Malika Sharma

Dr. Naowarat Cheeptham (Ann)

BIOL 4130

April 7, 2023

Transposable Elements: A Means for Estimating Genomic Diversity

Introduction:

Evolution is the unifying principle of all biology and biological organisms. Because understanding evolution can help us solve biological problems that impact our lives, it is crucial that we learn as much as we can about this process. Understanding our evolution has been an ongoing investigation that researchers around the world have been trying to do for many years. A huge step in this investigation was made when we analyzed the results from the Human Genome Project in 2003. This project gave us new insight on how we came to be and simultaneously emphasized the biological importance of transposable elements.

Transposable elements (TEs), also known as "jumping genes," are DNA sequences that move from one location on the genome to another and were first identified more than 50 years ago by geneticist Barbara McClintock (Pray, 2008). In the next several decades it became apparent that not only do TEs "jump," but they are also found in almost all organisms and approximately 45% of the human genome is comprised of TEs (Ayarpadikannan and Kim, 2014). Many studies have revealed that TEs play various roles in processes, including genome evolution, gene expression regulation, genetic instability, and cancer disposition. TEs are a subject of interest worldwide, not only in terms of their clinical aspects but also in basic research, such as evolutionary tracking (Ayarpadikannan and Kim, 2014).

What are transposable elements?

TEs are DNA sequences that can integrate into the genome at a new site within the cell of its origin. Sometimes, the change in their positions creates or reverses mutations, thereby altering the cell's genotype (Ayarpadikannan and Kim, 2014). TEs consist of two major classes: DNA transposons and retrotransposons. DNA transposons can move and inserting into new genomic sites (cut-and-paste mechanism). Retrotransposons replicate by forming RNA intermediates, which are then reverse-transcribed to make DNA sequences and inserted into new genomic locations (transposition) (Ayarpadikannan and Kim, 2014). DNA transposons consist of a transposase gene that is flanked by two *Terminal In-verted Repeats* (TIRs) (Fig. 1) (Muñoz-Lópezand and García-Pérez, 2010).

Figure 1. A transposable element consists of a transposase gene flanked by Terminal Inverted Repeats (Muñoz-Lópezand and García-Pérez, 2010).

Detailed Protocol (Muñoz-Lópezand and García-Pérez, 2010)***:***

1. Two transposase enzyme molecules recognize the TIRs and bind to them, forming the *Single-End Complex* (SEC) (Fig. 2).
2. Both transposases cleave the 5'-ends of the TIRs by hydrolysis to liberate the non-transferred strands (5'-P extremes) (Fig. 3).
3. The two transposase molecules interact and bring together the transposon ends to form the *Paired-End Complex* (PEC) generating a transposase dimer (Fig. 2). At this point, the phosphodiester bond undergoes a hydrolysis to produce the transferred strands (3'-OH extremes) (Fig. 3).

4. The PEC binds to target DNA forming the *Target Capture Complex*, at which insertion takes place (Fig. 2). The target TEs is any TA dinucleotide. The 5'-end in the target DNA undergoes a nucleophilic attack from the transposon *transferred strands* 3'-OH. The gaps in the transposon 5'-ends are filled by the host, generating canonical Target Site Duplications (TSDs) flanking the new transposon insertion (Fig. 3).

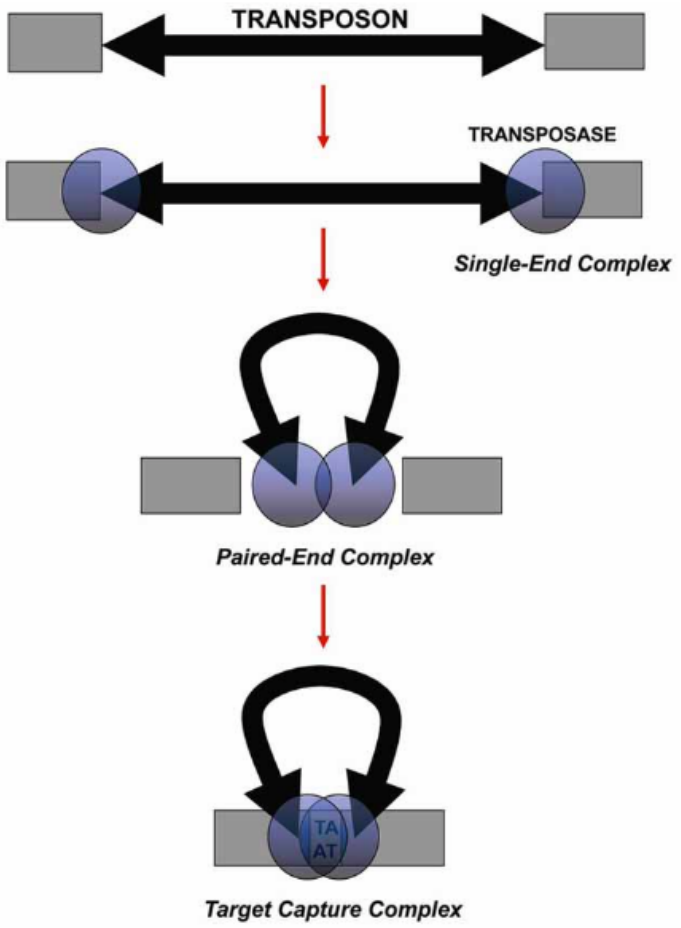

Figure 2. Representation of the transposition mechanism performed by the transposable elements (Muñoz-Lópezand and García-Pérez, 2010).

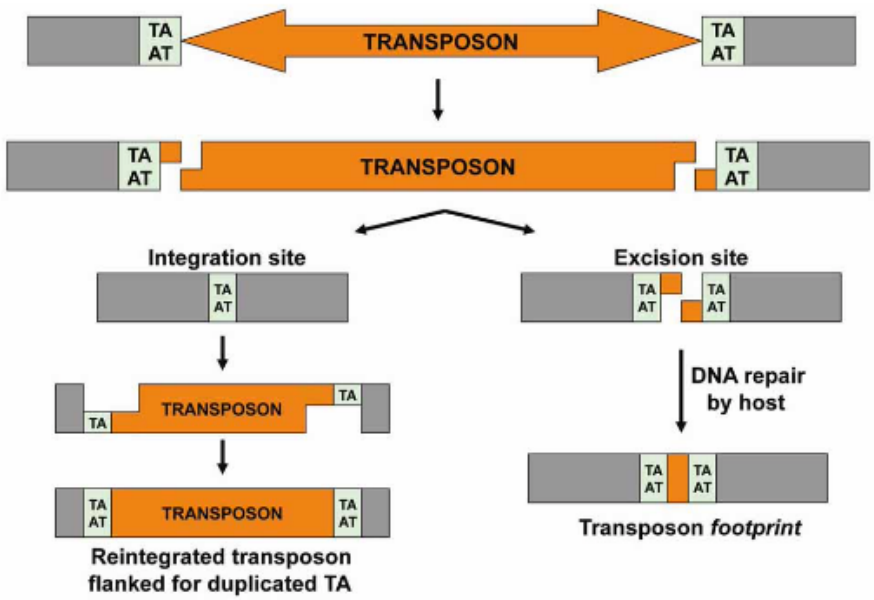

Figure 3. Representation of the cut-and-paste mechanism in which the transposon is excised from one site and integrated at a TA target dinucleotide (Muñoz-Lópezand and García-Pérez, 2010).

Objectives of Transposable Elements:

Notably, structural variations (insertion, deletion, inversion, duplication, translocation) in the human genome are the primary cause of inter-individual variability (Ayarpadikannan and Kim, 2014). When a TE inserts within an exon, it may change the codes for an irregular peptide, or it may even cause missense or nonsense mutations. On the other hand, if it is inserted into an intronic region, it may cause an alternative splicing event by introducing novel splice sites, disrupting the established splice site (Ayarpadikannan and Kim, 2014). In some instances, TE insertion into intronic regions can even cause mRNA destabilization, reducing gene expression. The many possibilities that can occur from the insertion/deletion of a TE shows how important TEs are to mutations.

Application and effectiveness:

There are many applications of TEs that can help us learn more about evolution and our biology. For example, studies have proposed that specific TE insertions have created variants that can potentially be used as DNA markers in human population studies, as well as in forensic analyses (Ayarpadikannan and Kim, 2014).

Moreover, being able to identify TEs in a genome is of great use to evolutionary biologists. By identifying the location of TEs in a genome, we can understand and estimate the biological changes that have occurred overtime to an organism and species. One way that this process has been done is by Polymerase Chain Reaction (PCR). For instance, to find active retro-transposons, PCR amplification using primer pairs on conserved *pol* gene of rice was performed. The specific amplification of *pol* gene in the callus of rice led to the detection of active retrotransposon *Tos17* (Miyao and Yamanouchi, 2022).

Another tool used by biological and computational scientists to detect TE in the genome and transposition events of TE has been using next generation sequence (NGS) data. However, detection of newly transposed events by TEs from NGS data is difficult, due to their multiple distribution sites over the genome containing older TEs. The previously used mechanism, *Transposon Insertion Finder* (TIF), detects TE transpositions on the reference genome from NGS short reads using end sequences of target TE (Miyao and Yamanouchi, 2022). However, TIF requires the sequence of target TE and is not able to detect transpositions for TEs with an *unknown* sequence (Miyao and Yamanouchi, 2022). Therefore, it has been difficult to detect TE transposition by the standard NGS analysis flow because multiple copies of TEs are already on the chromosome. But fortunately, researchers found a new tool that can be used instead. The new

algorithm Transposable Element Finder (TEF) enables the detection of TE transpositions, even for TEs with an unknown sequence (Miyao and Yamanouchi, 2022).

Advantages and Limitations:

The characteristics of TEs, such as abundance in the genome, high sequence identity, and ability to move, make them major contributors to genomic instability and variety (Ayarpadikannan and Kim, 2014). TE mobilization can promote gene inactivation, modulate gene expression, or induce illegitimate recombination. In other words, as helpful as TEs are to understand our genome and thus our evolution, it does also have some drawbacks. Both insertions and excisions of TEs can cause genomic instability, thus causing many human diseases, including genetic disorders, psychiatric problems, and cancer (Ayarpadikannan and Kim, 2014). Although some TE insertions have caused harmful mutations, most have contributed to genetic diversity.

Some TEs are reported to be the cause of many genetic disorders, such as hemophilia, Apert syndrome, familial hypercholesterolemia, and colon and breast cancer. For example, the breast cancer 2 (BRCA2) gene, associated with breast and ovarian cancers, has been reported to be disrupted by multiple TE insertions (Ayarpadikannan and Kim, 2014).

Other times, TEs have no effect on their host. Strategies have been developed by host and TEs to minimize the deleterious impact of transposition, and to reach equilibrium (Muñoz-Lópezand and García-Pérez, 2010). For example, some transposons tend to insert in nonessential regions in the genome, such as heterochromatic regions, where insertions will likely have a minimal deleterious impact.

Conclusions:

TE insertions contributed markedly to variation and increased the speed of evolution. Furthermore, they increase the recombination rate, in addition to affecting genes and their expression. Thus, DNA transposons are useful tools to analyze the regulatory genome, identify genes and pathways implicated in disease or pathogenesis, and even contribute to gene therapy. It is of great importance that researchers continue to conduct in-depth studies on the role of TEs in evolution and clinical aspects, and the epigenetic control of gene expression as it will be of paramount importance in uncovering novel mechanisms that can be targeted for therapeutic intervention (Ayarpadikannan and Kim, 2014).

References:

- Ayarpadikannan, S., and Kim , H.S. 2014. The Impact of Transposable Elements in Genome Evolution and Genetic Instability and Their Implications in Various Diseases. Genomics and Information 12(3): 98-104.
- Miyao, A., and Yamanouchi, U. 2022. Transposable element finder (TEF): finding active transposable elements from next generation sequencing data. BMC Bioinformatics 23:500.
- Muñoz-López, M., and García-Pérez, J.L. 2010. DNA Transposons: Nature and Applications in Genomics. Current Genomics 11, 115-128.
- Pray, L.A. 2008. Transposons: The Jumping Genes. Nature Education 1(1):204