Transposon biology – A historical perspective

Lakshminarasimhan Krishnaswamy

Ever since Gregor Mendel postulated his laws of genetics, the field of biology has made immense progress. Beginning early 1900s, there was a boom period in genetics research for over half a century. The discovery of the structure of the DNA by Watson and Crick in 1953 sowed the seeds for molecular biology, and in less than 50 years biology has evolved into the genomics era. Through a steady but quick progression in this field, different areas of research have been of foremost interest at various times. Interestingly, ever since the discovery of transposons by Barbara McClintock in the 1940s, the study of transposon biology has undergone a parallel progression through these years. This note traces the history of transposon research through the genetics, molecular biology and genomics era.

McClintock's discovery

Barbara McClintock is possibly the most enigmatic of modern scientists. Though she is well known for her discovery of transposable elements in maize, her research interest has spanned over several aspects of maize genetics.

In the 1920s during her graduate studies at Rollins Emerson's laboratory in Cornell, McClintock was part of a team that established the discipline of maize cytogenetics, and she characterized the ten maize chromosomes. In collaboration with Harriet Creighton, she observed the cytological correlation for the phenomenon of crossing-over. Between 1935 and 1941, at the University of Missouri, Columbia, she studied X-ray-irradiated maize to map genes. Here she observed ring chromosomes and her studies on deletion of chromosome ends contributed to the early understanding of telomere biology. In one of her strains, McClintock observed spontaneous breakage and fusion of chromosome arms, which repeated over somatic and germinal cell divisions, and she described the 'breakage–fusion–bridge' cycle.

McClintock moved to Cold Spring Harbor in 1942 and continued to use the 'breakage–fusion–bridge' strains to map genes. She identified two interesting loci – *Dissociator* (*Ds*) and *Activator* (*Ac*), which could transpose or change positions on the chromosome. In one case, she observed frequent chromosome break occurring at the *Ds* locus on chromosome-9 in *Ac*-dependent manner. Interestingly, this *Ds* element could move in the genome even to different chromosomes. In another case, the *Ds* locus seemed to regulate the expression of neighbouring genes. The change in position of the *Ds* element correlated with the expression of the *C* gene, and resulted in variegation of the kernel colour. Based on her observations, McClintock proposed that the *Ac* and *Ds* were 'controlling elements' that regulated the expression of other genes.

At a time when the true nature of the gene was yet unclear, and genes were thought to be static structure in the nucleus, McClintock's discovery of mobile genetic elements was revolutionary. Though many scientists could realize the significance of her observation, McClintock's emphasis on the role of transposable elements as regulators of gene expression drew indifferent reaction from the scientific community.

New transposable elements

In the wake of the discovery of *Ac/Ds*, few more transposable element systems were identified in maize. In 1936, Marcus Rhoades observed frequent mutability of anthocyanin pigmentation in the aleurone of the endosperm in Mexican Black corn, giving a dotted phenotype. He determined that the unstable phenotype of the *a1* gene that was responsible for the aleurone pigmentation was controlled by the *Dotted* (*Dt*) gene. McClintock proposed that *Dt* was a controlling element analogous to the *Ac/Ds* system, and Nuffer established that the dotted phenotype observed in Rhoades' plants was caused by the activity of the *Dt* transposable element. Similarly, study of the mutability phenomenon in maize leaf, kernel, anther and husk by Peterson, identified the *Enhancer-Inhibitor* (*En/I*) element at the *pale green mutable* (*pg-m*) locus. Almost simultaneously, McClintok identified the *Suppressor-mutator* (*Spm/dSpm*) element. Later it was shown that these two systems, *En/I* and *Spm/dSpm*, are identical.

Controlling elements or gene disruption by insertion?

The years following the discovery of the *Ac/Ds* elements were marked by genetic characterization of newer transposons in maize. However, there was an uncertainty among the biologists whether transposons caused alteration in phenotype of other genes because they disrupt other genes by insertion, or, they cause alteration in phenotype by their 'controlling' influence. When the limitations of maize genetics to resolve this puzzle was clearly felt, research in bacterial genetics was breaking new grounds. In 1961, Jacob and Monod proposed the operon model for regulation of expression of the β-galactosidase gene in *Escherichia coli*. Here the *Lac I* gene regulates the expression of the *Lac-operon* in trans. Based on this observation, McClintock suggested a parallel between regulation of gene expression in bacteria and the effect of the 'controlling elements' in maize. However, two other papers in the following years seemed to add credence to the 'disruption of gene expression by insertion of the transposon' model, as advocated by Peterson. First is a study on bacteriophage-induced mutations in *E. coli*, where Taylor demonstrated that the viral episome caused frequent mutations by inserting into the bacterial genome. A more emphatic support was from the discovery of the *IS2* insertion sequences element causing mutation by insertion into the gal operon in *E. coli*.

In the following decade, more transposable elements, including *IS* and *Tn* elements in bacteria, *Ty* elements in yeast and *copia*-like elements, *Foldback* elements and *P* elements in *Drosophila* were identified, mainly due to their ability to cause mutations, rearrange the genome, and regulate or modify the expression of other genes. The power of genetics in these systems, their simplicity and small size of genome, coupled with the advantage of emerging molecular techniques enabled the characterization of many transposable elements at the genetic and molecular level.

Molecular studies on transposons – the new era

Though evidence from transposable elements from other systems were in favour of the insertion model of mutagenesis and gene regulation, the maize elements were still considered to be 'controlling elements' for lack of direct molecular evidence from the maize elements.

The early 1980s was a significant period for transposon biology. In 1983 came the first evidence that insertion of a *Ds* element caused a mutation in the *Adh1* gene in maize. This was followed by the cloning of *Ac* and then *Ds* element from the maize *waxy* gene. The same year McClintock was recognized with the Nobel Prize for the discovery of transposable elements made about 40 years earlier.

The cloning and molecular studies on the maize *Ac/Ds* transposon were followed by characterization of several transposons from various organisms. A significant breakthrough that spiced up transposon research at this juncture was the identification of retrotransposons in yeast. For a long time, the similarity between the *copia* elements of *Drosophila* and retroviruses had been well known. A. J. Flavell isolated extrachromosomal linear *copia* elements, and J. D. Boeke *et al.* proved that the *TyH3* element transposition in yeast passes through an RNA intermediate. These two papers established that a class of transposons transposed via an RNA intermediate. Soon retrotransposons were reported in other organisms, including plants.

Towards the end of the 1980s transposon biology was not limited to 'molecular stamp collecting' of transposable elements, but opened up newer avenues of research. Advancements in molecular techniques and the emerging genomics tools enriched our understanding of the function and complexity of the genomes of eukaryotes, and the effect of transposons on genome complexity.

The 1990s – molecular era and genomics era synergy

Transposon biology has seen tremendous advancement in the past 15 years. We have identified new transposons, understood the mechanisms of transposition at the molecular level to some extent, studied the effect of transposition on the host genome stability and dynamics, and also gathered knowledge on how in certain instances the transposon sequences contribute to some cellular functions. In addition, we have managed to engineer transposons to serve as tools in functional genomics research. The vastness of the knowledge we have accumulated makes it impossible either to trace it in a chronological order or to dwell in detail on any topic. Therefore, the diversity of transposable elements, and a brief summary of our understanding and utility of transposons gathered in this period is presented in nutshell. The reader is requested to refer specialized literature for detailed discussion on these topics.

Transposon diversity

During the early days of transposon biology, transposons were classified into two – autonoumous and non-autonomous elements. Autonomous elements are those that encode functional transposase and other factors essential for transposition, whereas non-autonomous elements lack one or more of the molecules required for transposition and therefore cannot transpose independent of the corresponding autonomous element.

In the early 1980s, the advent of molecular techniques enabled identification of newer transposons, which required classification on different parameters. Eukaryotic transposons are now classified into two classes based on whether transposition involves an RNA intermediate or not.

Class 1

In Class 1 transposons, the transposon is transcribed into RNA, which is reverse-transcribed by the transposable element-encoded reverse transcriptase. The cDNA is inserted at different loci by the transposase. Class 1 elements are also called retrotransposons, and include both autonomous and non-autonomous elements.

Retrotransposons are further classified as LTR elements and non-LTR elements, based on their transposition mechanism and structure. LTRs are named after the 'Long Terminal Repeats' in direct orientation at either end of the transposon. The LTRs bear structural similarity to retroviruses. The autonomous elements encode at least two genes: the *gag* gene encodes a capsid-like protein and the *pol* gene encodes a polyprotein with various activities, including protease, reverse transcriptase, RNaseH and integrase. LTR elements are further classified into *copia*-like and *gypsy*-like elements. The non-LTR retrotransposons lack the long terminal repeats; instead they terminate by simple sequence repeats. They are divided into two: long interspersed nuclear elements (LINES) and short interspersed nuclear elements (SINES). Their coding regions include a *gag*-like protein, an endonuclease and reverse transcriptase. All SINES described yet have an internal RNA *polIII* promoter near the 5′-end. Some of the SINE elements share homology with certain LINES at the 3′-tail sequence, and could parasitize on their transposition machinery.

Class 2

The prokaryotic insertion sequences (IS elements) and eukaryotic DNA transposons transpose by direct cut-and-paste mechanism without any RNA intermediate. The transposase recognizes the terminal inverted repeat sequences (TIRs) on either ends of the transposable element, cuts it out and inserts at a new site. In order to insert the transposon, a staggered cut is made at the target site which causes duplication of a few bases – target site duplication (TSD), which now flank the inserted element. When the transposon is excised, the flanking direct repeats are joined to leave a 'footprint'. DNA transposons are grouped into several superfamilies and families based on sequence similarity, length of the TIRs, and length of the TSD. *hAT*, *CACTA*, *Mutator*, *PIF/Harbinger*, *Tc1/mariner* are well-represented class 2 transposable element groups in eukaryotes.

Newer transposons

The advent of the genomics era has introduced a few new transposon families.

MITES: Miniature inverted-repeat transposable elements (MITES) were first identified a decade ago through computer-assisted searches for repeated sequences in the grass genome. These elements are characterized by their very small size (less than 600 bp) and terminal inverted

repeats. MITES are abundantly represented in the genomes of plants and animals, and are mostly inactive, although active MITES have been reported in rice.

Helitrons and polintons: These DNA transposons were identified by computational analysis of genome sequence of different organisms. Helitrons are presumed to transpose by rolling circle replicon, similar to some plasmids, bacterial IS, single-stranded bacteriophage, and plant geminiviruses. Polintons are yet another new type of DNA transposons discovered recently by computational analysis of genome sequences. Polintons are widely found in several invertebrate and vertebrate genome sequences, and even protists and fungi, but are not found in the mammalian genome sequence. The predicted mechanism of polinton transposition is the most complex known for any eukaryotic DNA transposon, where a single strand of the transposon is excised, the second strand synthesized and the transposon inserted in the host genome. Because of the novelty of the mechanism of transposition, helitrons and polintons cannot be clearly placed in any of the above mentioned classes.

Role of transposons in genome rearrangement

The role of transposons as an insertional mutagenic agent is well documented. In addition, transposons play a significant role in various kinds of genome rearrangements. It has been shown that the repair mechanism following double-strand break caused by transposon excision could lead to intrachromosomal or ectopic homologous recombination. In addition, transposon sequences distributed all over the genome could provide sequence homology for ectopic recombination. In a normal transposon, the transposase recognizes the TIRs at either end of the element. If the transposase recognizes two TIRs that are reversed or in direct orientation in reference to each other, such non-standard or aberrant transposition reaction could lead to chromosome breakage, inversions or deletions in the genome. In fact, chromosome breakage caused by an aberrant transposition of a *Ds* element on chromosome-9 in maize led to the identification of *Ac/Ds* elements by McClintock. Such rearrangements have considerable impact in restructuring the genome. Non-standard transposition reaction at reversed transposon ends has been shown to result in the formation of chimeric gene in maize.

Effects of transposon on the host

Transposon activity in the cells results in increased mutation rates and enhances the adaptation of the organism to changing environments. Transposons can inactivate or alter the expression genes when they insert into the coding region or the regulatory elements of genes, and thereby act as a source of variation in organisms. Transposable elements could contribute to epigenetic silencing of flanking genes when they are targetted for methylation by the host defense mechanism. Transposon insertions result in duplication of a short sequence at the target site, which is seen as a 'footprint' after excision of the element. Such footprint could modify the gene coding sequence as shown in the case of the *Ascot* transposon from *A. immersus*. Here, integration of the transposon into the *b2* spore colour gene led to colourless spores. The type of footprint produced upon excision varied and different types of revertants were obtained with speckled, banded, spread, blotchy or double-belted spore phenotypes.

There are also examples where transposons are directly involved in the host-cell functions. Retroposons *HeT-A* and *TART* in *Drosophila* have been implicated in the rescue of damaged telomeres and chromosome-end maintenance. Transposons sometimes contribute to the formation of introns, or promote exon-shuffling. In *Arabidopsis*, it appears that the transposon *DAYSLEEPER* has been domesticated by the host to regulate global gene expression and is essential for normal plant development. Retrotransposition of an engineered human retrotransposon *Line1* in rat neuronal precursor cells has been shown to influence the expression of neuronal genes and influence neuronal cell fate. In primates, fusion of the transposase domain of a mobile element to a histone methyl transferase gene is supposed to have given rise to *SETMAR*, a chimeric gene.

Transposons sometimes seem to be beneficial to the host and are considered as 'pacemakers of evoluton'; while in other instances they are likely parasitic or 'selfish DNA'. The debate whether transposons are 'angels' or 'imps' has been going on for over 25 years. But, of late a synthetic view of mutualistic relationship between the host genome and transposable elements, tipping towards parasitic relation in some cases, has been suggested.

Transposon as tools in functional genomics

Although the goal of fundamental research is to gain knowledge, often it leads to applications that fuel the advancement of science and technology. This has been true with transposons biology too. Transposons have been widely used to cause mutations in genes by insertion, which has helped us identify and characterize several genes. For instance, the *Ac/Ds* transposon system from maize has been used in several heterologous systems, including *Arabidopsis* to generate libraries of insertional mutants, thereby simplifying mapping of the mutant gene. Transposons have also been used in enhancer trap and promoter trap constructs to help identify enhancer and promoter elements in the genome. Aberrant transposition reactions cause genome rearrangements like deletion, inversion and translocation. The deletions caused by such a transposition could be useful in knocking out a few genes, and in studying the effect of alteration in gene dosage. Such rearrangements are also useful in studying the function of chromosomal elements such as the centromere and neocentromere. Further, transposons have been used as gene therapy vectors in mammalian cells.

What is ahead?

One key feature of progress in science is it that is unpredictable. However, it is tempting to project certain outcomes based on the current trend in this field. In fact, research is ongoing in some of these aspects and the results could be realized in just a few years.

As the whole genome sequence from more organisms become available, it would enable us mine out newer transposons. How many of these will be active is an open question. However, with the success in resurrection of two synthetic transposons – *Sleeping beauty* and *Frog prince*, reconstructed from inactive transposons in fish and frog genome respectively, it is likely that more inactive

transposons could possibly be engineered to activity. These newer transposons will enrich the geneticists' tool kit for functional genomics. A deeper insight into the mechanisms of transposition could help us engineer precise deletions, inversions and other rearrangements in the genome. The most significant application would be precise integration of the transposon at a defined target sequence. Current techniques for generating transgenic plants suffer from poor gene targetting efficiencies. Similarly, our effort for efficient delivery and integration of gene therapy constructs in mammals is also plagued with problems. Viral vectors are efficient gene therapy vectors, but they often lead to immunological complications, or genotoxic effects due to non-specific integration. Recent success in the use of the transposons *Sleeping beauty*, *Frog prince*, *Tol-1* and *piggyBack* in mammalian gene transfer gives the hope that we are not far from the day when transposons could be used for precise integration of transgene at a specific target site, in both plants and animals. Evidently, transposon biology holds a lot of promise for the years to come.

Recommended reading

1. http://www.cshl.edu/History/mcclintock.html (The life history of Barbara McClintock is described here.)
2. Peterson, P. A., *Maydica*, 2002, **47**, 147–167; 2005, **50**, 321–338.
3. Craig, N. L. *et al.* (eds), *Mobile DNA II*, AMS Press, 2002 (Reviews on various aspects of Mobile DNA elements, including transposons.).

ACKNOWLEDGEMENTS. I thank Dr Thomas Peterson and Xianyan Kuang for helpful comments on the manuscript. Research in Dr Peterson's lab, Department of Genetics, Development and Cell Biology, Iowa State University, Ames, Iowa, USA, is supported by grants from the NSF. I apologise for not including a comprehensive list of references for all the research discussed here, due to space limitations. A short list of references for further reading is suggested here. However, the reader may contact me for a comprehensive list of references.

Lakshminarasimhan Krishnaswamy did his graduate studies from the Department of Genetics, Development and Cell Biology, Iowa State University, Ames, Iowa 50011, USA. He is currently at the Division of Biological Sciences, University of Missouri – Columbia, Missouri 65211, USA.
e-mail: krishnaswamyl@missouri.edu

Edited and published by P. Balaram, Current Science Association, Bangalore 560 080.
Typeset by WINTECS Typesetters (Ph: 2332 7311), Bangalore 560 021 and Printed at Lotus Printers, Bangalore (Ph: 2320 9909)