly, Karpen et al. (1996) proposed that DNA–protein structures inherent to heterochromatin in Drosophila could produce a self-complementary chromosome “landscape” that ensures partner recognition and alignment by “best-fit” mechanisms. Specific coiling patterns that could promote pairing, showing apparent denser and weaker zones presumably reflecting more or less condensed chromatin, were observed at stages before meiotic prophase in the homologous chromosome domains of wheat (Figure 1A in Schwarzacher, 1997).

The existence of a nuclear matrix, or chromosomal skeleton, or both, following models for the cytoskeleton, remains controversial. Numerous papers describe features that appear under conditions that are far from the in vivo situation so that the relevance of such nuclear scaffolds, matrices, cages, and compartments remains questionable. As with any responsive and precisely regulated system, even small changes in hydration, ion concentration, and toxicity during experimentation are certain to have major effects on structure (Jackson and Cook, 1995). There are, moreover, complex controls on traffic between cytoplasm and nucleus (Jackson and Cook, 1995). Most of the major cytoskeletal proteins, including tubulins and actin, have been found within the nucleus, but their function and significance are unclear. It is accepted that the lamins (intermediate filaments) have a key role at the periphery of the nucleus and also extend deep into the nuclear volume. The nuclear pore complex and its dynamics have been worked out in detail (Allen et al., 1998; Goldberg et al., 1999), and it is clear that proteins permeate several microns from the pore complex into the nucleus in yeast, insects, and vertebrates. Advanced microscopic methods and antibody technology show that there are unexpected subnuclear localization patterns of many proteins (including transcription factors) in Arabidopsis and other species.

The higher-order structure of the chromatin fiber and the organization of chromatin domains in the nucleus appear to have a profound influence on gene expression. Good evidence exists that in most interphase nuclei, individual chromosomes occupy discrete domains, but the internal structure of these territories and the relation of their organization to presumptive higher-order functional compartments are difficult to investigate (Heslop-Harrison and Bennett, 1990; Sadoni et al., 1999). However, there is reasonable evidence in mammals that active genes tend to locate on the periphery of the territories, where RNA transcripts are formed. The chromosome domains may be elongated or subspherical, and DNA fibers may stretch away—possibly many microns—from the surface of the domain.

Although perhaps not a model for all aspects of nuclear behavior, incontrovertible evidence for nuclear compartmentalization is provided by the nucleoli. Nucleoli are subspherical compartments in the nucleus; there are no defined boundaries to nucleoli, although their composition is very different from the rest of the nucleus, and they move and fuse during interphase of the cell cycle. Soon after cell division in most species, multiple nucleoli (each originating from one rRNA locus) are the norm, but they often fuse to a smaller number during development of the cell. They are located at different positions within the nucleus, depending on cell type: peripheral and very close to the nuclear envelope in pollen mother cells at early meiotic prophase, but more central in other cell types.

Genomics, Chromosomes, Evolution, and the Nucleus

Genomes evolve at the level of the chromosome, chromosome segment, gene, and DNA sequence. Biotechnologists
and plant breeders aim to control and direct evolution, although limiting the impact of experimentally imposed genome evolution is an objective for the conservation of biodiversity and the environment. As Capy (1998) has stated, studies of the molecular basis of genome evolution are still young, but identification of the many processes in genome evolution, from molecular events to population dynamics, shows the impressive plasticity of the genome and the rapid amplification and fixation of advantageous novelties. An understanding of the functional and genetic bases of the major sources of variation at the genomic level (including retroelements; Kumar and Bennetzen, 1999) will have important applications. An appreciation of the types of changes that have occurred during species evolution will enable us to understand what can be done in plant breeding with respect to the changing environment.

Chromosome organization has a fundamental influence on processes as diverse as chromosome pairing, segregation, gene organization, and expression and has a direct impact on the aims of plant breeders in understanding genome evolution and genetics. The current model of the chromosome in the nucleus (Figure 8) is very different from that of five years ago. We now have complete sequences of chromosomes, and we can build a picture of the organization of

![Figure 8](image-url)

**Figure 8.** A Model of the Sequence and Structural Organization of Plant Chromosomes. See the text for details.
the different sequence motif types, each with characteristic locations along the linear structure (Schmidt and Heslop-Harrison, 1998). We now know about tandem repeats in relation to chromosome structure at telomeres and centromeres, about the abundance of intercalary tandem repeats, and in plant species with larger genomes than Arabidopsis, about largely dispersed retroelements that represent 50% of the DNA in the genome. The observed organization may be a consequence of the varying constraints on the evolution of each sequence motif, with many evolutionary changes reflective of the amplification or depletion of sequence motifs at particular genomic locations. The current chromosome model will be helpful for designing strategies for gene cloning, evolutionary studies, gene transfer, and perhaps gene expression.

The functional interphase nucleus is divided into compartments that are dynamic with respect to changes in chromosome structure and locations of genomic components (Figure 8). Subdomains of interphase chromosome territories are related to the different domains of mitotic chromosomes (Zink et al., 1998). We can suggest that there is a proteinaceous framework to the interphase nucleus, with many enzymatic and structural processes involved in chromatin modification and packaging. Together, these structures undoubtedly contribute to the observed and complex interactions between different genes and their regulation and inheritance of expression patterns. Nevertheless, this interphase model is based largely on vertebrate and insect systems; it must be treated with caution because the linearly compartmentalized mammalian chromosome (Holmquist, 1992; Saccone et al., 1993) shows differences in evolution and structure from the plant metaphase chromosome, as revealed by banding and molecular cytogenetics methods (Figure 8). There are parallels between plants and animals in chromatin modification and packaging in the nucleus (Figure 8, bottom), but details of the modulation and the significance of these structures remain to be studied.

Bennetzen et al. (1998) has shown that knowledge of the structure of the (grass) genome is useful in gene discovery and isolation. Comparative analysis of different species over narrow as well as wider taxonomic ranges has been instrumental in understanding genome organization in general. Promising technological developments to investigate the nucleus include in silico analyses, high-throughput DNA and protein studies, digital imaging methods, and in vivo methods to study dynamic chromosome structures and interactions. Such technologies will ensure that advances in understanding the behavior, function, and evolution of the chromosome in the nucleus will come at an increasing rate.

ACKNOWLEDGMENTS

I am particularly grateful to Trude Schwarzacher for her close involvement with the manuscript, and I thank my many collaborators, cited below, including Thomas Schmidt, Alexander Vershinin, and those in the European Community (EC) projects TEBIODIV and ANACONGEN. I thank EC Framework IV projects, the Palm Oil Research Institute of Malaysia, and the Biotechnology and Biological Sciences Research Council for funding.

REFERENCES


Comparative Genome Organization in Plants: From Sequence and Markers to Chromatin and Chromosomes

J. S. Heslop-Harrison

*Plant Cell* 2000;12;617-635

DOI 10.1105/tpc.12.5.617

This information is current as of July 12, 2017

| References | This article cites 177 articles, 50 of which can be accessed free at: /content/12/5/617.full.html#ref-list-1 |
| eTOCs | Sign up for eTOCs at: http://www.plantcell.org/cgi/alerts/ctmain |
| CiteTrack Alerts | Sign up for CiteTrack Alerts at: http://www.plantcell.org/cgi/alerts/ctmain |
| Subscription Information | Subscription Information for *The Plant Cell* and *Plant Physiology* is available at: http://www.aspb.org/publications/subscriptions.cfm |

© American Society of Plant Biologists
ADVANCING THE SCIENCE OF PLANT BIOLOGY