Transposable Elements, Epigenetics, and Genome Evolution

Nina V. Fedoroff

Transposable genetic elements (TEs) comprise a vast array of DNA sequences, all having the ability to move to new sites in genomes either directly by a cut-and-paste mechanism (transposons) or indirectly through an RNA intermediate (retrotransposons). First discovered in maize plants by the brilliant geneticist Barbara McClintock in the mid-1940s, they were initially considered something of a genetic oddity (1, 2). Several decades later, TEs acquired the anthropomorphic labels of “selfish” and “parasitic” because of their replicative autonomy and potential for genetic disruption (3, 4). However, TEs generally exist in eukaryotic genomes in a reversibly inactive, genetically undetectable form we now call “epigenetically silenced,” whose discovery can also be traced to McClintock’s elegant genetic studies (5, 6). As the underlying biochemical mechanisms emerged from obscurity and epigenetics became popular toward the end of the 20th century, it was proposed that epigenetic silencing evolved to control the proliferation of TEs and their perceived destructive potential (5, 6).

Today, we know that TEs constitute more than half of the DNA in many higher eukaryotes. We know, too, that the fingerprints of TEs and transposition are everywhere in their genomes, from the coarsest features of genomic landscapes and how they change through real and evolutionary time to the finest details of gene structure and regulation. My purpose here is to challenge the current, somewhat pejorative, view of TEs as genomic parasites with the mounting evidence that TEs and transposition play a profoundly generative role in genome evolution. I contend that it is precisely the elaboration of epigenetic mechanisms from their prokaryotic origins as suppressors of genetic exchanges that underlies both the genome expansion and the proliferation of TEs characteristic of higher eukaryotes. This is the inverse of the prevailing view that epigenetic mechanisms evolved to control the disruptive potential of TEs. The evidence that TEs shape eukaryotic genomes is by now incontrovertible. My thesis, then, is that TEs and the transposases they encode underlie the evolvability of higher eukaryotes’ massive, messy genomes.

Although my examples in this essay are largely from plants, I believe that the inferences drawn apply to higher eukaryotes in general, among which plants tend toward exaggeration in genome size, TE abundance, and epigenetic complexity. Perhaps because they have no recourse to behavioral responses in coping with stressful environments, plants appear to have honed genetic and epigenetic strategies for adaptation to a much greater extent than animals.

How Transposons Came to Be Called “Selfish” DNA

The invention of DNA sequencing techniques in the late 1970s and their subsequent mechanization led to an explosion of knowledge about the structure, gene content, and organization of genomes. The 1960s had seen the development of nucleic acid reassociation techniques whose application revealed the presence of much repetitive DNA in eukaryotic genomes (7, 8). As DNA sequencing became a reality, a good deal of discussion arose over the value of sequencing entire genomes, particularly that of humans (9–12), in view of the calculation that only a tiny fraction of the genome consisted of genes in the then-conventional sense of protein- and structural RNA–coding sequences and their associated regulatory sequences (13).

A pair of papers published in *Nature* in 1980 solidified the idea that much of eukaryotic DNA, including transposons, was “junk”—a designation conferred a decade earlier by Ohno, who argued that our genomes were replete with nonfunctional DNA (3, 4, 14). The objective of the *Nature* papers was to get beyond the then

---

Fig. 1. The C-value paradox. The range of haploid genome sizes is shown in kilobases for the groups of organisms listed on the left. [Adapted from an image by Steven M. Carr, Memorial University of Newfoundland]
still-prevalent view that every bit of an organism’s DNA has a specific function crafted by selection. Thus, both papers promoted Dawkins’ concept of “selfish DNA”—the notion that DNA capable of proliferating within a genome, as TEs do, may need no other explanation for its survival (15). Orgel and Crick asserted, “The spread of selfish DNA sequences within the genome can be compared to the spread of a not-too-harmful parasite within its host” (3).

The selfish DNA concept was initially offered to explain the long-standing C-value paradox that organisms of similar evolutionary complexity differ vastly in their DNA content (16), and this it did. The C value, which is the DNA content per haploid genome, varies widely among closely related organisms of apparently comparable complexity (Fig. 1); this has for some time been attributed to the repetitive portion of the genome (17). Such variation is especially striking in angiosperms, whose highest and lowest C values differ by a factor of 2000 (18, 19). The explanation of the C-value paradox does indeed reside largely in the profound differences among genomes in the abundance of TEs, primarily retrotransposons, even as gene numbers remain relatively constant. The Arabidopsis genome, for example, contains about 27,000 genes and 20 to 25 Mb of retrotransposons, whereas the maize genome contains about 40,000 genes and more than 1800 Mb of retrotransposon sequences (20–22). What the selfish DNA hypothesis does not attempt to explain, however, is how genomes can accumulate such vast amounts of repetitive sequences, given the ease of eliminating them by homologous recombination.

The Selfish DNA Label Stuck
Fortunately, genome sequencing raced forward, and today we have a vastly more complex understanding of genome structure and organization than we might have had if sequencing had been restricted to protein- and RNA-coding sequences. But we find ourselves neck deep in TEs: Transposon and retrotransposon sequences constitute two-thirds of our own genome and 85% of the corn genome (21, 23). Ideas about junk DNA have evolved substantially over the past two decades, with growing knowledge of the regulatory roles fulfilled by noncoding sequences and their transcripts (24, 25). However, the transposon monikers from the selfish DNA papers have persisted. Having cloned the first mammalian DNA methyltransferase gene and realizing that its methyltransferase domain resembles those of prokaryotic restriction methyltransferases, Bestor proposed in 1990 that eukaryotic DNA methylation had evolved to regulate gene expression in development and chromosome structure, much along the lines of earlier proposals (26–28). But by the end of the decade, he and his colleagues had concluded that because cytosine methylation is predominantly found in TEs, methylation was more likely to represent a nuclear defense system that had evolved precisely to “control” the destructive potential of “parasitic sequences,” mostly transposons and retrotransposons (5, 29). This view was widely accepted, and transposons are today almost universally referred to as “invaders,” “parasites,” or “parasitic sequences” (30–33).

Does the notion that epigenetic mechanisms evolved to control invading “parasitic” transposons still fit the facts in the light of the many that have accumulated since this hypothesis was first advanced? Perhaps not. The difficulty starts with the question of where such parasites might have come from. It turns out that genes encoding transposases, which all have certain common structural motifs in their catalytic cores, are present throughout eukaryotes (34) and can be traced back into prokaryotic organisms that do not have the elaborate epigenetic regulatory superstructure of eukaryotes (35). That is, transposons were around long before the eukaryotic lifestyle, with its bloated genomes, appeared on the evolutionary scene. This implies that transposons coevolved with all the rest of the eukaryotic genome’s inhabitants. Moreover, prokaryotic transposition is minimized and regulated by mechanisms that are similar to those in eukaryotes, including weak, enhancer-insensitive promoters, transposon-encoded regulatory proteins and regulatory RNAs, and DNA methylation (36–39). Yet prokaryotic genomes carry only modest numbers of transposons.

It is true that the ability of eukaryotic transposons and retrotransposons to accumulate in large numbers, together with their highly generic transposition mechanisms, means that the proliferation of a transposon introduced into a genome lacking it—whether by a genetic cross or a virus—makes it resemble an “invader” (40, 41). And indeed, there is growing appreciation that transposons are subject to horizontal transfer in eukaryotes, in some cases through host-parasite interactions (42–46). But the same is true of prokaryotic transposons (37). As well, new transposons can arise within a genome and silent transposons can be mobilized anew by a variety of physiological and genetic stresses, undergoing “bursts” of transposition to expand genomes over millennia before being silenced and decaying (47) or being silenced quickly within a generation (48).

---

**Fig. 2.** Generation and elimination of duplications by unequal crossing over. Broken lines trace the recombination event.
How Do Genomes Get so Fat?

There’s perhaps a deeper problem than the ancient origin of transposons. What distinguishes the organization of higher eukaryotic genomes from that of prokaryotes is the presence of vast amounts of duplicated DNA. It has long puzzled me that we almost universally take this for granted. But then, eukaryotes also have a markedly more complex, largely epigenetic system than do prokaryotes for managing the transcription, reproduction, and recombination of genetic material, as well as its distribution to daughter cells during mitotic and meiotic divisions. Which is cause and which is effect?

Prokaryotes can readily duplicate genome segments by virtue of small stretches of homology, but tandem duplications are rapidly lost unless retained by selection, and even then, they are generally interspersed with nonhomologous sequences (49–51). This is illustrated diagrammatically in Fig. 2. Absent either selection or a reduction in homology, tandem duplications are inevitably eliminated by homologous unequal intrachromosomal crossing over between duplicated sequences, generating one-copy organisms (an absorbing state) and organisms with increasing numbers of copies that in turn throw off singletons (50, 52). Organisms with many copies are quite unstable and are likely to be eliminated, either by virtue of the energetic drag of the extra DNA or by a population bottleneck. This is borne out by the observation that duplicate genes in prokaryotes are generally acquired by horizontal gene transfer rather than by duplication (53, 54).

The “selfish DNA” argument rests on the assumption that there exists a category of DNA that has little or no phenotypic effect, and hence is not subject to selective pressure, but can nonetheless multiply within the genome. This is not an unreasonable inference, because we know that eukaryotic genomes are packed with repetitive DNA of all kinds. But I find it quite remarkable that it passes unremarked. How did eukaryotes tip the balance between duplication and deletion that keeps genome size small in organisms in which homology-dependent recombination mechanisms predominate? And how can transposons, whose duplicative mechanisms create dispersed repetitive sequences, build up in large numbers, given the ability of homologous recombination among them to cause major, even catastrophic, chromosomal rearrangements?

What Epigenetic Mechanisms Do and How They Came to Be

I believe that the answer to these questions lies precisely in the epigenetic mechanisms that eukaryotes have elaborated to a much greater extent than prokaryotes. Repressive protein complexes, histone methylation, RNA interference (RNAi), and RNA-directed DNA methylation, as well as recombinational regulatory complexes, are among the epigenetic mechanisms that have so far surfaced (55–59). These serve a variety of structural and regulatory functions, but perhaps the essential one for understanding the evolution of eukaryotic genomes is the minimization of illegitimate and ectopic recombination among homologous sequences during DNA replication and the DNA break-repair processes that maintain genome and chromosome stability.

Heterochromatin, the highly compacted chromosome regions rich in repetitive DNA, is recombinationally inert (60, 61). Although not all eukaryotes use all of the known epigenetic mechanisms, even lower eukaryotes with relatively small genomes use RNAi to stabilize repetitive DNAs, such as ribosomal RNA genes and centromeric repeats (61–63). Disruption of the RNAi machinery disturbs the repair of double-strand breaks, stimulating repair by homologous recombination (66).

The evolutionary origins of the eukaryotic epigenetic regulatory machinery lie in bacterial systems that discriminate endogenous DNA from that acquired through horizontal gene transfer and bacteriophage infection (67–69). Although prokaryotic transposons can move both by conjugation (commonly on plasmids) and on bacteriophage, they do not appear to have been independently targeted for inactivation by either the restriction-modification system or the CRISPR (clustered regularly interspaced short palindromic repeat)–cas (CRISPR-associated) interference pathway (70). This recently discovered bacterial pathway confers sequence-specific immunity to phage and plasmids and exhibits parallels with eukaryotic RNAi systems, particularly the Piwi-interacting RNA...
system of Drosophila (71, 72). Cytosine methylation is widespread in both prokaryotes and eukaryotes, and eukaryotic DNA methylases evolved from bacterial restriction-modification methylases by acquiring new recognition and binding modules (69). The evolution of DNA methylases appears to have proceeded in parallel with that of histone-modifying enzymes and RNA-based silencing mechanisms, so that today they comprise intimately interconnected systems (68, 69). Some eukaryotes lack either DNA methylation or the RNAi machinery (or both), but nonetheless exhibit epigenetic silencing; hence, there is some redundancy among silencing mechanisms (73, 74). Precisely how these disparate systems came together is not yet known, but the evolutionary genius of linking RNAi feedback mechanisms to the heritability of the DNA methylation mechanism means, of course, that silencing can be inducible, sequence-specific, andheritable.

My contention is that it was precisely the evolution of prokaryotic epigenetic mechanisms, originally limiting recombination among horizontally exchanged sequences, to regulate homologous recombination within the eukaryotic genome that made it possible for genomes to grow. Interference with DNA methylation, histone modification, and the small RNA pathways of contemporary genomes generally destabilizes repetitive regions, both tandem and dispersed (48, 61, 62). The ability to suppress homologous recombination might well be what tipped the balance between duplication and deletion in favor of sequence endo-reduplication in general and transposon proliferation in particular. The fact that small diffusible RNA molecules are at the heart of the silencing machinery also means that new copies of transposons cannot evade regulation by moving to new locations where the ability to cause severe chromosomal disruptions through ectopic homologous recombination might consign them to the scrap heap of evolution. What I am suggesting, then, is that TEs accumulate in eukaryotic genomes because of, not despite, epigenetic silencing mechanisms. This is exactly the inverse of the “parasite control” hypothesis, which posits that epigenetic mechanisms arose to control invading, parasitic transposons (5).

The ability to retain duplicated sequences is also arguably a critical step in the evolution of multicellular organisms, underpinning the ability to diversify duplicates for expression in specific cells and tissues, at different developmental moments, and in response to different environmental stimuli (75). Equally key is the ability to program genes for differential expression by a variety of mechanisms, among which are the relatively stable ones involving DNA and histone modification, as well as the more labile small RNA–mediated and transcriptional mechanisms. On balance, then, the likelihood that contemporary eukaryotic genomes evolved in the context of epigenetic mechanisms seems vastly greater than the likelihood that they were invented as an afterthought to combat a plague of parasitic transposons.

Plant Genomes Do It More

Plants have a more complex and redundant array of epigenetic silencing mechanisms than animals, making use of multiple DNA methylation mechanisms, chromatin protein modification, and feedback mechanisms involving small noncoding RNAs (55, 58, 76). Mammals primarily methylate the C residues in the CG dinucleotide context, whereas plants methylate C residues in nucleotides within all sequence contexts (55,
DNA methylation stabilizes the silencing and inactivation of genes and other genetic elements in many eukaryotes, but is not universal; Drosophila and budding yeast represent well-investigated exceptions (55, 78). Chromosomal protein modification, particularly histone 3 methylation, is involved in guiding DNA methylases to their correct targets in both plants and animals (55, 79). In plants, unmethylated DNA is methylated by one DNA methyltransferase, whereas maintenance methylation involves two additional DNA methyltransferases (55). Figure 3 shows an overview of the several epigenetic mechanisms currently known to exist in plants.

Sequence specificity is imparted to DNA methylation through a mechanism called RNA-directed DNA methylation (RdDM) (57, 58). RdDM involves two unique plant RNA polymerases, Pol IV and Pol V, and is mediated by 24-nucleotide small interfering RNAs (siRNAs) (55, 57, 80). As illustrated in Fig. 4, RdDM is initiated by conversion of Pol IV–generated transcripts to RNA duplexes by an RNA-dependent RNA polymerase (RDR2). The duplexes are then cleaved into 24-nucleotide siRNAs by an RNAse III–family enzyme (DCL3) and the appropriate strand associates with the Argonaute family protein AGO4 (55, 80, 81). This leads to the formation of a complex comprising the AGO4-siRNA and a number of other proteins (including a DNA methylase, DRM2), which then triggers local DNA methylation (57, 80, 82). As well, activation of RdDM promotes histone deacetylation and methylation changes that lead to the establishment of chromatin structures that repress transcription (57, 80). Plant DNA demethylation is mediated by one of several DNA glycosylase activities that removes the 5-methylcytosine, after which the DNA backbone is cleaved at the abasic site and repaired (83).

Why and exactly how the different epigenetic systems evolved remains to be understood. But plants use epigenetic systems today in a variety of developmental contexts. Unlike higher animals, plants do not set aside a germ line early in development; this imposes more stringent requirements for maintaining genetic integrity, because differentiated genomes must eventually be reprogrammed for reproduction. So the elaboration of epigenetic mechanisms may have made possible the indeterminate lifestyle of many higher plants and their ability to reproduce in response to environmental signals. Epigenetic mechanisms, for example, regulate such environmentally responsive developmental transitions as vernalization, a cold-temperature requirement for germination or the transition to flowering, the reproductive phase. For example, Arabidopsis plants requiring vernalization grow vegetatively unless exposed to a period of cold to induce flowering, as illustrated in Fig. 5 (84, 85). Vernalization is mediated by cold-induced epigenetic silencing of the FLOWERING LOCUS C (FLC) gene that encodes a repressor of flowering. The silencing increases with the duration of the cold period, involves production of noncoding FLC transcripts, and results in histone modifications that inactivate transcription of the gene (86).

Although transposons are primary targets for epigenetic silencing, they are far from the only targets in plants. The first plant gene silencing mechanism understood at the molecular level was that underlying the long-known ability of a viral infection to cross-protect a plant against infection by a closely related virus (87–90). Then, in 1994, it was reported that a wholly artificial gene comprising a viroid cDNA became methylated and transcriptionally inactive in the tobacco genome, but only if viroid RNA replication had occurred, suggesting a feedback mechanism initiated by transcript overabundance (91). It was subsequently discovered in the early days of plant molecular modification that an introduced transgene encoding an
enzyme in the pigment biosynthetic pathway was subject to silencing (92–94). Later studies found that silencing entails both transcriptional and posttranscriptional mechanisms, and that these mechanisms share characteristics with those used by plants to control viral pathogens and are mediated by the production of siRNAs (95, 96). Sequence duplication also underlies a reversible silencing phenomenon, termed “paramutation,” in which an allele termed “paramutagenic” can heritably silence a susceptible allele termed “paramutable” of the same locus residing on the homolog (97). Paramutation involves a small RNA feedback mechanism and DNA methylation triggered by duplication of either coding or regulatory sequences (98, 99). Thus, the repetitive character of the sequence is also a common trigger for siRNA-mediated gene silencing and methylation (100).

The Contemporary Plant Genome Landscape
Despite the multiplicity of plant epigenetic silencing mechanisms, the fingerprints of transposition and recombination are evident at every level of plant genome structure, organization, and evolution. Maize genes are clustered in small groups separated by long, uninterrupted stretches of DNA consisting of retrotransposons (101, 102). Almost 85% of the contemporary 2.3-Gb maize (Zea mays or corn) genome comprises transposons, more than 75% of which are long terminal repeat (LTR) retrotransposons (21). Its roughly 40,000 genes, averaging about 3.3 kb in length, form small islands in a sea of more than a million transposons and retrotransposons belonging to almost 1300 different gene families.

In addition to forming very large blocks, retrotransposons exhibit a tendency to home to different neighborhoods. In maize, for example, gypsy and copia elements are over- and underrepresented in pericentromeric regions, respectively (21, 103). Within a retrotransposon block, younger elements are progressively nested within older elements, as illustrated in Fig. 6 for a short region near the maize adh1 gene (21, 102, 103). Such targeting can occur through the interaction of retrotransposon-specific proteins and chromatin proteins, which are themselves preferentially associated with certain types of sequences. An example is provided by the interactions of yeast Sir4p, a structural protein of heterochromatin, with a 6–amino acid motif of the Ty5 integrase protein that targets insertion into telomeric heterochromatin (104, 105). An Arabidopsis lyrata
centromeric retrotransposon was reported to insert preferentially into centromeres in *A. thaliana* (106). Because the centromeric sequences are quite different in the two species, targeting is likely to involve an interaction with the highly conserved centromere-specific structural proteins.

Unlike retrotransposons, which replicate through an RNA intermediate and reinsert DNA copies, DNA transposons move by a cut-and-paste mechanism, generally excising from just one newly replicated sister chromatid and reinserting into a site either nearby on the same chromosome or elsewhere in the genome (107). Because a copy of the transposon is retained at the donor site, such transposition events commonly give rise to additional transposon copies. DNA transposons account for a much smaller fraction of the plant genome than retrotransposons, are generally present in fewer copies, and tend to be associated with genic regions, some even inserting preferentially into genes (108). Mu transposons in maize favor recombinationally active regions of the genome (109), whereas Helitrons accumulate near but not inside each other (110). Such clustering may reflect the propensity of some TEs to move to nearby sites, long documented for the *Ac/Ds* (Activator/Dissociation) transposon family of maize (111).

**The Rapid Pace of Plant Genome Change**

Genome divergence through TE activity and recombination are ongoing processes that occur within species at surprisingly high rates. For example, a comparison among maize inbred lines revealed major differences within a region of just 150 kb surrounding the *bronz*e gene in both TE abundance and composition (Fig. 7) (112, 113). Speciation—the process by which subgroups of a reproductively compatible population become reproductively isolated—occurs by a variety of mechanisms, some of which involve both transposon mobilization and active genome restructuring. The genomes of newly formed plant species are necessarily similar, and the gene order is largely colinear. As the evolutionary distance increases, the colinearity declines rapidly, although the number and nature of the genes remain more or less constant (114, 115). Such evolutionary scrambling of genomes is illustrated in Fig. 8, which traces orthologous sequence blocks among familiar grass genomes (116).

It has long been known that genes change their chromosome locations, and it has been speculated that transposons mobilize large DNA segments because they are often found at the ends of inverted or transposed sequence (117–120). McClintock's initial studies on the *Ac/Ds* transposon family showed that chromosome breaks at the site of insertion of a nonautonomous *Ds* element could be resolved with attendant duplications, deletions, inversions, and translocations of large chromosomal segments (121, 122). More recent studies on *Ac/Ds*-associated rearrangements at the P locus identified transposition events that initiate at the 5′ end of one transposon and terminate at the 3′ end of a nearby transposon (123). Such “alternative transposition” events can generate a variety of rearrangements (depending on the relative orientation of the transposon ends) and can translocate large segments of intervening DNA.

The movement of genes is often mediated by a process that duplicates the gene and flanking sequences, leaving a copy at the original insertion site (115). Because translocated genome segments are commonly flanked by transposons, the movement of a chromosomal segment is likely to be initiated by a double-strand break at the new insertion site of a transposon and may be repaired through one of several known processes that repair double-strand breaks in plants, including synthesis-dependent strand annealing, template slippage, and unequal crossing over (114, 124). Such duplications can arise during mitotic chromosome replication, when transposition is known to occur, by the alternative transposition mechanism. The well-known tendency of transposons of the *Ac/Ds* family to undergo short-range transpositions from only one sister chromatid to an unreplicated site nearby gives rise to genic regions flanked by copies of the same transposon, facilitating subsequent mobilization of the intervening chromosome segment (111, 123).

**Genome Contraction and Divergence of Intergenic Sequences**

Genomes expand by sequence duplication, transposition, and retrotransposition, and they contract by deletion mediated by a variety of homologous and illegitimate recombination events (125–129). Solo LTR generation by unequal homologous recombination between theLTRs at the ends of a single retroelement is frequent in some plant species, particularly near genes and at the kinetochore (126). Such unequal events can also occur between adjacent elements, leading to the deletion of the DNA between two TE copies (125, 126). Not surprisingly, retrotransposon elimination by unequal and illegitimate recombination is most frequent in recombinationally active genome regions (130). In parts of the genome where TEs are abundant, homologous recombination is markedly suppressed; this is likely a causal factor in TE accumulation, as noted earlier (131, 132).

Autonomous DNA transposons commonly decay by internal deletions that reveal the operation of double-strand break repair mechanism that duplicates genetic information, now called “filler DNA,” by the invasion of a single strand into a duplex nearby and copying of a short sequence that is then inserted at the deletion breakpoint (124). McClintock identified and investigated a series of *Spm* derivatives that originated from an internally deleted, nonautonomous *Spm*, all of which were the result of further deletions within the same element at the same location (133, 134). These arose at a high frequency, but only in the presence of an autonomous element; this finding suggests that they were initiated either directly...
by the double-strand cleavage activity of
the transposase encoded by the autonomous
element or indirectly by secondary double-
strand breaks incurred during the transposition
reaction (135).

Analyses of intergenic regions in related
species have revealed that they turn over very
rapidly on an evolutionary time scale. Fine-
gained analysis of the sequence dynamics
shows that the intergenic volatility is indeed
mediated by transposons, which both delete
adjacent DNA sequences and insert filler
DNA from elsewhere in the process of repair-
ing the double-strand breaks in the DNA intro-
duced by transposases (128). Although dele-
tions commonly occur adjacent to a transpo-
son end, they can remove entire transposons.
Statistically significant clustering of such
deletions in the vicinity of transposons sug-
gests that they arise as a result of the double-
strand breaks that initiate transposition.

Traffic in Genes and Regulatory Sequences
At a much finer level of resolution, transpo-
sons contribute to creating genes, modify-
ing them, and programming and reprogram-
ning them. Many transposons and retroele-
ments contain captured gene fragments and
can be part of gene regulatory regions (136–
142). A classic example is provided by the
maize R locus that encodes a transcription
factor necessary for synthesis of anthocy-
aran pigments. The R-r allele comprises four
tandem duplications, including a complete
coding sequence and several truncated ones
direct and inverted order separated by a
Doppia transposon. The complete coding
sequence is responsible for pigment expres-
sion in the plant, whereas the several trun-
cated copies support pigment expression in
the seed (143).

The traffic in genes and regulatory
sequences is bidirectional: Transposons pick
up bits and pieces of genes that code for
proteins other than transposases, and trans-
posase genes are pressed into services other
than transposition (144, 145). For example,
the proteins encoded by the FAr1 and FHY3
genes of Arabidopsis are both related to the
MuDR family of transposases (146). FHY3
and FAR3 are transcription factors that regu-
late light-dependent chlorophyll biosyn-
thesis in development, the former also gag-
ing phytochrome signaling to the circadian
clock (147, 148). A familiar example is pro-
vided by the human immune system, which
uses recombinase proteins that evolved from
transposons to generate sequence diversity
through V(D)J recombination (149, 150).
Transposons provide the telomeres of some
organisms and jump in to replace them in
others (151, 152). Centromeres contain and
are often surrounded by transposons (22,
153–155). As well, transposons are central to
the epigenetic phenomenon of “imprinting”
that imbues genes with different expression
patterns depending on whether they were
transmitted through male or female gametes;
such differences arise during the major epi-
getic reprogramming events of gameto-
genesis (156–158).

Driving Evolution
Thus, transposases hold a special place in
the pantheon of genome sculptors. Arguably
the products of the most abundant genes on
Earth (159), transposases are transposon-encoded
enzymes that cleave transposon ends and
attach them to new sequences. The essential
elements are (i) very strict sequence recogni-
tion and precise cleavage at the donor site
and (ii) either a relaxed sequence specificity
or no sequence specificity at the target site
(160). The prokaryotic Mu element’s trans-
position mechanism appears to be paradigm-
atic and common to many members of the
transposase superfamily (161–165). Multim-
ers of the transposase form a transpososome
complex that recognizes the transposon’s
terminal inverted repeats and brings them
together with the target insertion site (Fig. 9)
(166, 167). The transposon ends are brought
into close juxtaposition with each other at the
donor site for a coupled reaction that cleaves
the transposon termini, introduces a stag-
gered cleavage at the target site, and trans-
fers the 3′ ends of the transposon to the over-
hanging 5′ ends at the target (168, 169).
The gapped duplexes at the two element ends
are then repaired to generate the target site duplica-
tion, whose length is a TE family charac-
teristic determined by the transposase. Sub-
sequent excision of the transposon generally
leaves behind an imperfect version of the
target site duplication, generating sequence
diversity (170, 171).

In both bacteria and plants, recognition
and cleavage of the two hemimethylated ter-
minal inverted repeats of a single transpos-
on ensures genome integrity and confines
transposition to just one of the two newly
replicated daughter strands or sister chro-
matids (36, 172). But like other aspects of
transposition, recombination, and DNA
repair, this process is error-prone and can be
fooled by such transposition events as the
insertion of one Ds transposon in inverted
orientation into the center of itself, giving
the chromosome-breaking double Ds trans-
poson that led McClintock to her momen-
tous discoveries about how transposons
move and restructure chromosomes (173,
174). The bottom line for genomes is that
the cleavage and resection of DNA by trans-
posases virtually guarantees sequence vari-
ation, genome scrambling, and the appear-
ance of transposons at rearrangement
breakpoints. Simply put, transposases drive
genoome evolution.
Genomic Shock and Transgenerational Epigenetic Inheritance

Both unpredictable stresses, such as irradiation, and predictable abiotic stresses, such as heat shock, elicit from genomes a highly programmed response intended to minimize the impact of the stress. McClintock coined the term “genomic shock” to refer to such a response (175). It is by now amply documented that plant transposons are activated in response to a variety of DNA-damaging agents and both biotic and abiotic stresses, as well as pathogen infection and the passage of plant cells through tissue culture (176–181). Other sources of natural chromosomal disturbance are provided by interspecific hybridization and allopolyplidization, both of which trigger the activation of transposons (182–184). This appears to be true as well in other eukaryotes, from yeast to flies to humans. Telomerases are relatives of retrotransposon-encoded reverse transcriptases, and transposons either comprise or can be filled in for missing telomeres in flies and yeast, respectively (152, 185–187).

Evidence is accumulating that both biotic and abiotic stresses induce a heritable increase in the ability of plants to withstand infection and tolerate stress (188–191). DNA damage, pathogen infection, and abiotic stresses also increase homologous recombination frequency and chromosomal rearrangements, both somatically and heritably (188, 192–195). Thus, responses to stress—whether from pathogens, environmental extremes, or damage to the genetic apparatus—evoke not just a transcriptional response, but also a profound and to some extent heritable change in the epigenetic framework.

Plant responses to stress—whether from pathogens, environmental extremes, or damage to the genetic apparatus—evoke not just a transcriptional response, but also a profound and to some extent heritable change in the epigenetic framework.

Heat-induced transcription and transposition of the Arabidopsis OSEN retrotransposon is rapidly silenced, becoming transgenerational only in plants with a compromised RdDM pathway (181). Thus, transcriptional activation by demethylation can also trigger a feedback mechanism that restores methylation and resiliences transposons. Recent years have seen progress in identifying the components of the restructuring response, but we do not yet know how cells and organisms perceive and initiate epigenetic reorganization in response to either genetic disruptions or environmental stresses.

Evolvability and Its Agents

I have argued that epigenetic mechanisms not only underpin the management of transcription and chromosome structure, but also provide the key to understanding the size and organization of eukaryotic genomes.

They assure the stability of chromosomes, including vast manegaries of TEs, and manage the replication and segregation of the genetic material in both mitosis and meiosis. My argument is that TEs accumulate because of, not despite, the epigenetic mechanisms that control homology-dependent recombination, whose dominance keeps the genomes of prokaryotes and many lower eukaryotes small. Absent the existence of such mechanisms, ectopic, homology-dependent recombination among dispersed TEs would rapidly eliminate them, either directly by intrachromosomal deletions or indirectly by creating nonviable chromosomes.

But although epigenetic mechanisms slow the pace of genome restructuring to an evolutionary time scale, the impact of transposons and retrotransposons on genes and genomes is inescapable. Indeed, their ability to move and to move sequences has shaped higher eukaryotic genomes, from the structuring and restructuring of genes and their regulatory sequences to the shaping and reshaping of the genomic landscape. It is becoming increasingly difficult to escape the conclusion that eukaryotic genome evolution is driven from within not just by the gentle breeze of the genetic mechanisms that replicate and repair DNA, but by the stronger winds (with perhaps occasional gale-force gusts) of transposon activity. The ability to evoke rapid genome restructuring is at the heart of eukaryotic evolvability—the capacity of organisms with larger and larger genomes to maintain evolutionary flexibility.

References

Transposable Elements, Epigenetics, and Genome Evolution
Nina V. Fedoroff

Science 338 (6108), 758-767.
DOI: 10.1126/science.338.6108.758

ARTICLE TOOLS http://science.sciencemag.org/content/338/6108/758
REFERENCES This article cites 197 articles, 61 of which you can access for free http://science.sciencemag.org/content/338/6108/758#BIBL
PERMISSIONS http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. 2017 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. The title Science is a registered trademark of AAAS.