14.5: On the Evolution of Transposons, Genes, and Genomes

We noted that transposons in bacteria carry antibiotic resistance genes, a clear example of benefits of transposition in prokaryotes. Of course, prokaryotic genomes are small, as is the typical bacterial transposon load. Yeast species also have low transposon load. But, what can we make of the high transposon load in eukaryotes?

To many, the fact that genes encoding proteins typically represent only 1-2% of a eukaryotic genome meant that the rest of the genome was informationally non-essential. Even though transposons turn out to be much of the non-coding DNA in some eukaryotic genomes, they seemed to serve no purpose other than their own replication. For many organisms, large amounts of transposon DNA were dubbed selfish DNA and their genes, selfish genes.

Are transposons just junk DNA, some kind of invasive or leftover genomic baggage? Given their propensity to jump around and potential to raise havoc in genomes, how do we tolerate and survive them? Is the sole ‘mission’ of transposons really just to reproduce themselves? Or are transposons tolerated because they are neither selfish nor junk? By their sheer proportions and activity in eukaryotic genomes, we will see that transposons have dispersed into, and re-shaped genomic landscapes. Do the consequences of transposition (relocation dispersal through a genome, structural alteration and mutation of genes) have any functional or evolutionary value? While all of these questions are a reasonable response to the phenomena of jumping genes, a rational hypothesis would be that, like all genetic change, transposons began by accident. But, their spread and ubiquity in genomes of higher organisms must in the long term have been selected by virtue of some benefit that they provide to their host cells and organisms. Let’s briefly looks at the evolutionary history of transposons to see if this assumption has some merit.

A. A Common Ancestry DNA and RNA (i.e., All) Transposons

Transposases catalyze the cut-and-paste as well as the replicative transposition of Class II (DNA) transposons. Integrases catalyze insertion of reverse-transcribed retrotransposons. Bottom line: both enzymes end by catalyzing
insertion of transposons into new DNA locations. So, it should not be surprising that class I and II transposons enzymes share similar amino acid sequence and domain structures. These similarities support a common ancestry of class I and II transposons. Sequence comparisons of transposable elements themselves reveal that they comprise distinct families of related elements.

This allows us to speculate on the origins of these families in different species. For example, the TC1/mariner (DNA) transposon is found in virtually all organisms examined (except diatoms and green algae). Based on sequence analysis, there is even an insertion element in bacteria related to the mariner element. This amount and diversity of conservation bespeaks an early evolution of the enzymes of transposition, and of transposition itself, within and even between species. Linear descent, or the ‘vertical’ transmission of transposons from parents to progeny, is the rule. However, the presence of similar transposons in diverse species is best explained by interspecific DNA (“horizontal”) transfer. That is, a transposon in one organism must have been the ‘gift’ of an organism of a different species! This is further discussed below. Clearly, moveable genes have been a part of life for a long time, speaking more to an adaptive value for organisms than to the parasitic action of a selfish, rogue DNA!

B. Retroviruses and LTR Retrotransposons Share a Common Ancestry

The ‘integration’ domain of retrotransposons and retroviruses share significant similarities as shown below.

The question raised by these observations is: Did transposons (specifically retrotransposons) arise as defective versions of integrated retrovirus DNA (i.e., reverse transcripts of retroviral RNA)? Or, did retroviruses emerge when retrotransposons evolved a way to leave their host cells. To approach this question, let’s first compare mechanisms of retroviral infection and retrotransposition.

In addition to the structural similarities between the enzymes encoded by retrotransposon and retroviral RNAs, LTR retrotransposons and retroviruses both contain flanking long terminal repeats. However, retrotransposition occurs within the nucleus of a cell while retroviruses must first infect a host cell before the retroviral DNA can be replicated and new viruses produced (check out Visualizing Retroviral Infection to see how immunofluorescence microscopy using antibodies to singlestranded cDNAs was used to track the steps of HIV infection!). A key structural difference between retrotransposons and most retroviruses is an ENV gene-encoded protein envelope surrounding retroviral DNA. After infection, the incoming retrovirus sheds its envelope proteins and viral RNA is reverse transcribed. After the reverse transcripts enter the nucleus, transcription of genes and translation of enzymes here.

• Retroviral DNA, like any genomic DNA, is mutable. If a mutation inactivates one of the genes required for infection and retroviral release, it could become an LTR retrotransposon. Such a genetically damaged retroviral integrate might still be transcribed and its mRNAs translated. If detected by its own reverse transcriptase, the erstwhile viral genomes would be copied. The cDNAs, instead of being packaged into infectious viral particles, would become a
source of so-called endogenous retroviruses (ERVs). In fact, ERVs exist, making up a substantial portion of the mammalian genome (8% in humans)… and do in fact, behave like LTR retrotransposons!

- Yeast TY elements transcribe several genes during retrotransposition (see the list above), producing not only reverse transcriptase and integrase, but also a protease and a structural protein called Gag (Group-specific antigen). All of the translated proteins enter the nucleus. Mimicking the retroviral ENV protein, the Gag protein makes up most of a coat protein called VLP (virus-like particle). VLP encapsulates additional retrotransposon RNA in the cytoplasm, along with the other proteins. Double-stranded reverse transcripts (cDNAs) of the viral RNA are then made within the VLPs. But, instead of bursting out of the cell, the encapsulated cDNAs (i.e., new retrotransposons) shed their VLP coat and re-enter the nucleus, where they can now integrate into genomic target DNA. Compare this to the description of retroviral infection. During infection, retroviral envelope proteins attach to cell membranes and release their RNA into the cytoplasm. There, reverse transcriptase copies viral RNA into double-stranded cDNAs that then enter the nucleus where they can integrate into host cell DNA. When transcribed, the integrated retroviral DNA produces transcripts that are translated in the cytoplasm into proteins necessary to form an infectious viral particle. The resulting viral RNAs are encapsulated by an ENV (envelope) protein encoded in the viral genome. Of course, unlike VLP-coated retrotransposon RNAs, the enveloped viral RNAs do eventually lyse the host cell, releasing infectious particles. Nevertheless, while VLP coated Ty elements are not infectious, they sure do look like a retrovirus!

Common mechanisms of retrovirus and retrotransposon replication and integration clearly support their common ancestry, but they do not indicate origins. On the one hand, the origin of ERVs from retroviruses might imply an origin of retrotransposons from retroviruses. On the other, transposons have been around since the earliest prokaryotic cells, but that retrotransposons arose with eukaryotes. In that case, Type II (DNA) transposable elements were around before retroviruses.

The phylogenetic analysis below is based on comparisons of retroviral and retrotransposon reverse transcriptase gene DNA sequences.
Comparisons of aligned DNA sequences permit evolutionary analyses that reflect phylogenetic relationships of genes (in this case, retrotransposon and viral genes), in much the same way the evolutionary biologists historically demonstrated evolutionary relationships of plants and animals by comparing their morphological characteristics. The data in the analysis supports the evolution of retroviruses from retrotransposon ancestors. From the ‘tree’, TY3 and a few other retrotransposons share common ancestry with Ted, 17.6 and Gypsy ERVs (boxed) in the “Gypsi-TY3 subgroup”. Further, this sub-group shares common ancestry with more distantly related retroviruses (e.g., MMTV, HTLV…), as well as the even more distantly related (older, longer diverged!) Copia-TY1 transposon sub-group. This and similar analyses suggest strongly that retroviruses evolved from a retrotransposon lineage [For a review of retroposon/retrovirus evolution, check Lerat P. & Capy P. (1999, Retrotransposons and retroviruses: analysis of the envelope gene. Mol. Biol. Evol. 19(9): 1198-1207).

C. Transposons Can Be Acquired by "Horizontal Gene Transfer"

As noted, transposons are inherited vertically, meaning that they are passed from cell to cell or parents to progeny by reproduction. But they also may have spread between species by horizontal gene transfer. This just means that organisms exposed to DNA containing transposons might inadvertently pick up such DNA and become transformed as the transposon becomes part of the genome. Accidental mobility of transposons between species would have been rare, but an exchange of genes by horizontal gene transfer would have accelerated with the evolution of retroviruses. Once again, despite the potential to disrupt the health an organism, retroviral activity might also have supported a degree genomic diversity useful to organisms.

255 Transposon Evolution