

# Quadruplex-forming DNA sequences spread by retrotransposons may serve as genome regulators

Eduard Kejnovsky<sup>1,2,\*</sup> and Matej Lexa<sup>2,3</sup>

<sup>1</sup>Department of Plant Developmental Genetics; Institute of Biophysics ASCR; Brno, Czech Republic; <sup>2</sup>Laboratory of Genome Dynamics; CEITEC—Central European Institute of Technology; Masaryk University; Brno, Czech Republic; <sup>3</sup>Faculty of Informatics; Masaryk University; Brno, Czech Republic

**Keywords:** LTR retrotransposons, DNA quadruplexes, TE transcription

**Abbreviations:** TEs, transposable elements; LTR, long terminal repeat

Transposable elements (TEs) are ubiquitous genome inhabitants in eukaryotes. Increasing evidence shows that TEs are involved in regulatory networks of eukaryotic cells and contribute to genome evolution. Recently, we reported that many plant long-terminal repeat (LTR) retrotransposons contain DNA quadruplex-forming sequences at precise positions inside their LTRs and that quadruplexes are better preserved in evolutionary younger elements. As quadruplexes can modulate molecular processes, quadruplexes found at specific distances upstream and downstream from the endogenous TE promoter can affect transcription of the element. Moreover, quadruplexes found in solo LTRs, as well as in 3' ends of 5'-truncated copies of LINE-1 elements, can affect expression of neighboring genes. Here, we propose that this way retrotransposons can serve as vehicles for spread of DNA quadruplexes. Quadruplexes can thus fulfill a dual regulatory role—to influence both the retrotransposons carrying them and the neighboring host genes, e.g., by direct effect on transcription or by modifying the local chromatin state. Additionally, four-stranded DNA structures may serve as hotspots for recombination-based genome rearrangements.

Transposable elements are a significant part of plant and animal genomes and greatly contribute to genome dynamics. Recently, we analyzed 18,377 LTR retrotransposons in 21 plant species and found that they often contain sequences with the potential to adopt four-stranded DNA structures—DNA quadruplexes.<sup>1</sup> Using circular dichroism and gel electrophoresis, we confirmed the ability of such sequences to form DNA quadruplexes. We showed that some motifs formed parallel-stranded quadruplexes while others adopted anti-parallel-stranded structures. Quadruplex-forming sequences were mostly localized in long-terminal repeats of LTR retrotransposons and found at specific distances from the TE promoter—upstream of promoter in minus strand and downstream of promoter in plus strand. Quadruplex-forming sequences were better preserved in presumably active evolutionary younger retrotransposons,

suggesting a role of four-stranded structures in the life-cycle of retrotransposons.

DNA quadruplexes can have an effect on many molecular processes in the cell including replication, transcription, chromatin remodeling or recombination (for review see Bochman et al.<sup>2</sup>). In transposable elements, the formation of a quadruplex upstream of their endogenous promoter in minus strand can open the DNA double helix and help transcription to proceed, while formation of a quadruplex downstream of the promoter in plus strand can inhibit or stop elongation of growing RNA strands.<sup>3</sup>

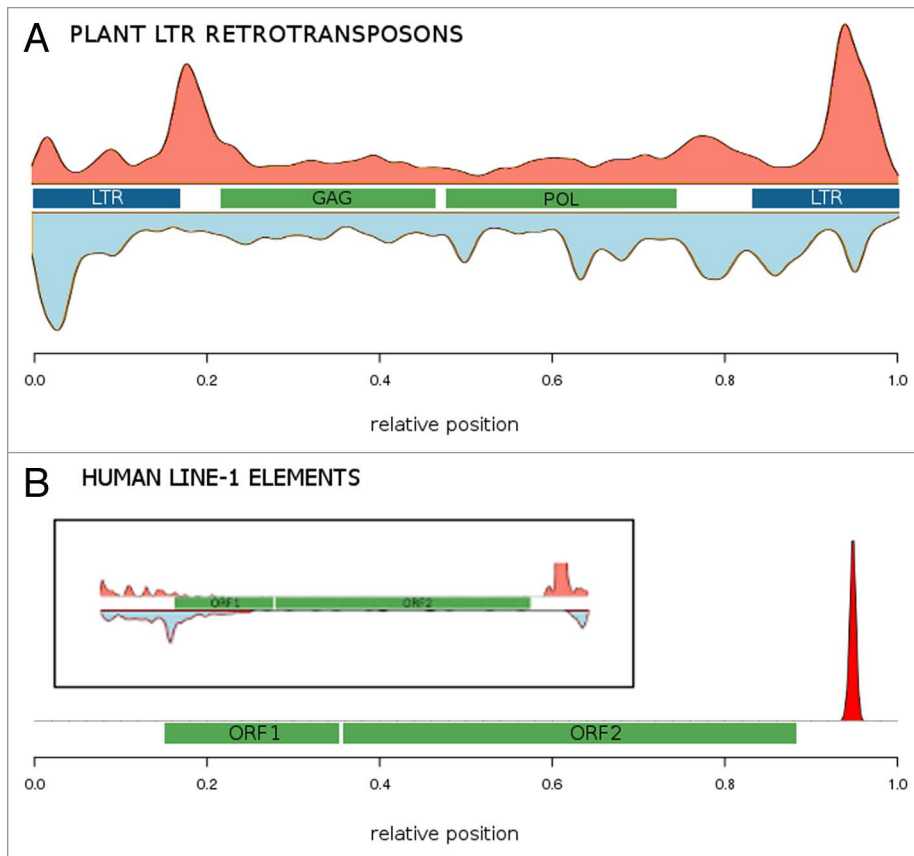
Quadruplex DNA can affect the whole genome as it represents a barrier for DNA replication. However, many cellular proteins like helicases can bind and unwind quadruplex DNA and allow replication to go through.<sup>2</sup> The function of one such helicase (FANCF) has clearly been demonstrated in mutants lacking the helicase protein. As a result, replication of quadruplex-containing regions was delayed.<sup>4–6</sup> Another protein, ATRX, belonging to a family of SWI/SNF chromatin modulators, binds quadruplexes and attracts H3.3 histone to nucleosomes, leading to changes of chromatin structure.<sup>7,8</sup> A silent state of the chromatin may then accumulate inside TEs or could spread from TEs to neighboring genomic regions. The quadruplexes flanking the TE promoter on both sides could serve as marks for changes in chromatin status, by interacting with ATRX and other chromatin/associated proteins. Interestingly, another protein, heterochromatin protein 1 (HP1) interacts with ATRX as well as with trimethylated Lys 4 and 9 in histone H3.<sup>9</sup> This has a signaling role in chromatin remodelling, possibly by marking boundaries between open and closed chromatin or regions that are transcriptionally silent or active.<sup>10</sup> Moreover, DNA elements with a less-stable duplex are present in the vicinity of the quadruplexes (promoters in LTRs and polyA sequences in LINE-1) which could help quadruplex formation.

LTR-borne quadruplexes therefore seem to be functioning as a link between the duplicative/dispersive activity of LTR retrotransposons and the epigenetic state of discrete genomic regions. Similar function has recently been proposed for LINE-1

\*Correspondence to: Eduard Kejnovsky; Email: kejnovsk@ibp.cz

Submitted: 01/17/2014; Accepted: 02/03/2014

Citation: Kejnovsky E, Lexa M. Quadruplex-forming DNA sequences spread by retrotransposons may serve as genome regulators. *Mobile Genetic Elements* 2014; 4:e28084; <http://dx.doi.org/10.4161/mge.28084>



**Figure 1.** Distribution of potential DNA quadruplex sequences (PQS) along (A) full-length plant LTR retrotransposons and (B) human LINE-1 elements. PQS in plus strand (PQS+) are shown in salmon, PQS in minus strand (PQS-) are shown in light blue. Main features of the transposable elements are shown in the middle: LTRs (blue), ORFs (green). The LTRs and GAG/POL ORFs are only shown symbolically as the data covered a large range of TE families with different architectures. The inset shows PQS density at 30× magnification of y-axis to better see minus strand PQS.

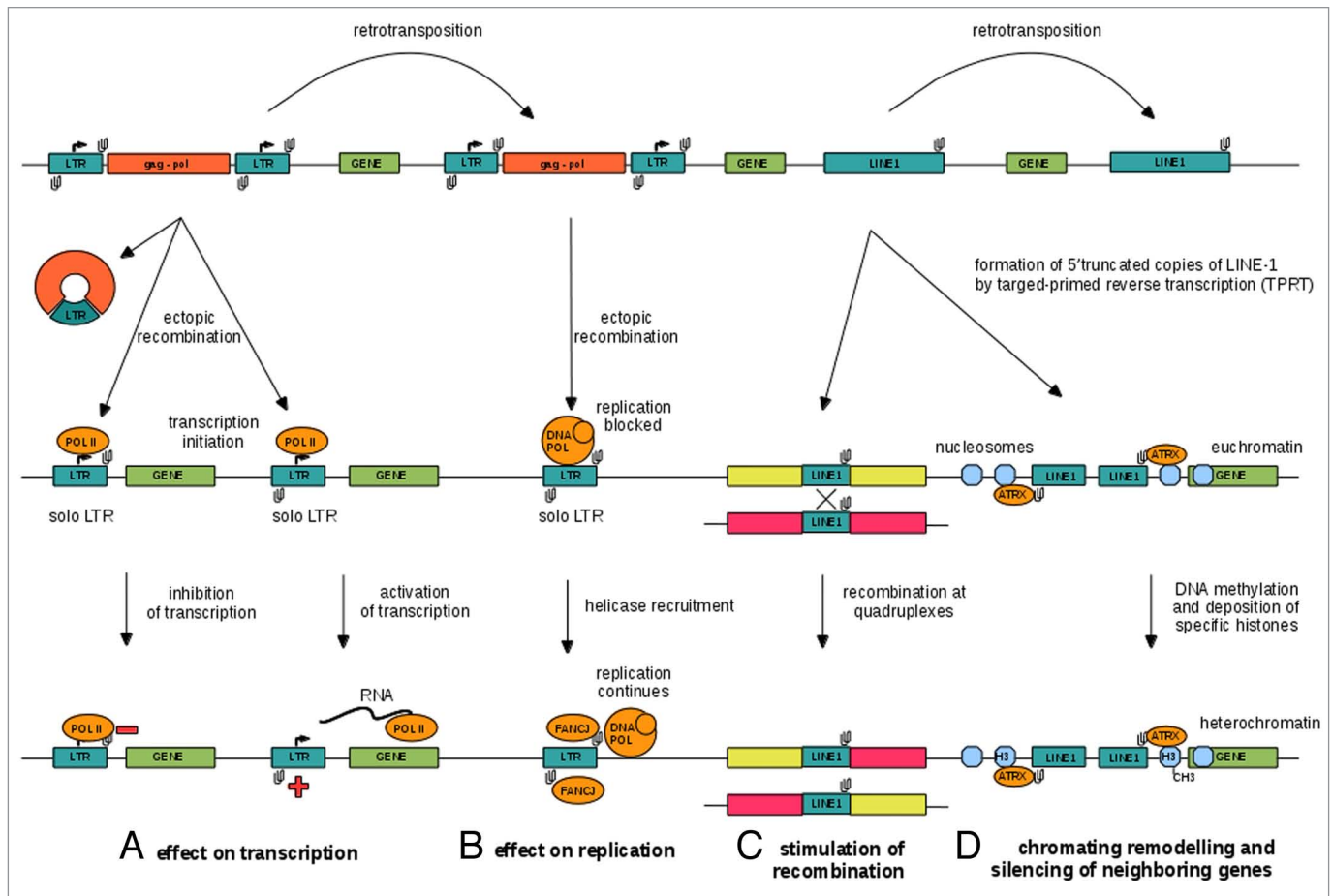
elements in human, where the 3'UTR quadruplexes may serve as markers for recruitment into perinuclear heterochromatin.<sup>11</sup> The association of LINE-1 elements with regional genome silencing was first proposed by Lyon<sup>12</sup>; LINE-1 elements act as “boosters” promoting expansion of heterochromatin during spread of the X chromosome-inactivation signal.<sup>13</sup> Quadruplexes are also sites of preferential recombination<sup>14</sup> which can lead to formation of chimeric TEs. They can also facilitate ectopic recombination of TE-related parts of the genome. Taken together, the maintenance of quadruplexes inside TEs and their spread in genome can influence the biology of TEs and contribute to whole-genome changes including silencing or reshuffling of genomic regions.

It is unclear whether DNA quadruplexes represent barriers to replication or transcription that are overcome by enzymes that have evolved to deal with these structures. On the contrary, these enzymes may exist to enable the formation and spread of quadruplexes. It is also becoming clear that quadruplexes are not only obstacles to DNA and/or RNA processing, they can serve as both negative and positive modulators of many cellular processes.

To investigate, how common quadruplexes are in various types of plant and animal TEs, we extended our analysis of LTR retrotransposons of 21 plant species<sup>1</sup> (Fig. 1A) to non-LTR retrotransposons in humans. While LTR retrotransposons are most common in plants, non-LTR retrotransposons are most abundant in humans. We found that potential quadruplex-forming sequences (PQS) are strongly accumulated at the 3' end of human long interspersed nuclear elements—LINE-1 elements (Fig. 1B). We also found that quadruplexes are preserved in solo LTRs—LTRs that remained in the genome after deletion of the rest of the retrotransposon by ectopic recombination. For example, in HERV-K family, where solo LTRs were identified,<sup>15</sup> we found quadruplexes in 13% of full-length HERV-K elements and in 10% of solo LTRs. In plants, we showed that LTR retrotransposon quadruplexes are present up to twice as often as predicted by a second-order Markov model.<sup>1</sup> In repeat masker data from the human UCSC Genome Browser,<sup>16</sup> we counted 12% of LINE-1 elements with at least one quadruplex sequence (86,477 PQS altogether) or about 23% of all 376,000 PQS that can be detected in the human genome.<sup>17</sup> Since LINE-1 elements form about 17% of the human genome, they carry more than the average share of the human PQS.

From our previous and current data, we cannot exclude that quadruplexes are gathered inside TEs because transposable elements are genomic regions where abundant secondary structures are less harmful for the genome and the host cell. However, the localization of quadruplexes in specific regions of TEs, as well as their better preservation inside younger elements more indicates the regulatory role of quadruplexes than their accumulation due to purifying selection.

Here, we propose a model of quadruplex spread by retrotransposons involving the possible role of quadruplexes in genome organization (Fig. 2). In our view, quadruplex-forming sequences appeared and were maintained inside long-terminal repeat of LTR retrotransposons and 3'UTR regions of LINE-1 non-LTR retrotransposons due to their contribution to the elements life cycle and their positive effect on the host. Solo LTRs, that are more abundant than full-length elements in many plant species like maize<sup>18</sup> and barley,<sup>19</sup> therefore represent an important store of quadruplexes with high regulatory potential. For example, the HERV-K family in humans contains one order of magnitude more solo LTRs than full-length elements.<sup>15,20</sup> Similarly, 3'UTRs regions of 5'end-truncated LINE-1 elements that make up the



**Figure 2.** A model of DNA quadruplex spreading by retrotransposition and the assumed biological functions of DNA quadruplexes in gene regulation and genome organization: (A) negative and positive effect on transcription via interfering with RNA extension or separating DNA strands, respectively, (B) inhibitory effect on DNA replication that can be overcome by binding of FANCD1-type helicases, (C) serving as hotspots for recombination and (D) the role in chromatin remodeling by attracting ATRX-like proteins that are associated with heterochromatin or heterochromatin boundaries.

majority of LINE-1 elements, are also a rich store of quadruplexes. Both solo LTRs and truncated LINE-1 elements are short and quadruplexes carried by these elements are delivered close to neighboring genes.

In conclusion, we propose that DNA quadruplexes are mainly spread in eukaryotic genomes by retrotransposons and play a dual regulatory role, affecting the retrotransposon life-cycle as well as the activity of neighboring genes, replication of DNA, chromatin status and can be involved in recombination-based genome rearrangements.

### References

- Lexa M, Kejnovsky E, Steflová P, Konvalinová H, Vorlíčková M, Vyskot B. Quadruplex-forming sequences occupy discrete regions inside plant LTR retrotransposons. *Nucleic Acids Res* 2014; 42:968-78; PMID:24106085; <http://dx.doi.org/10.1093/nar/gkt893>
- Bochman ML, Paeschke K, Zakian VA. DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet* 2012; 13:770-80; PMID:23032257; <http://dx.doi.org/10.1038/nrg3296>
- Tornaletti S, Park-Snyder S, Hanawalt PC. G4-forming sequences in the non-transcribed DNA strand pose blocks to T7 RNA polymerase and mammalian RNA polymerase II. *J Biol Chem* 2008; 283:12756-62; PMID:18292094; <http://dx.doi.org/10.1074/jbc.M705003200>
- Wu Y, Shin-ya K, Brosh RM Jr. FANCD1 helicase defective in Fanconi anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability. *Mol Cell Biol* 2008; 28:4116-28; PMID:18426915; <http://dx.doi.org/10.1128/MCB.02210-07>
- Schwab RA, Nieminuszczy J, Shin-ya K, Niedzwiedz W. FANCD1 couples replication past natural fork barriers with maintenance of chromatin structure. *J Cell Biol* 2013; 201:33-48; PMID:23530069; <http://dx.doi.org/10.1083/jcb.201208009>
- Tarsounas M, Tijsterman M. Genomes and G-quadruplexes: for better or for worse. *J Mol Biol* 2013; 425:4782-9; PMID:24076189; <http://dx.doi.org/10.1016/j.jmb.2013.09.026>
- Whitehouse I, Owen-Hughes T. ATRX: Put me on repeat. *Cell* 2010; 143:335-6; PMID:21029854; <http://dx.doi.org/10.1016/j.cell.2010.10.021>

### Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

### Acknowledgments

This research was supported by the Czech Science Foundation (grant P305/10/0930 to EK), grant AV0Z50040702 from the Academy of Sciences of the Czech Republic, by the project “CEITEC—Central European Institute of Technology” (CZ.1.05/1.1.00/02.0068) from European Regional Development Fund and by the project OPVK (CZ.1.07/2.3.00/20.0045).

8. Boyarchuk E, Montes de Oca R, Almouzni G. Cell cycle dynamics of histone variants at the centromere, a model for chromosomal landmarks. *Curr Opin Cell Biol* 2011; 23:266-76; PMID:21470840; <http://dx.doi.org/10.1016/j.ceb.2011.03.006>
9. Eustermann S, Yang J-Ch, Law MJ, Amos R, Chapman LM, Jelinska C, Garrick D, Clynes D, Gibbons RJ, Rhodes D, et al. Combinatorial readout of histone H3 modifications specifies localization of ATRX to heterochromatin. *Nat Struct Mol Biol* 2011; 18:777-82; PMID:21666677; <http://dx.doi.org/10.1038/nsmb.2070>
10. Goldberg AD, Banaszynski LA, Noh K-M, Lewis PW, Elsaesser SJ, Stadler S, Dewell S, Law M, Guo X, Li X, et al. Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* 2010; 140:678-91; PMID:20211137; <http://dx.doi.org/10.1016/j.cell.2010.01.003>
11. Belan E. LINEs of evidence: noncanonical DNA replication as an epigenetic determinant. *Biol Direct* 2013; 8:22; PMID:24034780; <http://dx.doi.org/10.1186/1745-6150-8-22>
12. Lyon MF. LINE-1 elements and X chromosome inactivation: a function for "junk" DNA? *Proc Natl Acad Sci U S A* 2000; 97:6248-9; PMID:10841528; <http://dx.doi.org/10.1073/pnas.97.12.6248>
13. Graham T, Boissinot S. The genomic distribution of L1 elements: the role of insertion bias and natural selection. *J Biomed Biotechnol* 2006; 2006:75327; PMID:16877820; <http://dx.doi.org/10.1155/JBB/2006/75327>
14. Boán F, Gómez-Márquez J. In vitro recombination mediated by G-quadruplexes. *ChemBiochem* 2010; 11:331-4; PMID:20014270; <http://dx.doi.org/10.1002/cbic.200900612>
15. Subramanian RP, Wildschutte JH, Russo C, Coffin JM. Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. *Retrovirology* 2011; 8:90; PMID:22067224; <http://dx.doi.org/10.1186/1742-4690-8-90>
16. Karolchik D, Barber GP, Casper J, Clawson H, Cline MS, Diekhans M, Dreszer TR, Fujita PA, Guruvadoo L, Haussler M, et al. The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res* 2014; 42:D764-70; PMID:24270787; <http://dx.doi.org/10.1093/nar/gkt1168>
17. Huppert JL, Balasubramanian S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res* 2005; 33:2908-16; PMID:15914667; <http://dx.doi.org/10.1093/nar/gki609>
18. SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, et al. Nested retrotransposons in the intergenic regions of the maize genome. *Science* 1996; 274:765-8; PMID:8864112; <http://dx.doi.org/10.1126/science.274.5288.765>
19. Vicient CM, Suoniemi A, Anamthawat-Jónsson K, Tanskanen J, Beharav A, Nevo E, Schulman AH. Retrotransposon BARE-1 and Its Role in Genome Evolution in the Genus *Hordeum*. *Plant Cell* 1999; 11:1769-84; PMID:10488242
20. Belshaw R, Dawson ALA, Woolven-Allen J, Redding J, Burt A, Tristem M. Genomewide screening reveals high levels of insertional polymorphism in the human endogenous retrovirus family HERV-K(HML2): implications for present-day activity. *J Virol* 2005; 79:12507-14; PMID:16160178; <http://dx.doi.org/10.1128/JVI.79.19.12507-12514.2005>