Literature Review for Current Knowledge of Transposable Element Evolution in Plants

Honours Thesis - Drew Lindsay

I. BACKGROUND

Transposable elements (TEs) are DNA sequences that can move and replicate within a genome. These elements are ubiquitous in all living organisms, and make-up nearly half of the human genome, but can reach as high as 90% some organisms, such as maize (Mills et al., 2007; SanMiguel et al., 1996). Furthermore, these figures are likely an underestimate as many ancient TEs may be unrecognizable due to mutation (Forterre, Filée and Myllykallio, 2004). For many decades after their discovery in the 1940's by Barbara McClintock, TEs were thought to be 'junk' DNA with no functional role in the host genome (Pennisi, 2012). Some even considered them to be parasitic, as their persistence and proliferation in the genome can disrupt the host genome (Orgel and Crick, 1980). However, recent research has shown that TEs are significant roles in gene regulation (Wang et al., 2013), rearrangement (Xuan et al, 2016) and recombination (Hallet, 1997), as well as mutagenic agents to disrupt gene function (Martienssen, 1998), or generate new functionality (Leprince et al., 2001), even from very early embryogenesis (Ge, 2016).

II. TE CLASSIFICATION

A. Class 1 TEs

These TEs, also known as retrotransposons, use an RNA intermediate to facilitate replication and movement in the genome (often referred to as a 'copy-and-paste mechanic'). This gave rise to the original notion of TEs being 'parasitic' (Orgel and Crick, 1980). If left unchecked, retrotransposons would eventually overwhelm the genome, and creating massive numbers of repeats with no functional purpose. DNA replication would become much costlier, and confer a significant disadvantage to the host (Gbadegesin, 2012). However, mechanisms exist to limit TE proliferation.

B. Class 2 TEs

The mechanism for transposition of these TEs, also known as DNA transposons, lack an RNA intermediate. This prevents direct copying, and giving a mechanism referred to as 'cut-and-paste' for movement. Because of this, they are typically less frequently in the genome (2% of all TEs in humans; 0.2% in *Saccharomyces cerevisiae*; 0.7% in *Drosophila melanogaster*), but can be widely variable even in closely related species (0.4% in *Entamoeba histolytica*; 96% in *Entamoeba invadens*) (Feschotte & Pritham, 2007). Replication of these TEs occurs during nuclear DNA replication, where the TE is duplicated into the new strand, is later transposed from the parent strand, and then inserted downstream of the replication fork. This results in two copies of the TE in the replicated strand.

C. Autonomy

An important feature of both TEs is autonomy. Some TEs lack the ability to transpose themselves (lack reverse transcriptase for retrotransposons, or transposase for DNA transposons), and must rely on the machinery from other transposons to transpose. This plays a significant role in proliferation and elimination of TEs.

III. TE ORIGINS

The evolutionary origin of TEs is difficult to study, due to their chaotic nature and prevalence in most organisms. We are confident of a close link to viruses, as well as potential methods of transfer between organisms.

A. Viral Similarity

There are several features in structure and function shared between viruses and TEs; This suggests they share some evolutionary history (Lerat & Capy, 1999). For instance, LTR retrotransposons and retroviruses both have Long Terminal Repeats (LTRs), *gag* (antigen) and *pol* (reverse transcriptase) genes. The *env* gene is sometimes shared, but will be incomplete or nonfunctional in the retrotransposon, preventing the encapsulation into a protein envelope. Further similarity occurs in reverse transcriptase-like sequences of retrotransposons, retroviruses and DNA viruses. Each show highly conserved regions, which suggests a common ancestor (Xiong & Eickbush, 2009).

Retroviruses show further similarity to TEs, as they integrate themselves into the host genome, and later excise themselves when conditions are favorable (Leaf-Loving, 2004). Typically, the retrovirus will only encode a single copy of itself into the genome, but some have been known to create several copies (Knouf et al., 2009). Some parvoviruses have gone even further and completely lost the ability of normal-viral propagation, instead adopting a transposon-like replication method and existing exclusively within the host genome (Fisher & Mayor, 1991).

The origin of TEs is rather unclear, but some theories have been put forward. The similarity to viruses strongly suggests there is some relationship, giving rise to 3 different hypotheses (Wessner, 2010):

i. Progressive Hypothesis

This hypothesis suggests that viruses evolved out of TEs in the genome. TEs are now known to be a significant evolutionary force (Galindo-González et al., 2017), and may have arisen through natural selection (Le Rouzic et al., 2007). The mobile nature of TEs may have captured a form of reverse transcriptase, and the resulting sequence may have allowed survival outside the cell. It could have acted as an endosymbiont, like mitochondria, providing some useful function to the cell to avoid selection during early evolution, and later became more parasitic.

ii. Regressive Hypothesis

Similar to mitochondria and other organelles, viruses may have derived from an ancient bacterium. The group of viruses known as nucleocytoplasmic large DNA viruses (NCLDVs) give significant evidence that this may have happened. There is an overlap in both physical size and genome length between the largest viruses and the smallest bacterium. The *Mimivirus* has a diameter of 750 nm and a genome length of 1.2 mbp, whereas a *Mycoplasma* bacteria has a size of around 250 nm, and *Candidatus Carsonella* has a genome length of 159 kbp (Wessner, 2010). Furthermore, the *Mimivirus* also encodes for a relatively large number of genes that are associated with translation, but are not used in its virus lifecycle. This could indicate that an ancestor could replicate itself, something that is not characteristic of a virus (Roult et al., 2004). TEs could be a further regression of the virus, losing the ability to move between hosts, and existing only within the host genome.

iii. Virus-first Hypothesis

Viruses are very simple compared to cells, so they may have existed first. There is strong evidence that RNA was the first type of replicating molecule and some RNA molecules can catalyze chemical reactions (Prangishvili et al., 2006). These two mechanisms could lead to a precursor of a virus, and possibly the first organism. Further evolution may have given some of these precursors an outer shell or membrane, while others gained the ability to infect. TEs could have arisen from one of these infective precursors that retained its replicative machinery, but lost its ability to move between organisms.

B. Horizontal Gene Transfer (HGT)

Genes are known to be commonly transferred between unrelated species. This includes TEs, and evidence has shown that they have been transferred between most types of organisms.

i. Bacteria-to-Eukaryote

This method is likely the most common method of HGT, and has shown that even essential genes can be transferred. Glycerol transporters in plants show strong evidence for a bacterial origin (Zardoya et al. 2002). The transfer of water and glycerol into the cell are often mediated by aquaporins (AQPs) and aquaglyceroporins (AQGPs), respectively. Plants lack the latter, but can still transport glycerol. Diversification of species would suggest that plant glycerol transporters should resemble proteins found in other eukaryotes, but a sister family of transporters has only been found as bacterial AQPs. It was estimated that ≈1.2 bya, very early in plant emergence, a single HGT event may have occurred to recruit bacterial AQPs to transport glycerol in plants. Furthermore, this method of HGT has been utilized as a vector for genetic modification. *Agrobacterium* is a pathogen known for inducing tumors in plants, via HGT to the host organism. It transfers plasmids into the host, which are integrated into the nuclear DNA, and the mechanism for transfer is well-understood (Gelvin, 2003). It has shown a board range of transformation not

only in plants, but also fungi and animals (Bundock et al. 1995; Kunik et al., 2001). TEs are commonly transferred in plants via this method of HGT (Diao et al., 2005).

ii. Organelle-to-Eukaryote

Intracellular gene transfer (IGT) is a special case of HGT where the transfer occurs from an organelle, often mitochondria, to the host eukaryote. This process occurs commonly in plants, but is virtually inactive in animals and fungi (Adams et al., 2002). Furthermore, HGT can occur between mitochondria of foreign species, creating a vector for eukaryote-to-eukaryote HGT (Won and Renner, 2003). There have been 3 types of transfer identified – recapture, duplicative and chimera. In recapture HGT, a gene that was previous lost to the nuclear genome is repopulated in the mitochondrial genome. This has been observed in *Actinidia*, *Lonicera* and *Betula* (Bergthorsson et al., 2003). Duplicative HGT, or possessing both foreign and native copies of a gene, has been shown in *Amborella* and *Gnetum* (Won and Renner, 2003). Chimeric HGT involves the partial transfer of a foreign gene, resulting in a gene that is half native/half foreign, and has only been observed in *Sanguinaria* (Bergthorsson, 2003). A repeatable model has been shown for IGT, but the mechanism for transfer is still unknown (Thorsness et al., 1990). Further evidence of transfer of TEs through this method have been documented (Diao et al., 2005).

iii. Eukaryote-to-Eukaryote

Some fungi have shown direct HGT between species, primarily through anastomosis. An analysis of genomic data from *Magnaporthiopsis incrustans* and many species in the genus of *Colletotrichum* have shown HGT-derived markers for 93 different genes over 89 distinct transfer events. (Qui, 2016). These genes are exclusive to these species, so HGT via a previously discussed method is unlikely. On a broader scale, there is strong evidence that TE transfer through HGT has occurred in recent evolutionary history. The Ty1-copia group and Ty3-gypsy group of retrotransposons are present only in plants, animals and fungus, and show significant sequence similarities (Ty1: Favell, 1992; Ty3: Doolittle et al., 1989). These sequences are similar enough to suggest a recent evolutionary divergence, rather than being inherited from an ancient common ancestor (Kidwell, 1992). Retrotransposons show a 10^4-10^6-fold higher mutation rate than normal DNA, so typical inheritance would suggest that the difference in sequence would be proportional to the distance of the ancestor. However, numerous species show a remarkably low divergence in sequences – *Drosophila melanogaster* and *Saccharomyces cerevisiae* show a 1% divergence for Ty1-copia group elements (Favell, 1992). Methods for transfer are still largely unknown.

IV. TE EVOLUTION

There is seemingly an on-going battle between TEs and the host genome. If TEs are left unchecked, they may overwhelm the genome, and make the cell less competitive. However, TEs can confer significant benefits to the cell, and increase the competitiveness of the cell. It is in the best interest of each party to limit the proliferation, and mechanisms exist for both to do this. It is best to consider TEs as individual organisms within the host, and will survive based on the principles of natural selection.

A. Structural Evolution

Much like pathogens, TEs need to modulate their rate of transposition to ensure survival and 'reproduction'. A virus that kills the host immediately will likely not reproduce, much like a TE that chokes the genome of the host. Some TEs contain motifs that will interact with itself or others to reduce transposition. These can include sequences that code for siRNA that will deactivate more active TEs or itself (Feschotte, 2008; Pray, 2008), act as 'sinks' for foreign transposase (Yang et al., 2009), and encourage methylation of the TE to reduce transposition rate (Lisch, 2009). Each mechanism is advantageous to the host, encouraging it to retain that TE.

Size may also play a significant role in survival. Non-autonomous TEs are much shorter than their autonomous counterparts, so they are less affected by random mutation, are less frequently targeted for silencing, and are less often transcribed (Hollister and Gaut, 2009). However, the lack of autonomy makes replication difficult. There is evidence that some TEs will co-evolve, and have a low-copy number autonomous element provide the replication machinery required for the high-copy number non-autonomous element. This can happen even within the same TE, with both autonomous and non-autonomous versions of the TE existing in the genome simultaneously (Le Rouzic et al., 2007).

Target selection for insertion sites is common for TEs (Craig, 1997). It is obvious that deleterious insertions into essential genic regions would be a poor choice for survival. Length is a large factor in target selection – the longer the sequence, the less likely it will occur in a coding region. LTR retrotransposons are typically quite large (up to 25 kbp), and a most often found in intergenic regions. DNA transposons (typically 1-5 kbp) are often found at the 5' end of genic regions (Dietrich et al., 2002). Miniature Inverted-repeat TEs (MITEs; a DNA transposon) and Short INterspersed Elements (SINEs; a retrotransposon), both with a sequence of about 400-500 bp, can be found in high copy numbers in both genic and intergenic sequences.

B. Host Interactions

Methylation is a common tactic used by the host to suppress TEs. However, TEs have evolved responses to this attack. Some families, like the MuDR and Spm families in maize, have transposases that can also demethylate (Lisch, 2009). This allows the family of TEs to reactivation other methylated (inactive) family members. Another method of methylation prevention is to capture gene fragments of essential genes. If the host targets the TE for methylation, it risks suppressing captured gene fragments. The pack-MULE family commonly contains multiple gene fragments, suggesting that this method can confer a survival advantage to the TE (Hanada et al., 2009).

Silencing pathways are often under control of epigenetic mechanisms, and can lead to sudden changes in transposition activity – Stresses can include infection, injury, hunger, or even temperature (Lisch, 2009). This could be particularly important for speciation events. Sudden changes in the environment will increase the rate of transposition, increasing potential mutations, and possibly allowing for adaptive evolution. This process could be a large

contributing factor to the C-value paradox; More stressful environments breed organisms with larger genomes (Craddock, 2016).

The host can also 'domesticate' a TE, inactivating it to become a functional gene. About 70 eukaryotic genes are known to have derived from TE sequences (Alzohairy et al., 2013; Zhao et al., 2016). Further evidence shows that the DNA binding domain of most plant-specific transcription factors may have arisen from this process (Yamasaki et al., 2013).

C. Removal of TEs by Recombination

TE proliferation is largely limited by deletion events. A mechanism for removal of LTR retrotransposons has been observed, and similar mechanisms may exist for other types of TEs (Tenaillon et al., 2010). Removal occurs via unequal recombination of the two LTRs in the TE. The LTRs exist at the boundaries of the TE, so everything in between them would be removed during recombination. The resulting sequence would be a solo LTR. In theory, it is possible for this process to occur between two TEs sharing an LTR, resulting in all DNA between the two TEs being removed. However, since the resulting sequence would be identical to that of a single TE recombination, it is likely impossible to identify (Le Rouzic et al., 2007).

D. Impact on Genome

The presence of TEs can have a drastic effect of genome function, and will influence the evolutionary trajectory of the species. Population processes ultimately determine the fate of the TE. Advantages can arise from altered gene function, chromosomal re-arrangements and emergence of novel sequences.

Insertion and excision of TEs can affect allelic diversity, which is a key metric for determining potential for adaptation to future environmental changes (Caballero, 2013). Location of insertion and removal can drastically change the phenotype of the individual, but is not limited to coding sequences. Promoter regions, introns and untranslated regions are all insertion targets that can have effects ranging from no change, to subtle regulatory changes, to the creation of a novel protein via alternative splicing, to the complete loss of function for the gene (Kidwell & Lisch, 2007).

The DNA cleavage and repair during transposition could cause massive changes to genome structure, and is known as alternate transposition. If two separate TE copies transpose and synapse during repair, alternate ends may be spliced together. This would cause a significant inversion of the genome. Further events, including duplication and deletions, have been observed and at lengths of over 100 kbp (Gray, 2000; Xuan et al., 2016).

Genes in the host genome can be captured and replicated during transposition. If the captured sequence contains a regulatory element, expression of the downstream gene can be greatly affected, and may have significant effect on evolution – Over 85% of binding site for E2F transcription factors found in Brassica are contained within TEs, so it must play a strong role in the evolution of this regulatory network (Zhao et al., 2016).

E. Rise of Novel Sequences

Gene capture with retrotransposons can lead to novel genes. The process occurs first by reverse transcription of the captured sequence, and then reinserting it into the genome. However, this new gene will not contain any introns. This could massively change gene function, as introns are essential for alternative splicing and commonly contain regulatory functions. Studies have shown that new gene functions can arise from these 'retrogenes', and may have contributed to hominid brain development (Burki & Kaessmann, 2004).

Exon shuffling is another form of gene capture that can create novel genes. It can occur when an exon is inserted into another gene, and either duplicated or removed from the host gene to affect alternate splicing (Kaessmann, 2009). Mechanisms have been proposed for Long Interspersed Elements (LINEs; Ejima & Yang, 2003), Helitron TEs (Zhao et al., 2016), and LTRs (Wang et al., 2006). Though combinations will be somewhat random, as they are still affected by targeting site specificity, complex genes involving multiple imported sequences have been found (4 foreign exons in human tissue plasminogen activator; Kolkman & Stemmer, 2004)

F. Modelling

Analyzing TE proliferation and elimination in the genome is difficult. Very few TEs are actively mobile in the host genome (<0.05% in humans; Mill et al., 2007), but still compose a significant portion of the genome. This creates difficulty in determining rates of transposition, and TE removal can remove large numbers of TE without any trace.

i. Mathematical Models/Simulation Modelling

These models make assumptions for several factors, including selection against TEs, self-regulation to TEs, degradation of TEs by mutation, the resulting fitness of the host after insertion, and the mating systems of hosts (Le Rouzic et al., 2007). These factors are modelled over many generations of finite population sizes via computer simulation. A common pattern in most simulations is a cyclical balance of proliferation and removal of autonomous and non-autonomous TEs (Tenaillon et al., 2010). This suggests that an equilibrium in selection and reproduction may not exist. However, these models have difficulty with competition between TEs, mainly due to lack of knowledge about these interactions.

ii. Comparative Genomics

Empirical data can show locations of TE insertions, but it is difficult to analyze selection. Exons are typically uncluttered by TEs, but could result from either TEs inserting into these regions and being subsequently removed, or the TEs evolving an insertion bias to avoid them (Tenaillon et al., 2010). The location of insertion is also important for recombination, but it is difficult to tell if low recombination regions could be a result of the presence of certain TEs, or the TEs could have a bias for low recombination regions (Gaut, 2007). Methylation of nearby genes has also been shown to affect gene expression. In A. thaliana, expression of a gene is negatively correlated to the density of TEs in close proximity (Hollister & Gaut, 2009). This suggests a trade-off between

silencing the TE and rate of gene expression. However, this may be a form of self-regulation by the TE, or simply a defensive mechanism from the host. These are complicated questions, and will require the cooperation of many fields to answer.

V. SUMMARY

Transposable elements remain a difficult field of research. They compose a significant portion of most genomes, and seemingly conform to the rules of natural selection. Evidence has shown they evolve over time and react to their environment, as well as work in unison with the host to optimize survival. Proliferation and mutation can cause rapid evolutionary changes in the host, but later deletion events may leave no trace of this. We know TEs are a significant evolutionary force, and able to create new novel genes, but there is still no consensus on where they originated from. Improvements in sequencing and bioinformatics techniques will hopefully shed further light on these enigmatic elements.

VI. REFERENCES

Adams, K., Qiu Y.L., Stoutemyer, M., Palmer,J.D. (2002) Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proceedings of the National Academy of Sciences, 99*, 9905-9912

Alzohairy, A. M., Gyulai, G., Jansen, R. K., & Bahieldin, A. (2013). Transposable elements domesticated and neofunctionalized by eukaryotic genomes. *Plasmid, 69(1),* 1-15.

Bergthorsson, U., Adams, K., Thomason, B., & Palmer, J. (2003). Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature*, *424(6945)*, 197-201. doi: 10.1038/nature01743

Bundock, P., den Dulk-Ras, A., Beijersbergen, A., & Hooykaas, P. J. (1995). Trans-kingdom T-DNA transfer from Agrobacterium tumefaciens to Saccharomyces cerevisiae. *The EMBO journal, 14(13)*, 3206.

Burki, F., & Kaessmann, H. (2004). Birth and adaptive evolution of a hominoid gene that supports high neurotransmitter flux. *Nature genetics, 36(10),* 1061-1063.

Caballero, A., & García-Dorado, A. (2013). Allelic diversity and its implications for the rate of adaptation. *Genetics, 195(4),* 1373-1384.

Charlesworth, B., & Charlesworth, D. (1983). The population dynamics of transposable elements. *Genetical Research, 42(01),* 1-27.

Craddock, E. M. (2016). Profuse evolutionary diversification and speciation on volcanic islands: transposon instability and amplification bursts explain the genetic paradox. *Biology Direct, 11(1)*, 44.

Craig, N. L. (1997). Target site selection in transposition. *Annual review of biochemistry, 66(1),* 437-474.

Diao, X., Freeling, M., & Lisch, D. (2005). Horizontal transfer of a plant transposon. *PLoS Biol, 4(1),* e5.

Dietrich, C. R., Cui, F., Packila, M. L., Li, J., Ashlock, D. A., Nikolau, B. J., & Schnable, P. S. (2002). Maize Mu transposons are targeted to the 5′ untranslated region of the gl8 gene and sequences flanking Mu target-site duplications exhibit nonrandom nucleotide composition throughout the genome. *Genetics, 160(2),* 697-716.

Doolittle, R. F., Johnson, M. S., & Mcclure, M. A. (1989). Origins and Evolutionary Relationships of Retroviruses. *The Quarterly Review of Biology*, *64(1),* 1-30. doi:10.1086/416128

Ejima, Y., & Yang, L. (2003). Trans mobilization of genomic DNA as a mechanism for retrotransposon-mediated exon shuffling. *Human molecular genetics, 12(11),* 1321-1328.

Feschotte, C., & Pritham, E. J. (2007). DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet., 41*, 331-368.

Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. *Nature Reviews Genetics*, *9*(5), 397-405.

Fisher, R. E., & Mayor, H. D. (1991). The evolution of defective and autonomous parvoviruses. *Journal of theoretical biology, 149(4),* 429-439.

Flavell, A. J. (1992). Ty1-copia group retrotransposons and the evolution of retroelements in the eukaryotes. *Transposable Elements and Evolution Contemporary Issues in Genetics and Evolution*, 258-274. doi:10.1007/978-94-011-2028-9_19

Forterre, P., Filée, J., & Myllykallio, H. (2004). Origin and Evolution of DNA and DNA Replication Machineries. *The Genetic Code and the Origin of Life,* 145-168. doi:10.1007/0-387-26887-1_10

Galindo-González, L., Mhiri, C., Deyholos, M. K., & Grandbastien, M. A. (2017). LTR-retrotransposons in plants: Engines of evolution. *Gene, 626.* 14-25.

Gaut, B. S., Wright, S. I., Rizzon, C., Dvorak, J., & Anderson, L. K. (2007). Recombination: an underappreciated factor in the evolution of plant genomes. *Nature Reviews Genetics, 8(1),* 77-84.

Ge, S. X. (2016). Exploratory bioinformatics analysis reveals importance of "junk" DNA in early embryo development. *Ge BMC Genomics, 18:200*. doi:10.1101/079921

Gelvin, S. B. (2003). Agrobacterium-mediated plant transformation: the biology behind the “Gene-Jockeying” tool. *Microbiology and molecular biology reviews, 67(1),* 16-37.

Gbadegesin, M. A. (2012). Transposable elements in the genomes: parasites, junks or drivers of evolution?. *African journal of medicine and medical sciences*, *41*, 13-25.

Gray, Y. H. (2000). It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. *Trends in Genetics, 16(10),* 461-468.

Hallet, B. (1997). Transposition and site-specific recombination: adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. *FEMS Microbiology Reviews, 21*(2), 157-178. doi:10.1016/s0168-6445(97)00055-7

Hanada, K., Vallejo, V., Nobuta, K., Slotkin, R. K., Lisch, D., Meyers, B. C., ... & Jiang, N. (2009). The functional role of pack-MULEs in rice inferred from purifying selection and expression profile. *The Plant Cell, 21(1),* 25-38.

Hollister, J. D., & Gaut, B. S. (2009). Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome research, 19(8),* 1419-1428.

Kaessmann, H. (2010). Origins, evolution, and phenotypic impact of new genes. *Genome research, 20(10),* 1313-1326.

Kidwell, M. G. (1992). Horizontal transfer of P elements and other short inverted repeat transposons. *Genetica, 86*(1-3), 275-286. doi:10.1007/bf00133726

Kidwell, M. G., & Lisch, D. (1997). Transposable elements as sources of variation in animals and plants. *Proceedings of the National Academy of Sciences, 94(15),* 7704-7711.

Knouf, E. C., Metzger, M. J., Mitchell, P. S., Arroyo, J. D., Chevillet, J. R., Tewari, M., & Miller, A. D. (2009). Multiple integrated copies and high-level production of the human retrovirus XMRV (xenotropic murine leukemia virus-related virus) from 22Rv1 prostate carcinoma cells. *Journal of virology, 83(14)*, 7353-7356.

Kolkman, J. A., & Stemmer, W. P. (2001). Directed evolution of proteins by exon shuffling. *Nature biotechnology, 19(5),* 423-428.

Kunik, T., Tzfira, T., Kapulnik, Y., Gafni, Y., Dingwall, C., & Citovsky, V. (2001). Genetic transformation of HeLa cells by Agrobacterium. *Proceedings of the National Academy of Sciences, 98(4),* 1871-1876.

Le Rouzic, A., Boutin, T. S., & Capy, P. (2007). Long-term evolution of transposable elements. *Proceedings of the National Academy of Sciences, 104(49),* 19375-19380.

Leaf-Loving, N.K., (2004) Retrovirus Integration into the Human Genome. *PLoS Biology, 2(8),* e281. http://doi.org/10.1371/journal.pbio.0020281

Leprince, A.S., Grandbastien, M.A., Meyer, C., (2001). Retrotransposons of the Tnt1B family are mobile in Nicotiana plumbaginifolia and can induce alternative splicing of

the host gene upon insertion. *Plant Molecular Biology, 47*, 533–541. doi:10.1023/A:1011846910918.

Lerat, E., & Capy, P. (1999). Retrotransposons and retroviruses: analysis of the envelope gene. *Molecular Biology and Evolution, 16(9),* 1198-1207. doi:10.1093/oxfordjournals.molbev.a026210

Lisch, D. (2009). Epigenetic regulation of transposable elements in plants. *Annual review of plant biology, 60*, 43-66.

Martienssen, R. A. (1998). Functional genomics: Probing plant gene function and expression with transposons. *Proceedings of the National Academy of Sciences, 95*(5), 2021-2026. doi:10.1073/pnas.95.5.2021

Mills, R. E., Bennett, E. A., Iskow, R. C., & Devine, S. E. (2007). Which transposable elements are active in the human genome? *Trends in Genetics, 23*(4), 183-191. doi:10.1016/j.tig.2007.02.006

Orgel, L. E., & Crick, F. H. (1980). Selfish DNA: the ultimate parasite. *Nature, 284*(5757), 604-607. doi:10.1038/284604a0

Pennisi, E. (2012). "ENCODE Project Writes Eulogy for Junk DNA". *Science, 337(6099),* 1159–1161. doi:10.1126/science.337.6099.1159.

Prangishvili, D., Forterre, P., & Garrett, R. A. (2006). Viruses of the Archaea: a unifying view. *Nature Reviews Microbiology, 4(11),* 837-848.

Pray, L. (2008). Discovery of DNA structure and function: Watson and Crick. *Nature Education, 1(1),* 100.

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., ... & Claverie, J. M. (2004). The 1.2-megabase genome sequence of Mimivirus. *Science, 306(5700),* 1344-1350.

Robertson, H. M. (2002). Evolution of DNA transposons in eukaryotes. *American Society of Microbiology, Mobile DNA II*, 1093-1110

Sanmiguel, P., Tikhonov, A., Jin, Y., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., . . . Bennetzen, J. L. (1996). Nested Retrotransposons in the Intergenic Regions of the Maize Genome. *Science, 274*(5288), 765-768. doi:10.1126/science.274.5288.765

Tenaillon, M. I., Hollister, J. D., & Gaut, B. S. (2010). A triptych of the evolution of plant transposable elements. *Trends in plant science, 15(8),* 471-478.

Thorsness, P. E., & Fox, T. D. (1990). Escape of DNA from mitochondria to the nucleus in Saccharomyces cerevisiae. *Nature, 346(6282),* 376.

Qiu, H., Cai, G., Luo, J., Bhattacharya, D., & Zhang, N. (2016). Extensive horizontal gene transfers between plant pathogenic fungi. *BMC Biology, 14(1).* doi:10.1186/s12915-016-0264-3

Wang, W., Zheng, H., Fan, C., Li, J., Shi, J., Cai, Z., ... & Lu, Z. (2006). High rate of chimeric gene origination by retroposition in plant genomes. *The Plant Cell, 18(8),* 1791-1802.

Wessner, D. R. (2010) The Origins of Viruses. *Nature Education, 3(9)*, 37

Won, H., & Renner, S. S. (2003). Horizontal gene transfer from flowering plants to Gnetum. *Proceedings of the National Academy of Sciences, 100(19),* 10824-10829.

Xiong, Y., & Eickbush, T. H. (1988). Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. *Molecular Biology and Evolution, 5(6),* 675-690.

Xuan, Y. H., Peterson, T., & Han, C. D. (2016). Generation and Analysis of Transposon Ac/Ds-Induced Chromosomal Rearrangements in Rice Plants. *Chromosome and Genomic Engineering in Plants: Methods and Protocols,* 49-61.

Yamasaki, K., Kigawa, T., Seki, M., Shinozaki, K., & Yokoyama, S. (2013). DNA-binding domains of plant-specific transcription factors: structure, function, and evolution. *Trends in plant science, 18(5),* 267-276.

Yang, G., Nagel, D. H., Feschotte, C., Hancock, C. N., & Wessler, S. R. (2009). Tuned for transposition: molecular determinants underlying the hyperactivity of a Stowaway MITE. *Science, 325(5946)*, 1391-1394.

Zardoya, R., Ding, X., Kitagawa, Y., & Chrispeels, M. J. (2002). Origin of plant glycerol transporters by horizontal gene transfer and functional recruitment. *Proceedings of the National Academy of Sciences, 99(23)*, 14893-14896. doi:10.1073/pnas.192573799

Zhao, D., Ferguson, A. A., & Jiang, N. (2016). What makes up plant genomes: The vanishing line between transposable elements and genes. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 1859(2),* 366-380.