NEW KIDS ON THE BLOCK:
SELF-SYNTHESIZING DNA TRANSPOSONS

Kapitonov V.V.*, Jurka J.
Genetic Information Research Institute, Mountain View, California, USA
*Corresponding author: e-mail: vladimir@girinst.org

Key words: transposons, molecular evolution, genomics, DNA polymerase, integrase, cysteine protease, ATPase, computational biology

SUMMARY

Motivation: Transposable elements constitute quite significant static and dynamic components of eukaryotic genomes. Moreover, transposable elements serve as efficient tools in genetic engineering. Therefore, identification and studies of transposable elements are important for researchers working in different fields of molecular biology.

Results: Based on computational studies of eukaryotic genomes, we discovered a novel class of DNA transposons, called Polintons, characterized by a unique set of proteins including a protein-primed family B DNA polymerase, retroviral integrase, cysteine protease, and ATPase. Polintons are also characterized by 6-bp target site duplications, long terminal inverted repeats, and 5'-AG and TC-3' termini. According to a transposition model discussed here, a Polinton DNA molecule excised from the genome by the Polinton-encoded integrase serves as a template for extrachromosomal synthesis of its double-stranded DNA copy by the Polinton-encoded DNA polymerase and is inserted into genome by the integrase.

Availability: http://www.girinst.org/repbase/.

INTRODUCTION

Genomes of most eukaryotes are populated by DNA copies of parasitic elements known as transposable elements (TEs) capable of reproducing themselves in the host genome in a non-Mendelian fashion. Despite an enormous diversity of eukaryotic TEs, they belong to only two types called retrotransposons and DNA transposons (Craig et al., 2002). While a retrotransposon is transposed via reverse transcription of its mRNAs, a DNA transposon is transposed via transfer of its genomic copy from one genomic site to another. Transposition of a retrotransposon is catalyzed by reverse transcriptase (RT) and endonuclease (EN) domains of a polyprotein encoded by itself or by other retrotransposons. All retrotransposons can be further divided into two subclasses called LTR and non-LTR retrotransposons. An mRNA molecule expressed during transcription of the genomic non-LTR retrotransposon is reverse transcribed and inserted in the genome. An LTR retrotransposon may carry three open reading frames (ORFs) coding for the gag, env, and pol proteins, the latter is composed of the RT, EN, and aspartyl protease domains. The endonuclease domain in LTR retrotransposons is usually called integrase (INT) and is distantly related to the DDE transposases (TPase) encoded by Mariner DNA transposons (Capy et al., 1997).

DNA transposons identified so far in eukaryotes belong to two classes characterized by the so-called “cut and paste” (Craig, 1995) and “rolling-circle” (Kapitonov, Jurka, 2001) mechanisms of transposition. Unlike retrotransposons, which synthesize their DNA
copies using their own RNA-dependent DNA polymerase (RT), DNA transposons cannot synthesize DNA. Instead, they multiply using the host replication machinery (Kapitonov et al., 2004). A typical autonomous mariner, hAT, piggyBac, P, Merlin, Transib DNA transposon encodes a single protein called transposase, which acts as an endonuclease and catalyses transfer of transposon DNA strands from one genomic site to another (Craig et al., 2002; Kapitonov, Jurka, 2003; Feschotte, 2004; Kapitonov, Jurka, 2005). In the En/Spm, MuDR, Harbinger, and Helitron superfamilies, an autonomous transposon usually encodes TPase and DNA-binding proteins (Kapitonov, Jurka, 2001; Craig et al., 2002; Kapitonov, Jurka, 2004).

RESULTS AND DISCUSSION

Polintons are widespread in protists, fungi and animals including entamoeba, trichomonas, soybean rust, sea urchin, sea anemone, sea squirt, fishes, chicken, lizard, frog, insects, and worms (Kapitonov, Jurka, 2006). Autonomous Polintons are 10–20 kb long and encode up to ten different proteins including a family B DNA polymerase (POLB), retroviral-like integrase, adenoviral-like protease (PRO), and putative ATPase (ATP). Polintons are the most complex DNA transposons in eukaryotes. Based on structural and evolutionary characteristics of these transposons, we developed a model of Polinton transposition. We also discuss implications of our findings, including likely origin of Polintons from a linear plasmid and evolution of adenoviruses from an ancient Polinton.

Given known distant similarities between retroviral INTs and “cut and paste” TPases, one can expect “cut and paste” transpositions of Polintons catalyzed by their INT. However, arguments listed below strongly suggest that transposition of Polintons follows a completely different mechanism unseen previously in transposons. First, a perfect conservation of all functional motifs in the extremely diverged POLBs indicates that the DNA-DNA polymerase and proofreading activities are necessary for transposition of Polintons. Second, POLB in Polintons belongs to the group of protein-primed DNA polymerases encoded by genomes of bacteriophages, linear plasmids and adenoviruses. Third, all these genomes and Polintons are characterized by terminal inverted repeats that are usually several hundred bps long. Fourth, their termini are composed of short 1-3-bp tandem repeats, which are thought to be necessary for the slide-back mechanism in protein-primed DNA synthesis (Mendez et al., 1992).

Table 1. General properties of Polintons in different species

<table>
<thead>
<tr>
<th>Species</th>
<th>Copies per genome</th>
<th>Encoded proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frog, Zebrafish</td>
<td>~100</td>
<td>POLB, INT, PRO, ATP, PW, PY, PZ</td>
</tr>
<tr>
<td>Lancelet, Sea squirt/urchin</td>
<td>50–100</td>
<td>POLB, INT, PRO, ATP, PW, PY, PZ</td>
</tr>
<tr>
<td>Red flour beetle</td>
<td>~50</td>
<td>POLB, INT, PRO, ATP, PW, PY, PZ</td>
</tr>
<tr>
<td>Nematode</td>
<td>~100</td>
<td>POLB, INT, PRO, ATP, PW, PY, PZ</td>
</tr>
<tr>
<td>Nematostella vectensis</td>
<td>~100</td>
<td>POLB, INT, PRO, ATP, PW, PY, PZ</td>
</tr>
<tr>
<td>Trichomonas vaginalis</td>
<td>~1000</td>
<td>POLB, INT, ATP, ATP1, PTV1-PTV6</td>
</tr>
<tr>
<td>Entamoeba invadens</td>
<td>50–100</td>
<td>POLB, INT, ATP, ATP1, PTV6</td>
</tr>
</tbody>
</table>

We proposed that Polintons form a novel class of DNA transposons propagated through protein-primed self-synthesis by POLB, according to the next model. First, during host genome replication, the integrase-catalyzed excision of Polinton from the host DNA leads to an extrachromosomal single-stranded Polinton that forms a racket-like structure. Second, the Polinton POLB replicates the extrachromosomal Polinton. Given the arguments listed above, initiation of the replication requires the terminal protein (TP) that binds a free 5’ end of Polinton. It is thought that N-terminal domains of proteins, whose C-terminal parts serve as POLB, encode TP in some linear plasmids. Therefore, it
is likely the N-terminal 400–600-aa domain of the Polinton POLB serves also as TP. After the double-stranded Polinton is synthesized, the INT molecules bind its termini and catalyze its integration in the host genome.

Polintons are present in genomes of species that belong to diverse eukaryotic kingdoms including opisthokonts (metazoa and fungi), heterokonts (oomycetes), alveolates (ciliates), amoebozoa (entamoeba), and parabasalids. Given the conserved complex structure of Polintons, their monophyletic origin is most likely. Although Polintons are much more complex (up to eight conserved proteins) than known eukaryotic TEs and resemble viruses (adenoviruses and BmDNV-2), we did not find any Polinton protein similar to viral capsid or envelope proteins, which are necessary for the infectious transmission of viruses. Moreover, we are not aware of any viruses capable of spreading over different kingdoms. Most likely, Polintons emerged in a common ancestor of modern species from the eukaryotic crown, approximately a billion years ago. As we reported here, Polintons share their main structural characteristics with “selfish” linear plasmids, bacteriophages and adenoviruses that multiply using their protein-primed DNA polymerases. Linear plasmids can be split into two groups: (i), plasmids that exist in mitochondria of plants and fungi; (ii), plasmids that exist in the yeast cytoplasm. While it is likely that mitochondrial linear plasmids evolved from bacteriophages during the evolution of mitochondria from bacteria, different equally plausible scenarios puzzle understanding of the evolution of cytoplasmic plasmids. Although Polintons represent a previously unknown link between cytoplasmic plasmids/adenoviruses and mitochondrial plasmids/bacteriophages, many aspects of evolution of Polintons and cytoplasmic linear plasmids remain unclear. Acquisition of the integrase by a protein-primed replicating genome of an ancient virus or linear plasmid was the most certain stage of the evolution. The Polinton INT has evolved from an INT encoded by an LTR retrotransposon. Thus, it might have been acquired after integration of an ancient LTR retrotransposon into the ancestral linear genome. However, we cannot rule out the origin of the Polinton INT from a DNA transposon. For instance, the Tdd-4 transposon from the slime mould Dictyostelium discoideum genome is a DNA transposon characterized by its 145-bp TIRs, 5-bp TSDs, and a TPase that is similar to INTs encoded by LTR retrotransposons (Wells, 1999).

Polintons are characterized by a highly patchy distribution in different species. In insects, Polintons are present in flies and beetles but absent in mosquitoes and bees. In fungi, they are present in basidiomycetes (soybean rust) and glomeromyces (G. intraradices) but absent in ascomycetes (including the completely sequenced yeast genome). We interpret this patchiness as a frequent loss of Polintons from genomes. Due to the high complexity of Polintons, their transposition may be tightly regulated and may explain their small numbers in most studied genomes.

ACKNOWLEDGEMENTS

This work was supported by the National Institutes of Health grant 5 P41 LM006252–08.

REFERENCES

Capy P. et al. (1997) Do the integrases of LTR-retrotransposons and class II element transposases have a common ancestor? Genetica, 100, 63–72.


