Transposons and genome evolution in plants

Nina Fedoroff*

The Pennsylvania State University, University Park, PA 16803

Although it is known today that transposons comprise a significant fraction of the genomes of many organisms, they eluded discovery through the first half century of genetic analysis and even once discovered, their ubiquity and abundance were not recognized for some time. This genetic invisibility of transposons focuses attention on the mechanisms that control not only transposition, but illegitimate recombination. The thesis is developed that the mechanisms that control transposition are a reflection of the more general capacity of eukaryotic organisms to detect, mark, and retain duplicated DNA through repressive chromatin structures.

The 50 years that have elapsed since the publication of Stebbins' "Variation and Evolution in Plants" have seen extraordinary changes in our understanding of how genomes are structured and how they change in evolution. The book's publication date roughly coincides with the first reports by Barbara McClintock that there are genetic elements capable of transposing to different chromosomal locations in maize plants (1). The book contains a brief mention of Marcus Rhoades' observation that a standard recessive a_1 allele of a gene in the anthocyanin biosynthetic pathway can become unstable and revert at a high frequency to the dominant A_1 allele in a background containing a dominant *Dt* ("dotted") allele (2). But transposable elements were not yet common fare, nor was it known that *Dt* is a transposon.

Today we know that transposons constitute a large fraction—even a majority—of the DNA in some species of plants and animals, among them mice, humans, and such agriculturally important plants as corn and wheat. Given what we now know about genome organization, it is paradoxical that the discovery of transposable elements lagged so far behind the discovery of the basic laws of genetic transmission. And it is equally curious that even when they were discovered, acceptance of their generality and recognition of their ubiquity came so slowly. It is perhaps an understatement to say that McClintock's early communications describing transposition were not widely hailed for their explanatory power. Indeed, McClintock commented in the introduction to her collected papers that the response to her first effort in 1950 to communicate her discovery of transposition in ". . . a journal with wide readership . . . ," specifically the *Proceedings of the National Academy of Sciences*, convinced her that ". . . the presented thesis, and evidence for it, could not be accepted by the majority of geneticists or by other biologists" (3). By contrast, the explanatory power of Watson and Crick's 1953 *Nature* paper on the structure and mode of replication of nucleic acids was recognized immediately (4).

An informative parallel is provided by the contrast between the immediate recognition of the importance of Darwin's theory of evolution and the long delay between Mendel's articulation of the laws of heredity and their wide acceptance in evolutionary thinking (5). It can be speculated that this was because Darwin's theory provided immediate explanations in the realm of the perceptible, whereas the hereditary mechanisms underlying variation were obscure. Variation, in Darwin's view, was continuous. Geneticists sharing his view formed the "biometrical school," devoted to the statistical analysis of inheritance. It was not at all clear how the simple rules derived by Mendel for the hereditary behavior of "differentiating characters" bore on the problem of evolution (5). The relevance of discontinuous variation or the production of "sports," as morphological mutations were called, was even less obvious, because the biometric approach treated offspring as statistical combinations of parental traits. Thus the idea that the study of mutations was central to understanding evolution was close to unimaginable a century ago.

Equally unimaginable at mid-20th century was the idea that transposable elements are essential to understanding chromosome structure and evolution, much less organismal evolution. The efforts of Bateson and other geneticists had firmly established Mendelian "laws" as the central paradigm of genetics and the identification and mapping of genetic "loci" through the study of mutant alleles was proceeding apace. Because genetic mapping is predicated on the invariance of recombination frequencies, there was plentiful evidence that genes have fixed chromosomal locations. Written at this time, Stebbins' book in general and in particular his third chapter, titled "The Basis of Individual Variation," clearly acknowledges the existence of many chromosomal differences among organisms in a population, including duplications, inversions, translocations, and deletions. At the same time, the book reflects the prevailing view that these ". . . are not the materials that selection uses to fashion the diverse kinds of organisms which are the products of evolution" (2). Instead, Stebbins concludes that the majority of evolutionarily important changes in physiology and morphology are attributable to classical genetic "point" mutations.

Another half century has elapsed and the geneticist's "black box," sprung open, spills nucleotide sequences at an ever accelerating pace. Our computers sift through genomes in search of genes, knee-deep in transposons. How could we not have seen them before? The answer is as straightforward as it is mysterious and worthy of consideration: they are invisible to the geneticist. Well, almost invisible. And of course it depends on the geneticist.

The Discovery of Transposition

The study of unstable mutations that cause variegation dates back to De Vries, who formulated the concept of "ever-sporting varieties" and eventually came to the conclusion that these types of mutations do not obey Mendel's rules (6). The first person to make substantial sense of their inheritance was the maize geneticist Emerson, who analyzed a variegating allele of the maize *P* locus during the first decades of this century (7–9). His first paper on the subject opens with the statement that variegation ". . . is distinguished from other color patterns by its

This paper was presented at the National Academy of Sciences colloquium "Variation and Evolution in Plants and Microorganisms: Toward a New Synthesis 50 Years After Stebbins," held January 27–29, 2000, at the Arnold and Mabel Beckman Center in Irvine, CA.

Abbreviation: LTR, long terminal repeat.

*E-mail: nvf1@psu.edu.

incorrigible irregularity'' (7). What follows is a brilliant analysis of ''freak ears'' containing large sectors in which the unstable *P* allele has either further mutated or reverted. Emerson was able to capture the behavior of unstable mutations in the Mendelian paradigm by postulating that variegation commenced with the temporary association of some type of inhibitor with a locus required for pigmentation. Emerson's suggestion was that normal pigmentation was restored upon loss of the inhibitor.

Several prominent geneticists, among them Correns and Goldschmidt, dismissed unstable mutations as a special category of ''diseased genes'' (10, 11). It was their view that little could be learned from the study of such mutations that was relevant to the study of conventional genes. But the drosophilist Demerec and the maize geneticist Rhoades shared Emerson's view that there was no difference in principle between stable and unstable mutations. Indeed, the Rhoades mutation cited in Stebbin's volume illustrates the important point that instability is conditional. Rhoades' experiments had revealed that a standard recessive allele of the maize *A1* locus, isolated decades earlier and in wide use as a stable null allele, could become unstable in a different genetic background. The key ingredient of the destabilizing background was the presence of the *Dt* locus, which caused reversion of the *a1* allele to wild type both somatically and germinally (12, 13).

In the late 1930s, McClintock had begun to work with broken chromosomes and by the early 1940s she had devised a method for producing deletion mutations commencing with parental plants, each of which contributed a broken chromosome 9 lacking a terminal segment and the telomere. Searching for mutants in the progeny of such crosses, she observed a high frequency of variegating mutants of all kinds (14). She noted that although reports of the appearance of new mutable genes were relatively rare in the maize literature, she already had isolated 14 new cases of such instability and observed more. She chose to follow the behavior of a locus, which she called *Dissociation* (*Ds*), for its propensity to cause the dissociation of the short arm of chromosome 9 at a position close to the centromere, although she soon appreciated that chromosome dissociation required the presence of another unlinked locus, which she designated the *Activator* (*Ac*) locus (14, 15).

By 1948, she had gained sufficient confidence that the *Ds* locus moves to report: ''It is now known that the *Ds* locus may change its position in the chromosome'' (11, 16). The relationship between the chromosome-breaking *Ds* locus and variegation emerged as McClintock analyzed the progeny of a new variegating mutation of the *C* locus required for kernel pigmentation. She carried out an extraordinary series of painstakingly detailed cytological and genetic experiments on this new mutation, *c-m1*, whose instability was conditional and depended on the presence of the *Ac* locus (11, 16, 17). She showed that the origin of the unstable mutation coincided with the transposition of *Ds* from its original position near the centromere on chromosome 9 to a new site at the *C* locus and that when it reverted to a stable wild-type allele, *Ds* disappeared from the locus. Having established that *Ds* could transpose into and out of the *C* locus germinally, she inferred that somatic variegation reflects the frequent transposition of *Ds* during development. Transposition explained both Emerson's and Rhoades' earlier observations. McClintock and Rhoades were good friends, of course, and it is evident from their correspondence that McClintock immediately saw the parallels between the behavior of the *c-m1* mutation and Rhoades' *a1* mutation (Lee Kass, personal communication).

The *Ac* and *Ds* elements are transposition-competent and transposition-defective members of a single transposon family. In the ensuing years, McClintock identified and studied a second transposon family, called *Suppressor-mutator* (*Spm*) (18, 19). Her studies on these element families were purely genetic, and she was able to make extraordinary progress in understanding the transposition mechanism because she studied the interactions between a single transposition-competent element and one or a small number of genes with insertions of cognate transposition-defective elements (20). Two points about this early history of transposition merit emphasis. First, the active elements were denumerable and manageable as genetic entities, despite their propensity to move. Second, the number of different transposon families and family members uncovered genetically was (and still is) small. Hence the genetic impact of transposable elements was limited. McClintock recognized that the high frequency of new variegating mutations in her cultures was linked to the genetic perturbations associated with the presence of broken chromosomes (14, 21). Her inference, extraordinarily prescient, was that transposons are regular inhabitants of the genome, but genetically silent.

Plant Transposons in the Age of Genomics

With the cloning of the maize transposons, first the *Ac* element in my laboratory and later the cognate *En* and *Spm* elements in Heinz Saedler's and my laboratories, the picture began to change (22–24). To begin with, it became obvious immediately that the maize genome contains more copies of a given transposon than there are genetically identifiable elements. Although most of these sequences are not complete transposons, there are nonetheless more complete transposons than can be perceived genetically (25). Importantly, it was clear almost immediately that a genetically active transposon could be distinguished from one that was genetically silent by its methylation pattern (25, 26). Both of these observations bear on the genetic visibility of transposons.

As maize genes and genome segments began to be cloned and sequenced, the discovery of new transposons accelerated. Although the transposons that McClintock identified and studied were DNA transposons, both *gypsy*-like and *copia*-like retrotransposons were soon identified in the maize genome and subsequently in many other plant genomes (27–32). It has also become evident that non-long terminal repeat (LTR) retrotransposons are abundant in maize, as well as other plant genomes (33, 34). Many additional maize transposon families have been identified through their sequence organization and their presence in or near genes (35–40). We now know that transposons and retrotransposons comprise half or more of the maize genome (41).

What Do Transposons Do?

Commencing with McClintock's elegant analyses of transposon-associated chromosomal rearrangements and extending into the literature of today, the range of transposon-associated genetic changes has continued to expand (18). Insertion of plant transposons, like almost all known transposons, is accompanied by the duplication of a short flanking sequence of a few base pairs (42). Plant transposons excise imprecisely, generally leaving part of the duplication at the former insertion site (42). The consequences of insertion and excision of a transposon therefore depend on the location within the coding sequence and excision of an insertion from an exon commonly results in either an altered gene product or a frame-shift mutation. Transposon insertions can alter transcription and transcript processing, and there are cases in which transposons are processed out of transcripts by virtue of the presence of splice donor and acceptor sequences (43–45). Transposons also can promote the movement of large segments of DNA either by transposition or by illegitimate recombination (46, 47).

The Paradox

One might think that given their abundance, transposable elements would rapidly randomize genome order. Yet the results of a decade of comparative plant genome studies has revealed that

gene order is surprisingly conserved between species. Close relationships among genomes have been demonstrated in crop plants belonging to the *Solanaceae*, and the *Graminae*, between *Brassica* crops and *Arabidopsis*, among several legumes, and others (48–50). The synteny among the genomes of economically important cereal grasses is so extensive that they are now represented by concentric circular maps (49). There are rearrangements, but a relatively small number of major inversions and transpositions is required to harmonize the present day maps. Such maps, of course, are crude representations of the genome, and rearrangements can emerge as the level of resolution increases (48, 51, 52). The frequency of rearrangements also can differ markedly and there is evidence that rearrangements are more prevalent just after polyploidization (49, 53). Even within a conservative lineage, however, some gene families are more heterogeneous in composition and map distribution than others (54).

Synteny and Divergence

What are the useful generalizations? First, synteny can extend down to a very fine level, but it is far from perfect. A detailed sequence comparison of the small region around the maize and sorghum *Adh1* loci reveals a surprising amount of change in a constant framework (52). The sorghum and maize genomes are 750 and 2,500 Mbp, respectively. The *Adh1* gene sequences are highly conserved, and complete sequencing revealed that there were seven and 10 additional genes in the homologous regions of maize and sorghum, respectively. The region of homology extends over about 65 kb of the sorghum genome, but occupies more than 200 kb in the maize genome. The gene order and orientation are conserved, although three of the genes found in the sorghum *Adh1* region are not in the maize *Adh1* region. The genes are located elsewhere in the maize genome, suggesting that they transposed away from the *Adh1* region (52). Although homology is confined largely to genes, there are also homologous intergenic regions. There are simple sequence repeats and small transposons, called MITES as a group, scattered throughout this region in both sorghum and maize. MITES are found primarily between genes, but several are in introns. The small MITE transposons are found neither in exons nor in retrotransposons. There are three non-LTR retrotransposons in the maize *Adh1* region and none in the sorghum *Adh1* region (52).

The major difference between the maize and sorghum *Adh* regions is the presence of very large continuous blocks of retrotransposons in maize that are not present in sorghum. Although most blocks are between genes, one appears to be inside a gene sequence. They are present in many, but not all intergenic regions. There is a relatively long stretch of almost 40 kb containing four genes in maize and seven genes in sorghum, which contains no retrotransposon blocks in maize and in which there is about 10 kb of extensive homology, some genic and some intergenic. Thus synteny extends down to a relatively fine level and includes both genic and intergenic sequences.

Plant Genomes Expand

A second generalization is that plant genomes grow. Genome sizes among flowering plants vary dramatically over almost 3 orders of magnitude, from the roughly 130 Mbp *Arabidopsis* genome to the 110,000 Mbp *Fritillaria assyriaca* genome (55). Genome size variation greatly exceeds estimates of differences in gene numbers (56). This, of course, is the celebrated C-value paradox (57). Plant genomes expand by several mechanisms, including polyploidization, transposition, and duplication. Thus, for example, a fine-scale comparison of the *Arabidopsis thaliana* and *Brassica nigra* genomes reveals that the *Brassica* genome contains a triplication of the much smaller *Arabidopsis* genome, as well as chromosome fusions and rearrangements (50). There is evidence that the maize genome is a segmental allotetraploid (58). It is estimated that up to 70% of flowering plants have polyploidy in their lineages (59). Thus replication of whole genomes or parts of genomes is a common and important theme in plant genome evolution.

Transposition

Transposition is also a major cause of plant genome expansion. To begin with, transposition generates DNA. Retrotransposition results from transcription of genomic retrotransposons, followed by insertion of reverse transcripts into the genome at new sites (60). Plant transposons generate additional copies of themselves by virtue of excising from only one of two newly replicated sister chromatids and reinserting into as yet unreplicated sites (20). Absent countering forces, genome expansion is an inevitable consequence of the properties of transposable elements. The accumulation of retrotransposon blocks between genes is a major factor in the size difference between the maize genome and those of its smaller relatives (61, 62). Retrotransposon blocks occupy 74% of the recently sequenced 240-kb maize *Adh* region (52). These blocks contain 23 members of 11 different retrotransposon families, primarily as complete retrotransposons, but also occasionally as solo LTRs (52). Within these blocks, retrotransposons are commonly nested by insertion of retrotransposons into each other (61, 62).

What is perhaps most surprising about the maize retrotransposon blocks that have been characterized is that they grow quite slowly. The transposition mechanism assures that retrotransposon ends are almost always identical when an element inserts, hence the divergence between the LTRs of a single element reflects the age of the insertion. Bennetzen and his colleagues found that the sequence difference between the LTRs of a given element is almost invariably less than the sequence difference between the LTRs of the element into which it is inserted. Using these differences to order and date the insertions, they inferred that all of the insertions have occurred within roughly the last 5 million years, well after the divergence of maize and sorghum (62). Importantly, no retrotransposons have been found in the corresponding *Adh1* flanking sequence in sorghum (52, 62). This raises the possibility that retrotransposon activity may differ between closely related lineages.

Amplification and Rearrangement

New copies of transposons and retrotransposons provide new sites of homology for unequal crossing over. Evidence that transposable elements are central to the evolutionary restructuring of genomes has accumulated in every organism for which sufficient sequence data exist. Exceptionally detailed examples of the role of transposition, retrotransposition, amplification, and transposon-mediated rearrangements in the evolution of a contemporary chromosome are provided by recent studies on the human Y chromosome (47, 63–65). Although the level of resolution is not yet sufficient in many cases to determine the molecular history of each duplication, it is evident that many, if not a majority of plant genes belong to gene families ranging in size from a few members to hundreds (66–69). *R* genes, for example, comprise a superfamily of similar *myc*-homologous, helix–loop–helix transcriptional activators of genes in anthocyanin biosynthesis (70–72). Detailed analysis of the *R-r* complex, a well-studied member of the *R* superfamily, reveals a history of transposon-catalyzed rearrangement and duplication (73).

There also may be other genetic mechanisms that drive genome expansion. A recent analysis of the behavior of maize chromosomal knobs reveals that the pattern of segregation under the influence of a ''meiotic drive'' locus of as yet unknown function results in the preferential transmission of chromosomes with larger knobs over chromosomes with smaller knobs (74). Maize knobs are blocks of similar short tandemly repeated sequences, ranging from as few as 100 copies to as many as 25,000

per site (75). Their structure and dispersed occurrence further suggest that they are transposable (74–76). The combination of transposability and preferential transmission of chromosomes with expanded knobs thus provides an additional mechanism for genome expansion.

Genome Contraction

Are there genetic mechanisms that contract genomes? Careful analysis of the relative deletion frequency and length in drosophilid non-LTR retrotransposons supports the inference that there are more deletions per point mutation in *Drosophila* than in mammals and that the average deletion size is almost eight times larger (77). Thus mechanisms that contract genomes by preferential deletion may exist, as well. Bennetzen and Kellogg have argued that despite ample evidence for the operation of mechanisms that expand genomes in plants, there is little evidence that plant genomes contract (56). The maize intergenic regions that have been analyzed, for example, comprise predominantly intact retrotransposons, rather than solo LTRs, which can arise by unequal crossovers between the repeats at retrotransposon ends and are common in other genomes (56). However, it also is known that both the *Ac* and *Spm* transposons of maize frequently give rise to internally deleted elements, and *Ac* ends are very much more abundant in the maize genome than are full-length elements, suggesting deletional decay of transposon sequences (22, 24, 25, 78). So it would not be surprising to find mechanisms that preferentially eliminate sequences. And indeed, preferential loss of nonredundant sequences early after polyploidization has been detected in wheat (79).

Controlling Transcription, Recombination, and Transposition

Despite our growing awareness of the abundance of plant transposable elements and the role they have played in shaping contemporary chromosome organization, the fact is they eluded discovery for the first half century of intensive genetic analysis. Thus what is perhaps the most striking observation about transposable elements is not their instability, but precisely the opposite: their stability. Not only are insertion mutations in genes infrequent, but retrotransposition events are so widely separated that the time interval between insertions in a particular region of the genome can be counted in hundreds of thousands to millions of years (62). Chromosomes containing many hundreds of thousands of transposable elements are as stable as chromosomes containing few. By what means are such sequences prevented from transposing, recombining, deleting, and rearranging?

The transposon problem can be viewed as one aspect of a larger problem in genome evolution: why does duplicated DNA persist? Duplications are a by-product of the properties of the DNA replication and recombination machinery. Short stretches of homology suffice to give rise to duplications by slippage during replication, homology-dependent unequal crossing-over, and double-strand breakage/repair (80, 81). But duplications are problematical. Once a duplication exists, the mechanisms that generated it also permit unequal crossing over between identical repeats (82–84). Prokaryotes readily duplicate genetic material, but do not retain duplications (83, 85). Thus the ability of genomes to expand by duplication is predicated on their ability to sequester homologous sequences from the cell's recombination machinery and retain them, which may necessitate the invention of mechanisms to recognize and differentially mark duplications. Some lower eukaryotes, including *Neurospora crassa* (86, 87) and *Ascobolus immersus* (88, 89), have the capacity to recognize and mark duplicated sequences by methylating them. Sequence methylation silences transcription, enhances the mutability of the duplicated sequence, and inhibits recombination (87, 90).

Some years ago, Adrian Bird pointed out that there are two evolutionary discontinuities in the average number of genes per genome (91). The first is an increase between prokaryotes and eukaryotes and the second is between invertebrates and vertebrates. He suggests that with a given cellular organization there may be an upper limit on the tolerable gene numbers imposed by the imprecision of the biochemical mechanisms controlling gene expression. He suggested that the transcriptional "noise reduction" mechanisms that arose at the prokaryote/eukaryote boundary were the nuclear envelope, chromatin, and separation of the transcriptional and translational machinery, as well as RNA processing, capping, and polyadenylation to discriminate authentic from spurious transcripts. He proposed that genome-wide DNA methylation is the novel "noise reduction" mechanism that has permitted the additional quantal leap in gene numbers characteristic of vertebrates.

Homology-Dependent Gene Silencing

The results of both classical and contemporary studies on the silencing of redundant gene copies in plants suggests that both methylation and other epigenetic mechanisms reflect a much more fundamental ability to recognize and regulate gene dosage (92). McClintock understood that transposable elements exist in a genetically intact, but cryptic form in the genome and she carried out genetic analyses of *Spm* transposons undergoing epigenetic changes in their ability to transpose (93). We later found that the genetically inactive *Spm* transposons are methylated in critical regulatory sequences (26). It also has been reported that the large intergenic retrotransposon blocks in maize are extensively methylated (94).

The discovery that the introduction of a transgene can lead to the transcriptional silencing and methylation of both the introduced gene and its endogenous homolog brought gene silencing mechanisms under intense study (92, 95). Genes can be silenced both transcriptionally and posttranscriptionally consequent on the introduction of additional copies. Posttranscriptional silencing appears to be caused by RNA destabilization, whereas transcriptional gene silencing involves DNA methylation (92, 96). There is also some evidence that posttranscriptional silencing triggers DNA methylation (97). The results of recent studies on the classical epigenetic phenomenon of *R* locus paramutation in maize have revealed that local endoreduplication of a chromosomal segment both triggers silencing and can render the endoreduplicated locus capable of silencing an active allele of the gene on a homolog (98). Similar observations have been made with transgenes, as well as endogenous gene duplications at different chromosomal locations in tobacco and *Arabidopsis* (99, 100).

A connection between gene silencing and chromatin structure has come from the analysis of mutants altered in methylation and in transcriptional gene silencing (92, 101). Both approaches have identified alleles of the *ddm1* locus, which encodes a protein with homology to known chromatin remodeling proteins. This suggests that the repressive mechanisms of DNA methylation and chromatin structure are linked in plants, as they are in animal cells (102, 103). Evidence also is accumulating that double-stranded RNA mediates gene silencing, both in plants and in a variety of other organisms (104, 105). Analyses of mutants altered in posttranscriptional gene silencing in *Neurospora* have identified an RNA-dependent RNA polymerase, as well as a RecQ helicase-like protein, homologs of which are known to be involved in DNA repair and recombination (106, 107).

The Origin of Transposons and Methylation

Although it is popular to assert that transposons are genomic "parasites" and that DNA methylation evolved to control them, I suggest that the evidence supports neither notion (108). The idea that transposons as parasitic, selfish DNA comes from a

couple of essays written two decades ago, one by Doolittle and Sapienza (109) and one by Orgel and Crick (110). These essays sought rightly to free us from the then prevalent notion that genome structure is optimized by phenotypic selection. But the persistence of the moniker "selfish DNA" has become an impediment to further understanding of the origin, historical contribution, and contemporary role of transposons in chromosome structure.

Transposons may be an inevitable by-product of the evolution of sequence-specific endonucleases. Complete transposons have been shown to arise from a single cleavage site and an endonuclease gene (111). Although the successful constitution of a transposon from the recognition sequences used in Ig gene rearrangement and the RAG1 and RAG2 proteins was interpreted as evidence that the V(D)J recombination system evolved from an ancient mobile DNA element, the fact is that the critical components of a transposon and a site-specific rearrangement system are the same (112). Thus questions about the origin of certain kinds of transposons may devolve to questions about the association of sequence-specific DNA binding domains with endonuclease domains.

Although the majority of methylated sequences in a genome can be transposable elements, the view that DNA methylation evolved to control transposons seems implausible in the light of evidence that duplications of any kind trigger methylation in organisms that methylate DNA (108, 113, 114). And organisms that do not methylate DNA also have mechanisms for detecting duplications and sequestering repeats (115–117). Genome expansion by duplication is predicated on preventing illegitimate recombination between duplicated sequences. Although different eukaryotic lineages appear to have invented different mechanisms, what is common to repeat-induced silencing in all eukaryotes is the stable packaging of DNA into "repressive" chromatin. It may be that the evolution of mechanisms that recognize, mark, and sequester duplications into repressive chromatin structures, among which some involve DNA methylation, were the prerequisites for expansion of genomes by endoreduplication at all scales. The additional benefit of such "repressive" mechanisms in minimizing spurious transcription could be secondary sequelae. Because sequence duplication is inherent in transposition, the ability to recognize and repress duplications would serve to minimize both the activity and the adverse impacts of transposons, rendering them genetically invisible and favoring their gradual accumulation.

An important and as yet underappreciated property of compacted, inactive genomic regions is their ability to impose their organization on adjacent, as well as nonadjacent, active regions, often in a homology-dependent manner. This is evidenced in position effect variegation in *Drosophila*, an organism that does not methylate its DNA, as well as in plant paramutation, which involves DNA methylation (98, 117). What has been learned recently from analyzing gene silencing and paramutation suggests that it does not take many tandem duplications to trigger the formation of a compacted, silenced region. A silenced region then may become a "sink" for insertions within it, as well as a silencer for homologous sequences located adjacent to it or elsewhere in the genome (117, 118).

Conclusions

The key to understanding the prevalence of transposons in contemporary genomes, as well as their genetic invisibility, therefore may lie not in transposons themselves, but in the much more fundamental capacity of eukaryotic organisms to recognize and sequester duplications. Whether transposons, retrotransposons, and other repetitive elements accumulate extensively in a given evolutionary lineage may depend on several factors, among them the efficiency of repressive mechanisms and the rate at which the sequences undergo mutational and deletional decay. For example, methylation of C residues enhances the mutability of CG base pairs, hence methylation accelerates the divergence rate of newly arising duplications. This happens in an extreme form in *Neurospora*, in which many methylated CGs are mutated in the span of a single generation, and at more measured rates in plants and mammals, in which the mutability can be detected by virtue of a marked deficiency of the base pairs and triplets that are normally methylated (62, 119, 120).

The burgeoning analyses of genomes also makes it evident that repressive mechanisms are imperfect. However slowly, genomes are inexorably restructured by transposition and rearrangements arising from ectopic interactions between dispersed transposons. Thus there is little remaining doubt that transposons are central to genome evolution. What is less clear is the relationship between genome restructuring and morphological change. We know that the magnitude of the morphological differences between species does not necessarily reflect the magnitude of the genetic or chromosomal differences between them. It recently has become evident, for example, that the marked morphological and developmental differences between teosinte and maize are attributable to a very small number of genes and that for some genes, the differences are regulatory, rather than structural (58, 121). It is also well known that genes are expressed differently depending on their chromosomal position. But what remains to be discovered is the extent to which chromosomal restructuring contributes to organismal evolution.

1. McClintock, B. (1945) *Am. J. Bot.* **32,** 671–678.
2. Stebbins, L. G., Jr. (1950) *Variation and Evolution in Plants* (Columbia Univ. Press, New Tirj).
3. McClintock, B. (1987) *The Discovery and Characterization of Transposable Elements* (Garland, New York).
4. Watson, J. D. & Crick, F. H. C. (1953) *Nature (London)* **171,** 737–738.
5. Carlson, E. A. (1966) *The Gene: A Critical History* (Saunders, Philadelphia).
6. de Vries, H. (1905) *Species and Varieties: Their Origin by Mutation* (Open Court, Chicago), 2nd Ed.
7. Emerson, R. A. (1914) *Am. Nat.* **48,** 87–115.
8. Emerson, R. A. (1917) *Genetics* **2,** 1–35.
9. Emerson, R. A. (1929) *Genetics* **14,** 488–511.
10. Goldschmidt, R. (1938) *Physiological Genetics* (McGraw–Hill, New York).
11. Fedoroff, N. V. (1998) in *Discoveries in Plant Biology,* eds. Kung, S.-D. & Yang, S.-F. (World Scientific, Singapore), pp. 89–104.
12. Rhoades, M. M. (1936) *J. Genet.* **33,** 347–354.
13. Rhoades, M. M. (1938) *Genetics* **23,** 377–397.
14. McClintock, B. (1946) *Year Book Carnegie Inst. Washington* **45,** 176–186.
15. McClintock, B. (1947) *Year Book Carnegie Inst. Washington* **46,** 146–152.
16. McClintock, B. (1948) *Year Book Carnegie Inst. Washington* **47,** 155–169.
17. McClintock, B. (1949) *Year Book Carnegie Inst. Washington* **48,** 142–154.
18. McClintock, B. (1951) *Year Book Carnegie Inst. Washington* **50,** 174–181.
19. McClintock, B. (1954) *Year Book Carnegie Inst. Washington* **53,** 254–260.
20. Fedoroff, N. (1989) in *Mobile DNA,* eds. Howe, M. & Berg, D. (Am. Soc. Microbiol., Washington, DC), pp. 375–411.
21. McClintock, B. (1978) *Stadler Genet. Symp.* **10,** 25–48.
22. Fedoroff, N., Wessler, S. & Shure, M. (1983) *Cell* **35,** 243–251.
23. Pereira, A., Schwarz-Sommer, Z., Gierl, A., Bertram, I., Peterson, P. A. & Saedler, H. (1985) *EMBO J.* **4,** 17–23.
24. Masson, P., Surosky, R., Kingsbury, J. A. & Fedoroff, N. V. (1987) *Genetics* **117,** 117–137.
25. Fedoroff, N., Furtek, D. & Nelson, O. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 3825–3829.
26. Banks, J. A. & Fedoroff, N. (1989) *Dev. Genet.* **10,** 425–437.
27. Shepherd, N. S., Schwarz-Sommer, Z., Blumberg vel Spalve, J., Gupta, M., Wienand, U. & Saedler, H. (1984) *Nature (London)* **307,** 185–187.
28. Flavell, A. J. (1992) *Genetica* **86,** 203–214.
29. Purugganan, M. D. & Wessler, S. R. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 11674–11678.
30. White, S. E., Habera, L. F. & Wessler, S. R. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 11792–11796.
31. Suoniemi, A., Schmidt, D. & Schulman, A. H. (1997) *Genetica* **100,** 219–230.

32. Suoniemi, A., Tanskanen, J. & Schulman, A. H. (1998) *Plant J.* **13,** 699–705.
33. Schwartz-Sommer, Z., Leclercq, L., Gobel, E. & Saedler, H. (1987) *EMBO J.* **6,** 3873–3880.
34. Noma, K., Ohtsubo, E. & Ohtsubo, H. (1999) *Mol. Gen. Genet.* **261,** 71–79.
35. Spell, M. L., Baran, G. & Wessler, S. R. (1988) *Mol. Gen. Genet.* **211,** 364–366.
36. Bureau, T. E. & Wessler, S. R. (1992) *Plant Cell* **4,** 1283–1294.
37. Bureau, T. E. & Wessler, S. R. (1994) *Plant Cell* **6,** 907–916.
38. Bureau, T. E. & Wessler, S. R. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 1411–1415.
39. Wessler, S. R., Bureau, T. E. & White, S. E. (1995) *Curr. Opin. Genet. Dev.* **5,** 814–821.
40. Bureau, T. E., Ronald, P. C. & Wessler, S. R. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 8524–8529.
41. Bennetzen, J. L., SanMiguel, P., Chen, M., Tikhonov, A., Francki, M. & Avramova, Z. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 1975–1978.
42. Schwarz-Sommer, Z., Gierl, A., Cuypers, H., Peterson, P. A. & Saedler, H. (1985) *EMBO J.* **14,** 591–597.
43. Kim, H. Y., Schiefelbein, J. W., Raboy, V., Furtek, D. B. & Nelson, O., Jr. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 5863–5867.
44. Wessler, S. R. (1989) *Gene* **82,** 127–133.
45. Giroux, M. J., Clancy, M., Baier, J., Ingham, L., McCarty, D. & Hannah, L. C. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 12150–12154.
46. Courage-Tebbe, U., Doring, H. P., Fedoroff, N. & Starlinger, P. (1983) *Cell* **34,** 383–393.
47. Schwartz, A., Chan, D. C., Brown, L. G., Alagappan, R., Pettay, D., Disteche, C., McGillivray, B., de la Chapelle, A. & Page, D. C. (1998) *Hum. Mol. Genet.* **7,** 1–11.
48. Tanksley, S. D., Ganal, M. W., Prince, J. P., de Vicente, M. C., Bonierbale, M. W., Broun, P., Fulton, T. M., Giovannoni, J. J., Grandillo, S., Martin, G. B., *et al.* (1992) *Genetics* **132,** 1141–1160.
49. Gale, M. D. & Devos, K. M. (1998) *Science* **282,** 656–659.
50. Lagercrantz, U. (1998) *Genetics* **150,** 1217–1228.
51. Tanksley, S. D., Bernatzky, R., Lapitan, N. L. & Prince, J. P. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 6419–6423.
52. Tikhonov, A. P., SanMiguel, P. J., Nakajima, Y., Gorenstein, N. M., Bennetzen, J. L. & Avramova, Z. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 7409–7414.
53. Song, K., Lu, P., Tang, K. & Osborn, T. C. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 7719–7723.
54. Leister, D., Kurth, J., Laurie, D. A., Yano, M., Sasaki, T., Devos, K., Graner, A. & Schulze-Lefert, P. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 370–375.
55. Bennett, M. D., Smith, J. B. & Heslop-Harrison, J. S. (1982) *Proc. R. Soc. London Ser. B.* **216,** 179–199.
56. Bennetzen, J. L. & Kellogg, E. A. (1997) *Plant Cell* **9,** 1507–1514.
57. Thomas, C. A. (1971) *Annu. Rev. Genet.* **5,** 237–256.
58. White, S. & Doebley, J. (1998) *Trends Genet.* **14,** 327–332.
59. Leitch, I. J. & Bennett, M. D. (1997) *Trends Plant Sci.* **2,** 470–476.
60. Howe, M. & Berg, D., eds. (1989) *Mobile DNA* (Am. Soc. Microbiol., Washington, DC).
61. SanMiguel, P., Tikhonov, A., Jin, Y. K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P. S., Edwards, K. J., Lee, M., Avramova, Z., *et al.* (1996) *Science* **274,** 765–768.
62. SanMiguel, P., Gaut, B. S., Tikhonov, A., Nakajima, Y. & Bennetzen, J. L. (1998) *Nat. Genet.* **20,** 43–45.
63. Saxena, R., Brown, L. G., Hawkins, T., Alagappan, R. K., Skaletsky, H., Reeve, M. P., Reijo, R., Rozen, S., Dinulos, M. B., Disteche, C. M., *et al.* (1996) *Nat. Genet.* **14,** 292–299.
64. Lahn, B. T. & Page, D. C. (1999) *Nat. Genet.* **21,** 429–433.
65. Lahn, B. T. & Page, D. C. (1999) *Science* **286,** 964–967.
66. Michelmore, R. W. & Meyers, B. C. (1998) *Genome Res.* **8,** 1113–1130.
67. Riechmann, J. L. & Meyerowitz, E. M. (1998) *Biol. Chem.* **379,** 633–646.
68. Martienssen, R. & Irish, V. (1999) *Trends Genet.* **15,** 435–437.
69. Rabinowicz, P. D., Braun, E. L., Wolfe, A. D., Bowen, B. & Grotewold, E. (1999) *Genetics* **153,** 427–444.
70. Ludwig, S. R., Habera, L. F., Dellaporta, S. L. & Wessler, S. R. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 7092–7096.
71. Perrot, G. H. & Cone, K. C. (1989) *Nucleic Acids Res.* **17,** 8003.
72. Consonni, G., Geuna, F., Gavazzi, G. & Tonelli, C. (1993) *Plant J.* **3,** 335–346.
73. Walker, E. L., Robbins, T. P., Bureau, T. E., Kermicle, J. & Dellaporta, S. L. (1995) *EMBO J.* **14,** 2350–2363.
74. Buckler, E. S. T., Phelps-Durr, T. L., Buckler, C. S., Dawe, R. K., Doebley, J. F. & Holtsford, T. P. (1999) *Genetics* **153,** 415–426.
75. Ananiev, E. V., Phillips, R. L. & Rines, H. W. (1998) *Genetics* **149,** 2025–2037.
76. Ananiev, E. V., Phillips, R. L. & Rines, H. W. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 10785–10790.
77. Petrov, D. A. & Hartl, D. L. (1998) *Mol. Biol. Evol.* **15,** 293–302.
78. Schwarz-Sommer, Z., Gierl, A., Berndtgen, R. & Saedler, H. (1985) *EMBO J.* **4,** 2439–2443.
79. Feldman, M., Liu, B., Segal, G., Abbo, S., Levy, A. A. & Vega, J. M. (1997) *Genetics* **147,** 1381–1387.
80. Gorbunova, V. & Levy, A. A. (1997) *Nucleic Acids Res.* **25,** 4650–4657.
81. Liang, F., Han, M., Romanienko, P. J. & Jasin, M. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 5172–5177.
82. Anderson, R. P. & Roth, J. R. (1977) *Annu. Rev. Microbiol.* **31,** 473–505.
83. Perelson, A. S. & Bell, G. I. (1977) *Nature (London)* **265,** 304–310.
84. Koch, A. L. (1979) *J. Mol. Evol.* **14,** 273–285.
85. Romero, D. & Palacios, R. (1997) *Annu. Rev. Genet.* **31,** 91–111.
86. Selker, E. U. & Garrett, P. W. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 6870–6874.
87. Selker, E. U. (1997) *Trends Genet.* **13,** 296–301.
88. Rossignol, J. L. & Faugeron, G. (1994) *Experientia* **50,** 307–317.
89. Rossignol, J. L. & Faugeron, G. (1995) *Curr. Top. Microbiol. Immunol.* **197,** 179–191.
90. Maloisel, L. & Rossignol, J. L. (1998) *Genes Dev.* **12,** 1381–1389.
91. Bird, A. P. (1995) *Trends Genet.* **11,** 94–100.
92. Kooter, J. M., Matzke, M. A. & Meyer, P. (1999) *Trends Plant Sci.* **4,** 340–347.
93. McClintock, B. (1962) *Year Book Carnegie Inst. Washington* **61,** 448–461.
94. Bennetzen, J. L., Schrick, K., Springer, P. S., Brown, W. E. & SanMiguel, P. (1994) *Genome* **37,** 565–576.
95. Park, Y. D., Papp, I., Moscone, E. A., Iglesias, V. A., Vaucheret, H., Matzke, A. J. & Matzke, M. A. (1996) *Plant J.* **9,** 183–194.
96. Vaucheret, H., Beclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J. B., Mourrain, P., Palauqui, J. C. & Vernhettes, S. (1998) *Plant J.* **16,** 651–659.
97. Wassenegger, M., Heimes, S., Riedel, L. & Sanger, H. L. (1994) *Cell* **76,** 567–576.
98. Kermicle, J. L., Eggleston, W. B. & Alleman, M. (1995) *Genetics* **141,** 361–372.
99. Matzke, A. J., Neuhuber, F., Park, Y. D., Ambros, P. F. & Matzke, M. A. (1994) *Mol. Gen. Genet.* **244,** 219–229.
100. Luff, B., Pawlowski, L. & Bender, J. (1999) *Mol. Cell* **3,** 505–511.
101. Jeddeloh, J. A., Bender, J. & Richards, E. J. (1998) *Genes Dev.* **12,** 1714–1725.
102. Wade, P. A., Gegonne, A., Jones, P. L., Ballestar, E., Aubry, F. & Wolffe, A. P. (1999) *Nat. Genet.* **23,** 62–66.
103. Ng, H.-H., Zhang, Y., Hendrich, B., Johnson, C. A., Turner, B. M., Erdjument-Bromage, H., Tempst, P., Reinberg, D. & Bird, A. (1999) *Nat. Genet.* **23,** 58–61.
104. Waterhouse, P. M., Graham, M. W. & Wang, M. B. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 13959–13964.
105. Fire, A. (1999) *Trends Genet.* **15,** 358–363.
106. Cogoni, C. & Macino, G. (1999) *Nature (London)* **399,** 166–169.
107. Cogoni, C. & Macino, G. (1999) *Science* **286,** 2342–2344.
108. Yoder, J. A., Walsh, C. P. & Bestor, T. H. (1997) *Trends Genet.* **13,** 335–340.
109. Doolittle, W. F. & Sapienza, C. (1980) *Nature (London)* **284,** 601–603.
110. Orgel, L. E. & Crick, F. H. C. (1980) *Nature (London)* **284,** 604–607.
111. Morita, M., Umemoto, A., Watanabe, H., Nakazono, N. & Sugino, Y. (1999) *Mol. Gen. Genet.* **261,** 953–957.
112. Hiom, K., Melek, M. & Gellert, M. (1998) *Cell* **94,** 463–470.
113. Garrick, D., Fiering, S., Martin, D. I. & Whitelaw, E. (1998) *Nat. Genet.* **18,** 56–59.
114. Selker, E. U. (1999) *Cell* **97,** 157–160.
115. Pirrotta, V. (1997) *Trends Genet.* **13,** 314–318.
116. Sherman, J. M. & Pillus, L. (1997) *Trends Genet.* **13,** 308–313.
117. Henikoff, S. (1998) *BioEssays* **20,** 532–535.
118. Jakowitsch, J., Papp, I., Moscone, E. A., van der Winden, J., Matzke, M. & Matzke, A. J. (1999) *Plant J.* **17,** 131–140.
119. Selker, E. U. (1990) *Annu. Rev. Genet.* **24,** 579–613.
120. Wang, D. G., Fan, J. B., Siao, C. J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., *et al.* (1998) *Science* **280,** 1077–1082.
121. Doebley, J., Stec, A. & Hubbard, L. (1997) *Nature (London)* **386,** 485–488.