Learning Genetic Transposition with an Interactive Computer Program

Xiaoli Yang¹, Yifan Cai² and Charles Tseng³

¹, ² Department of Electrical and Computer Engineering, Purdue University Northwest, 2200 169th street, Hammond, Indiana, USA 46323
³ Department of Biological Sciences, Purdue University Northwest, 2200 169th street, Hammond, Indiana, USA 46323

¹ fyangx@pnw.edu, ² yifancai@me.com, ³ charlestseng219@gmail.com

ABSTRACT
Genetic transposition is a process in which a segment of genetic material moves from one site on a chromosome to another site, which can either be on the same chromosome or on a different chromosome, within the cell. This movable sequence of DNA is called a transposable element, or transposon. Transposition is an important property of life because it creates variations in the genome, which, in turn, allow life to adapt to changes in the environment. Transposable elements are also important in biomedicine, because they may be a vehicle through which bacteria spread antibiotic resistance genes. The study of transposition is difficult because of the heterogeneity of the transposable elements, i.e., there are not only transposons with different structures, but with different modes and mechanisms of moving from one site to another.

Because of its complex nature, transposition is difficult to grasp for many college students. In this paper, we use our platform “GeneAct” to develop an interactive computer program to facilitate and enhance the learning process. Several prokaryotic and eukaryotic transposons are used as examples in the program.

Keywords: Genetic Transposition, Prokaryotic Transposons, Eukaryotic Transposons, Interactive Computer Program.

1. INTRODUCTION

Genomes are dynamic. The cell’s DNA content, while being relatively constant and specific for each species, changes through mutation and recombination so that new genotypes emerge frequently. These processes create changes in the genome that are selected by the continuously changing environment.

Genetic transposition is a special kind of genetic recombination in which a particular sequence of DNA moves from one site on a chromosome to another site on a chromosome (either the same chromosome or a different chromosome). This type of movable DNA sequence is called a transposable element (TE), or transposon. In other words, a transposable element is a DNA sequence that can change its relative position within the genome of a single cell.

Transposable elements were first discovered by McClintock [1] in corn in the 1940’s, although her work was not recognized for many years. Today such elements are found in all groups of organisms. In many eukaryotes, transposons may constitute quite a large portion of the genome. In humans, for example, over 50% of the genome is composed of transposable elements, although the majority of them are defective, i.e., they have lost their ability to move. Transposable elements are especially important in biomedicine because they are thought to help spread antibiotic resistance genes in bacteria, as well as cause certain diseases [2].

Learning the subject of transposition is difficult due to the fact that there are many types of TEs with different molecular structures and mechanisms of transposition [3-6]. These elements are characterized by criteria, such as mode of transposition, mobility of the TEs, and/or whether or not transposition involves an RNA intermediate. To be sure, our understanding of the elements is still incomplete and the details of certain molecular mechanisms in transposition remain to be worked out. In this paper, we describe an interactive computer program that has been successfully used as a personalized tutorial to learn the subject.

2. SOFTWARE FOR THE INTERACTIVE PROGRAM

The interactive computer program is based on our unified framework “GeneAct,” which is designed to standardize and accelerate similar programs in learning genetics [7].
“GeneAct” contains functionalities for programmers to incorporate rich text (text with multi-formats), static images, animations, and interactive content, i.e., it is an application programming interface (API). As such, programmers only need to implement the building blocks and assemble them according to the rules defined in the framework [8]. Specifically, a script language was developed to display rich text, static images, and animations interpreted from a formatted text file. The script language is a lightweight language designed specifically to accelerate content development [9]. It includes a text parser to convert the formatted text into rich text, static images, and animations with corresponding positions, font types, font sizes, and colors – as well as a renderer to display them on the screen. In this way, the programmer only needs to write the content of the formatted text according to the tags defined in the script language, saving much time. The interactive functionalities in the framework define how the program responds to the user’s commands, e.g., dragging an object from one place to another place, pressing a button, or entering an input. These functionalities are organized and managed in the Interaction subsystem of the framework. This subsystem is implemented in collaboration with other subsystems, such as the Action, Record and Scheduler subsystems, under the coordination and supervision of a Director class in the framework. In order to handle multiple interactions efficiently, all the interactive objects are organized in a doubly linked list. Programmers implement the necessary blocks separately and save them in different subsystems. In the framework, the Director coordinates all the block functions.

3. CRITERIA FOR CHARACTERIZING MAJOR TYPES OF TRANSPOSALE ELEMENTS

Transposable elements are heterogeneous in their structure and mechanisms of movement. It is therefore necessary to use certain criteria to characterize the different types of TE. These criteria are introduced here to familiarize users with the basic terminology for understanding the subject.

3.1 Conservative Transposition vs. Replicative Transposition

There are two major modes of transposition. One is referred to as conservative (non-replicative) transposition, where the transposable element moves from one DNA site (original site) to another site (target host site) without replicating itself. The consequence of this mode of transposition is that the TE moves out of the original site and inserts itself into a new site. This process is also called the cut and paste method. The enzyme that is responsible for transposition is called transposase, and it is encoded by a gene within the element.

The other mode of transposition is known as replicative transposition, where the transposable element first replicates itself to form two copies: one copy remains at the original site, while the other copy is inserted into a new site (target). This process is also called the copy and paste method. The transposable elements that use this mode of transposition contain not only the transposase gene, but also the resolvase gene, which is used to resolve the cointegrate (combination of two DNA molecules, each containing a transposable element). In this way, the newly duplicated TE copies are separated, with one remaining at the original site and the other inserted into the target site.

3.2 Autonomous Transposable Elements vs. Non-autonomous Transposable Elements

An autonomous transposable element contains the transposase gene, enabling the element to move from one site to another. In many other transposable elements, however, the transposase gene is missing or defective, so that they have lost the ability to move themselves; these are the non-autonomous transposable elements. Non-autonomous transposable elements may nevertheless utilize transposases encoded by an autonomous element in the same cell, if one exists, to move.

The Ac (activator) transposable element in cells of corn kernels represents an autonomous transposable element, whereas the Ds (dissociation) transposable element represents a non-autonomous transposable element.

3.3 DNA Transposons vs. Retrotansposons

A DNA transposon is one that moves from the original site to the target site without involving an RNA intermediate. The DNA transposon contains a transposase (also called integrase or recombinase) gene and sometimes other genes (e.g., antibiotic resistance genes) flanked by special sequences (e.g., inverted repeats) at both ends, which serve as recombination sites.

Unlike DNA transposons, retrotansposons use an RNA intermediate for transposition. One type of retrotansposon is known as the virus-like retrotansposon because it behaves like a retrovirus (except that it does not move from cell to cell). This element contains two essential genes, one coding for transposase and the other for reverse transcriptase. The latter is used to synthesize cDNA from the RNA intermediate. Flanking the genes are long terminal repeats (LTRs).

Another type of retrotansposon is known as the poly-A retrotansposon. The poly-A retrotansposon contains two genes: ORF1 (open reading frame 1) and ORF2 (open reading frame 2), both of which are involved in RNA binding, reverse transcription, and insertion. These genes are flanked by a 5’UTR (untranslated region) and a 3’UTR, followed by a sequence of A-Ts with the A sequence
serving as a coding strand for the poly(A) tail of the mRNA.

It should be noted that various enzymes are involved in transposition. Some enzymes, such as transposase, resolvase, and reverse transcriptase, are coded by genes within the transposable element, while others, such as RNA polymerase, DNA polymerase, and DNA ligase, are supplied by the cell’s genes outside the transposable sequence.

Below are examples of some transposable elements that the interactive program uses to illustrate the multitude of structures and mechanisms involved in transposition in both prokaryotes and eukaryotes.

4. APPLICATION OF THE INTERACTIVE PROGRAM FOR UNDERSTANDING TRANSPOSABLE ELEMENTS

4.1 Prokaryotic Transposable Elements

Prokaryotic transposable elements are found in bacteria and plasmids. All of them are DNA transposable elements.

A. Simplest Transposable Elements: Insertion Sequences (IS)

The simplest transposable element in prokaryotes (bacteria and plasmids) is the insertion sequence (IS), which has only one gene, or at most a few genes, for mobilization and insertion. Insertion sequences are small (700 to 2500 bp) and contain a gene encoding the enzyme transposase, which is responsible for transposition.

The coding region (gene) in an insertion sequence is usually flanked by inverted repeats. An inverted repeat (IR) is a pair of nucleotide sequences, one further downstream than the other, that are reversed complements of one another (Fig. 1).

![Fig. 1. Example of inverted repeats (arrows show the same nucleotide sequence.)](image1)

Inverted repeats define the boundaries of transposable elements (Fig. 2). They also indicate regions capable of self-complementary base pairing (regions within a single sequence that can base pair with each other).

![Fig. 2. The structure of an insertion sequence](image2)

Another feature of an IS (as well as of more complicated transposons) is the presence of direct repeats (several bases long) flanking both sides of the IR. The direct repeat is not part of the original transposable element itself, but is created from the target DNA site during the process of transposition (Fig. 3).

In short, insertion sequences are short segments of DNA with one or few genes needed for transposition. The genes are flanked by IR’s at both ends, which are further bordered by direct repeats at the target site (Fig. 3).

Insertion sequences, like other TEs, move either by the conservative method or by the replicative method. During conservative transposition, no copying of the element takes place. Transposase is responsible for cleaving and inserting the IS element into the target DNA, while the cell’s DNA polymerase and DNA ligase fills the gap and seals the nick that result, respectively. The general mode of movement is similar to those found in some bacterial composite transposons (e.g., Tn10) to be elaborated in the next section.

![Fig. Error! No text of specified style in document. Generation of direct repeats flanking both ends of the transposon](image3)
for replicating the element into two copies. As in conservative transposition, transposase, in replicative transposition, makes staggered cuts at the target DNA site so that the IS can insert between the newly cut, single-stranded ends. The gaps are filled by the host DNA polymerase and the nicks are sealed by the host DNA ligase. The general mode of movement is similar to those found in some bacterial noncomposite transposons (Tn3) to be illustrated in the next section.

B. Bacterial Transposons

Bacterial transposons are larger and more complicated than IS elements. They not only contain the transposase gene, but other genes, such as antibiotic resistance genes, as well. Bacterial transposons can further be divided into composite transposons and non-composite transposons. The structure of the composite transposon is characterized by the presence of IS elements at either end of the transposon - elements, which carry additional genes, such as antibiotic resistance genes (Fig. 4). The mode of movement is either conservative or replicative.

In conservative transposition, the transposase binds to the ends of the transposon, making four initial cuts. Cuts are made at each end of the TE. One cut is made in a DNA strand at one end of the element, while the other cut is made in the complementary strand at the other end of the element. The other two cuts are made at the target DNA site, one in each strand, to create a staggered break. Joining the 3' ends of the element to the 5' ends of the target site generates a strand-transfer intermediate. The strand-transfer intermediate is released by nicking the ends of the element not initially cut. DNA synthesis and ligation (repair) is restricted to the gap at the flanking direct repeats. In this way, the TE is transferred from the original (donor) site to the target site. Fig. 5 shows stepwise the process of conservative transposition in the composite transposon Tn10.

**Step 1.** Transposases bind to each end of the element in the donor DNA and to a site on the target DNA, making four initial cuts: one at each end of the element (one cut on the upper strand and the other cut on the lower strand) and two at the target site to create a staggered cut.

**Step 2.** The donor DNA is coiled to show the two single-stranded cuts. The staggered cut at the target site is also shown.

**Step 3.** The element’s 3’ ends are joined to the target DNA’s 5’ ends to form a strand-transfer intermediate.

**Step 4.** The inserted element at the target site is straightened to show the two single-stranded gaps at both ends of the element.
Step 5. DNA polymerase fills the gaps, which is followed by ligation.

Fig. 5. The process of conservative transposition

Non-composite transposons do not contain IS elements; they have their own inverted repeats (Fig. 4b). Like composite transposons, non-composite transposons contain both transposase and antibiotic resistance genes. They usually move by the replicative method, and, as a result, contain the resolvase gene. As in conservative transposition, in replicative transposition the transposase makes four initial cuts, two at the element and two at the target site. At each end of the transposon, the 3' end of one strand is joined to the 5' extension of one strand at the target site. Ligation of the 3' ends of the element to the 5' ends of the target site forms a strand-transfer intermediate, joining the donor and target DNA molecules. The 3' ends of each strand at the target site (at the staggered cuts) serve as primers for synthesizing complementary DNA strands from the single stranded DNA template of the transposon. Replication followed by ligation leads to the formation of a cointegrate, which is a structure formed by the donor DNA molecule and the target DNA molecule. The cointegrate is then resolved by resolvase into two separate molecules, each containing a copy of the element. Fig. 6 shows stepwise the process of replicative transposition by the transposon Tn3.

Step 1. The transposases arrive at both ends of the TE and the target site.

Step 2. The transposases bind to both ends of the TE and the target site.

Step 3. The transposases make 4 cuts.
Step 4. The TE strands are joined to the corresponding strands from the target DNA.

Step 5. Using the free 3’ ends of the target DNA as primers, DNA polymerase replicates the TE, creating double stranded DNA, which is ligated.

Step 6. Resolvase separates the cointegrate, yielding two DNA molecules.

Step 7. Replicative transposition is completed and each DNA molecule contains a TE.

4.2 Transposable Elements in Eukaryotes

Unlike bacterial transposons, which are all DNA transposons, eukaryotic transposable elements may be DNA transposons or retrotransposons. The Ac-Ds system (DNA transposons) in corn will be used to illustrate autonomous and non-autonomous transposition in eukaryotes.

a) Autonomous and Non-autonomous Transposable Elements (Ac-Ds system in Corn)

Plant transposons also have IR sequences and direct repeats at the target site. As usual, these sequences flank the ends of the TE. The transposons can be autonomous or non-autonomous. Well-known examples include the Ac-Ds system in corn discovered by Barbara McClintock [1]. In this system, the Ac element is an autonomous transposon, which contains a functional transposase gene, while the Ds element is a non-autonomous transposon, which contains a defective transposase gene (Fig. 7). Thus, the transposition of the Ds element requires the use of the transposase encoded by the Ac element. Fig. 8 illustrates stepwise how the Ds element is driven by the Ac element, as well as how the Ac-Ds system affects corn kernel colors.
Step 1. The C gene (on chromosome 9), which is responsible for color development in the cells of a kernel, is activated.

Step 2. A Ds element is inserted into the C gene, changing it into a mutant (c gene) and resulting in the development of a colorless kernel.

Step 3. The transposase encoded by the Ac element removes the Ds element, reversing the mutation (the c allele returns to the wild type C allele). This explains the patches of color.

b) Retrotransposons

Retrotransposons (or RNA based transposons) are common in eukaryotes. They exist in many different forms. The virus-like retrotransposons may have originated from a degenerate virus, since they contain long terminal repeats (LTR) and genes such as the reverse transcriptase gene, as well as other genes that are often found in retroviruses (Fig. 9). Examples are the Ty element in yeast and the copia element in Drosophila.
Fig. 9. Structure of retrotransposons: a) virus-like retrotransposon and b) poly-A transposon

Fig. 10 illustrates stepwise the process of transcription, RNA processing, and cDNA synthesis. The cDNA from the TE element is then inserted into the target site.

Step 1. RNA polymerase travels to the promoter of the retrotransposon to initiate transcription (formation of pre-mRNA).

Step 2. RNA is processed by the spliceosome: introns are removed and exons are joined to form mRNA.

Step 3. Reverse transcriptase synthesizes cDNA from mRNA.

Step 4. cDNA is inserted into the target DNA by transposase (not shown).

Step 5. DNA polymerase fills in the gap, which is followed by ligation.
Another common transposon is the poly-A retrotransposon. As described earlier, a retrotransposon contains two ORFs (ORF1 and ORF2), which are involved in RNA binding, reverse transcription, and insertion. These genes are flanked by a 5'UTR and a 3'UTR, followed by a sequence of A-Ts with the A sequence serving as the coding strand for the poly(A) tail of the mRNA (Fig. 9). Examples includes LINEs (long interspersed elements) and SINEs (short interspersed elements).

LINEs are commonly found in eukaryotes. Their genes are transcribed into RNA using the cell’s RNA polymerase II. The reverse transcriptase coded by LINEs has greater specificity for its own RNA than for other RNA in the cell. The 5' UTR contains the promoter sequence, while the 3' UTR contains a polyadenylation signal (AATAAA) and a poly-A tail. Fig. 11 shows stepwise the transposition of a poly-A retrotransposon. The target site is used to prime reverse transcription. This replicative transposition mechanism is responsible for the rapid enlargement of the genome. Indeed, LINEs constitute about 17% of the human genome, of which only a small portion is active.

**Step 1.** The RNA polymerase travels to the promoter (P) for transcription.

**Step 2.** The mRNA arrives at the ribosome for translation.

**Step 3.** ORF1 and ORF2 form a protein complex.

**Step 4.** The ORF1-ORF2 protein complex travels to the mRNA.
Step 5. The ORF1-ORF2 complex carries the mRNA to the target DNA for cleavage and A-T base pairing.

Step 6. The protein complex synthesizes the first strand of the cDNA before the second strand.

Step 7. The cDNA and the target DNA are joined through the usual repair process.

Fig. 11. Transposition of a poly-A retrotransposon using the target site primed reverse transcription method

SINEs are similar to LINEs, but smaller. They do not encode a functional reverse transcriptase and thus rely on other mobile elements for transposition. The most common SINEs in primates are the \textit{Alu} sequences. \textit{Alu} elements are approximately 300 base pairs long, do not contain any coding sequences, and can be recognized by the restriction enzyme \textit{AluI} (hence the name). SINEs make up approximately 10\% of the human genome.

5. CONCLUSIONS

Transposable elements are present in all groups of organisms. Their structures and mechanisms of movement vary greatly, however. Transposable elements move to sites within and between the chromosomes. They also move between plasmids and bacterial chromosomes.
Since the key function of a transposable element is to move DNA from one site to another in the genome, transposable elements have historically been considered “junk DNA” or “selfish genes.”

How transposable elements originated is still unclear. There is evidence that some TEs may have originated from a common, universal ancestor, while others are thought to have been derived independently. Retrotransposons are similar to viruses in both structure and mechanism of action. Thus, it is conceivable that they share a common ancestor.

The mobility of transposable elements can significantly change the genome, creating variations for selection. The movement of TEs, however, may also disrupt genes. When TEs insert into coding or regulatory sequences, they may be quite harmful. In fact, the distribution of these elements has been implicated in genetic diseases such as hemophilia, Duchenne muscular dystrophy, and certain cancers.

The numerous copies of many transposons may be a factor in the imprecise pairing of homologous chromosomes during meiosis, resulting in unequal crossing over and chromosomal abnormalities such as deletions and duplications. Recombination between two elements in the same orientation results in a deletion, while recombination between two elements in opposite orientations generates an inversion.

Though many transposition events may be quite harmful, some transposition events may be beneficial. Bacterial transposons, for example, contain antibiotic resistance genes, which can spread through transposition to increase a bacterium’s resistance to antibiotics. In vertebrates, the mechanisms of transposition may have been adopted for use in the immune system (V(D)J recommendation system).

Transposable elements may constitute a great portion of the genome. This is especially true in large eukaryotic genomes. Replicative transposable elements are probably responsible for the increase in genome size in many species during the course of evolution. Since TEs are generally considered mutagens, proliferation of these elements is frequently detrimental to the organism. To counter the negative effects of TEs, mechanisms may have evolved to silence TEs at the mRNA level.

Despite major efforts, the detailed mechanisms of transposition remain to be worked out. Due to the heterogeneous nature of the elements, the study of this subject is certainly challenging. Understanding transposition is especially difficult for novices. This interactive program on transposition attempts to facilitate learning is suitable for undergraduate biology students, as well as advanced high school students interested in pursuing further studies in the field.

REFERENCES


AUTHOR PROFILES:

Xiaoli Yang is currently a professor in the Department of Electrical & Computer Engineering, Purdue University Northwest, USA. Her research interests are in virtual reality, software engineering, and its application in cognitive research and education.

Yifan Cai is currently a Load and Performance Software Engineer at Apple Inc. He received his Master of Engineering degree from the Department of Electrical & Computer Engineering, Purdue University Northwest, USA.

Charles Tseng is a Professor Emeritus of Biological Sciences, Purdue University Northwest, USA.