Epigenetika i pokretni genetički elementi

Despot Slade, Evelin

UNIVERSITY OF ZAGREB
FACULTY OF SCIENCE
BIOLOGY DIVISION

SVEUČILIŠTE U ZAGREBU
PRIRODOSLOVNO – MATEMATIČKI FAKULTET
BIOLOŠKI ODSJEK

Epigenetics and Transposable Elements

Epigenetika i pokretni genetički elementi

SEMINAR

SEMINARSKI RAD

Evelin Despot Slade
Undergraduate Study of Molecular Biology
(Preddiplomski studij molekularne biologije)
Mentor: Prof. dr. sc. Višnja Besendorfer

Zagreb, 2013.

TABLE OF CONTENTS

1. INTRODUCTION

The word ‘epigenetics’ was historically used to describe events that could not be explained by genetic principles [1]. Conrad Waddington in the late 1930s remarked that ‘One might say that the set of organizers and organizing relations to which a certain piece of tissue will be subject during development make up its “epigenetic constitution” or “epigenotype”...’ [2]. Today we refer to epigenetics as a study of stable, often heritable, changes that influence gene expression that are not mediated by DNA sequence and its mechanisms play crucial role in chromatin state regulation, thereby influencing processes such as gene expression, DNA repair and recombination [3]. Looking at eukaryotic organisms we can see that most epigenetic mechanisms are evolutionary conserved and several homologs of different epigenetic factors are present in plants and animals [4]. All somatic cells descended from a single progenitor contain near-identical genotype and during normal development they differentiate to acquire diverse biological function by expressing and repressing different set of genes; later this epigenetic marks are maintained through cell division to preserve cell identity [5]. One additional role of epigenetics is also to examine the influence of the environment in gene expression to determine how environment beside intracellular signals can influence the expression of genes [4].

Transposable elements (TEs) are mobile fragments of DNA that are repressed in both plant and animal genomes through epigenetic inheritance of silenced chromatin and expression states [6]. Most interesting characteristic is their ability to replicate themselves to extremely high genomic copy numbers. The 1983 Nobel Prize for Physiology and Medicine was awarded to Barbara McClintock for her discovery of genetic transposition [7, 8] and it all started with study of color patterns in maize kernels that led to first conclusion about ‘controlling elements’ that jump around the genome regulating gene expression [9]. Today, TEs are studied, examined and recognized as important genome components that help to shape their evolution with a profound impact on structure and function. Genome sequencing has revealed that transposable elements constitute a large fraction of most eukaryotic genomes where about half (Fig. 1.) of human repeat sequences is derived from TEs [10]. Recently developed highly sensitive alternative de novo strategy, P-clouds, which searches for clusters of high-abundance oligonucleotides that are related in sequence space, suggest that actually 66%-69% of the human genome is repetitive or repeat-derived

[11]. These date show the importance of understanding mechanisms of transposable elements because of their diverse impact on host transciptome.

Purpose of this summary is to bring transposable elements and epigenetic regulatory mechanisms in relationship and to give current opinions on the topic of their evolution. TEs have been underlying epigenetic phenomena for many years and there is important progress recently made in understanding these silencing mechanisms.

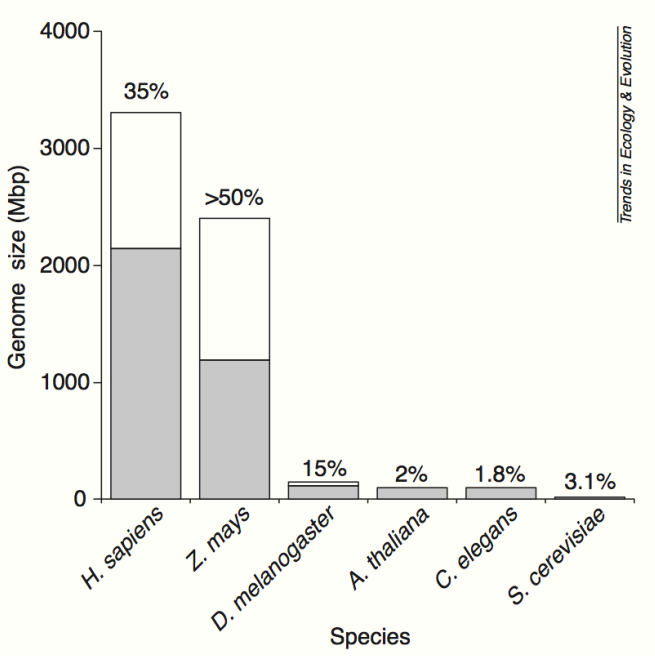

Figure 1. A histogram showing total genome sizes and the percentage occupied by transposable elements (open bars) for humans (data from 2000. - today is thought TEs make more than half of genome) and five model organisms (from [12])

2. TRANSPOSABLE ELEMENTS

2.1. From ‘junk’ and ‘selfish’ DNA to TEs

Since the radical suggestion that some genes might move along chromosomes in the 1950s by Barbara McClintock, our knowledge of transposable elements has vastly increased [13]. Susumu Ohno in 1972 used the term ‘junk’ DNA to refer to pseudogenes but with time meaning expanded to include all non-coding DNA. First idea about selfish DNA has sketched briefly but clearly by Dawkins in his book The Selfish Gene (1976). This term was widely used to describe pieces of genome with two distinct properties: (i) it arises when a DNA sequence forms additional copies of itself; (ii) it makes no contribution to the phenotype [14]. The history of these genomic elements shows one of the best examples of how scientific idea in biology emerge from first definitions and than evolve into new concept – TEs are no longer seen as ‘junk’ and ‘selfish’ pieces of DNA but rather as major components of genomes that have played a significant role in evolution [13].

2.2. Types and structure of TEs

Several criteria are used for classifying transposable elements and one is the requirement for an enzyme called reverse transcriptase that allows transposition of TEs. They are divided to one of two classes (Fig. 2.) according to their mechanism of transposition.

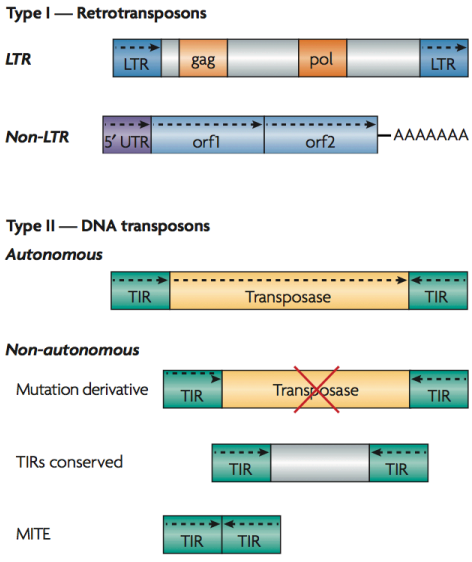

Figure 2. Types and structures of transposable elements (from [12])

Type I elements are called retrotransposons and they need a reverse transcription step. They are first transcribed from DNA to RNA and than the produced RNA is reverse transcribed into DNA. This step is usually catalyzed by a reverse transcriptase, which is often encoded by the TE itself. This copied DNA is than inserted at a new position in the genome. Retrotransposons can be divided into two types, on the basis of the presence or absence of direct repeats called long terminal repeats (LTRs). Ones with LTRs on their ends are closely related to retroviral proteins but they lack the envelope protein that is required to exit the cell. Among non-LTRs there are two families of elements; LINEs that encode for reverse transcriptase and SINEs that lack in same enzyme. Retrotransposons undergo duplicative transposition because their total number increases after each transposition with the potential to expand genomes [15].

Type II TEs, known as DNA transposons, don't need a reverse-transcription step for integration into a genome and instead they encode for transposase that recognizes the terminal inverted repeats (TIRs) that flank the TE, excises the TE out of the donor position and than integrates the transposon in the new acceptor site [15]. A main characteristic is that this mechanism does not involve an RNA intermediate. They use cut-and-paste transposition and the original gap is repaired without element replacement or they can use gap repair to fill with a copy of the transposon.

3. GENE EXPRESSION AND SILENCING

3.1. TEs influence on gene expression

The most prominent features of TEs are their invasiveness, the structural and functional consequences caused by their genomic insertions, and their potential ability to cross species boundaries [16]; in their active state TEs are seen as highly mutagenic, often targeting protein coding genes for insertion, causing chromosome breakage, illegitimate recombination and genome rearrangement [15]. It has been observed that TEs can also alter the regulation and expression of flanking genes in a variety of ways, either by altering polyadenylation and splicing patterns, or by acting as enhancer or promoter [17]. In *Drosophila* if inserted into the myosin heavy chain, every transposable element introduces a strong polyadenylation signal that defines novel terminal exons, which are then differentially recognized by alternative splicing apparatus [18]. TEs can exist in the genome in the form of cryptic elements, which remain intact and silent. On the other hand they can also affect the activity of genes as in *E. coli* where they can activate cryptic catabolic operons by small TEs called insertion sequences (IS elements) [19].

Chromatin is the combination of DNA and proteins that make up the contents of the cell nucleus. In dependence of its structure that is closely associated with the function we can distinguish two different levels of packaging. Euchromatin undergoes de-condensation in interphase, whereas heterochromatin has been defined as deeply staining chromosomal material that remains condensed [20]. Latter one can be found near centromeres and telomeres with attribute of repetitive and last-replicating regions. Finding that heterochromatin is composed of transposable elements that could regulate development [21] lies in two discoveries. Much older and already mentioned gene silencing in maize mediated by ‘controlling elements’ (TEs) and a phenomenon found in *Drosophila* called position-effect variegation where genes become silenced be heterochromatisation [22]. Looking at transposable elements more closely with understanding of their contribution to the function of heterochromatin it became possible to untangle their role in directing gen expression.

3.2. Epigenetic silencing of TEs – two different views

Maintaining transposable elements in their inactive state and avoiding their potentially harmful effect is one of cell's main goal. The genome has parallel evolved epigenetic 'defense' mechanisms to suppress their activity, and in such epigenetically inactive state TE retains the coding potential to mobilize itself but does not produce the necessary proteins because of a repressive chromatin environment [15]. In order to present epigenetic mechanisms of TE silencing it is necessary to look at transcriptional and post-transcriptional levels with a participation of host factors that regulate TE mobility. In hindsight, there was a prevailing view that epigenetic mechanisms evolved to control the disruptive potential of TEs but there is also one completely inverse suggestion that TEs actually accumulate in eukaryotic genomes because of epigenetic silencing mechanisms [23, 24].

Recently has been suggested a model based on *Arabidopsis thaliana* in which host silencing of TEs near genes has deleterious effect on neighboring gene expression [25, 26]. This is very interesting because in gene-rich regions of chromosome it consequently leads to preferential loss of methylated TEs. This problem is later described in the context of evolution.

3.2.1. Post-transcriptional silencing of TEs by RNAi

This type of silencing is widely found in *Caenorhabditis elegans*. dsRNA is post-transcriptional RNAi pathway first cleaved into siRNAs (small interfering RNAs) by a dicer-family protein. siRNA-guided transcript-cleavage complex (RISC) complex than loads siRNA and helps in finding complementary transcripts to siRNA that are consequentially cleaved (Fig. 3a.).

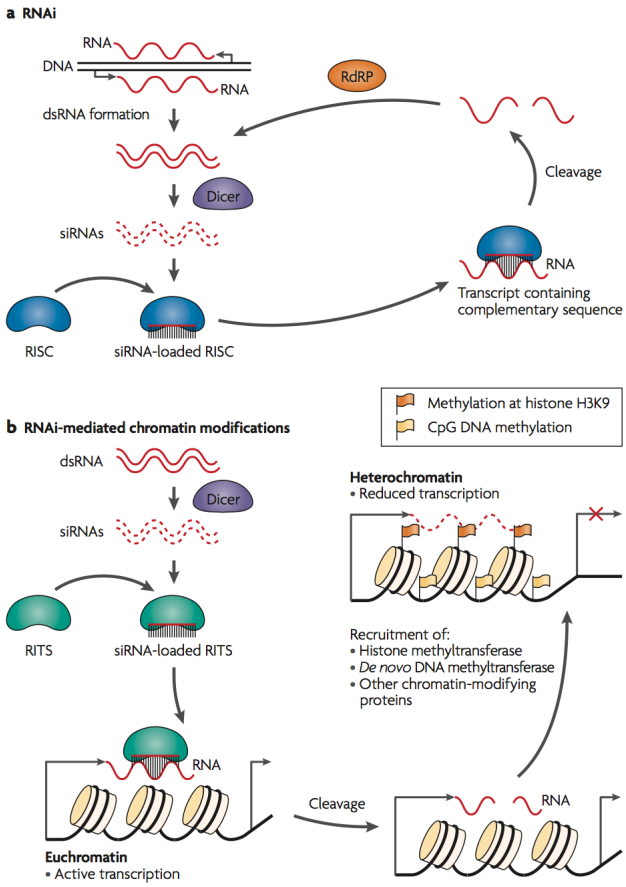

Figure 3. Mechanisms of transposable element silencing (from [12])

3.2.2. Chromatin modifications

This type of modification suppresses TE transcription. If there is a modification of histone amino-terminal tails than the binding of protein factors is altered. The *Arabidopsis* gene DDM1 is required to maintain DNA methylation levels and so in wild-type heterochromatin, transposons and silent genes are associated with histone H3 methylated at lysine 9 (H3K9), which is a signal for transcriptionally repressive and inactive chromatin [27]. DNA methylation on cytosine residues has the specific effect of reducing gene expression and has been found in every vertebrate examined. Methylation in a symmetrical context (CpG) can be retained upon DNA replication, providing a mechanism for inheritance of TE silencing. There is a subset of human LINE-1 (L1) retrotransposon elements that are targeted *de novo* for silencing trough methylation, which is efficiently maintained in asymmetric non-CG sites [28]. SWI/SNF is a nucleosome remodeling complex found in both eukaryotes and prokaryotes, and as a group of proteins they associate to remodel the way DNA is packaged. Ability to modify chromatin structure implies involvement in TE silencing.

3.2.3. RNAi-mediated chromatin modifications

In the fission yeast *Schizosaccharomyces pombe* pericentromeric heterochromatin has provided a model for this type of gene silencing. siRNA guide the cleavage of nascent transcripts, which are still attached to RNA polymerase II and the DNA strand (Fig. 3b.); this cleavage targets the region of chromatin for modification by the recruitment of H3K9 methyltransferase or other proteins that methylate the cytosine bases in dependence of organism species [15]. Described modifications of heterochromatin transposable elements make him condensed and inaccessible to transcription. Same principle of siRNA modification is also observed in *A. thaliana* but with numerous proteins and various type of RNAi.

3.2.4. Germline silencing

Under the term germline we think of genetic material that may be passed to a child. Best examples are the P elements that are found in *Drosophila melanogaster* [29]. They are actually DNA transposons that become active as a result of crossing between female without and male with P elements. Obtained progeny has a phenotype referred as hybrid dysgenesis with a temperature-dependent sterility, elevated mutation rates, and increased chromosome rearrangement and recombination. P elements in the maternal parent suppress this phenotype, suggesting the involvement of cytoplasmic factor, repressor that suppresses the activity of these elements.

4. GENOME EVOLUTION

4.1. TEs impact on evolution

Most TE-insertion mutations seem to exert a negative effect on host fitness, a growing list of evidences indicate that some TE-mediated genetic changes have become established features of host species genomes and having that in mind TEs could contribute significantly to organismic evolution [30]. One nice example is previously mentioned loss of silenced DNA in gen-rich regions. Here we have an evolutionary tradeoff in which the benefit of TE silencing means fitness cost via deleterious effect on the expression of nearby genes [25].

'C-value paradox' is the fact that organisms at the same general level of morphological complexity, with presumably same genetic requirements, often have genomes whose DNA content differ by orders of magnitude (Fig. 4.) [31]. The explanation largely reside in the profound differences among genomes in the abundance of TEs [23].

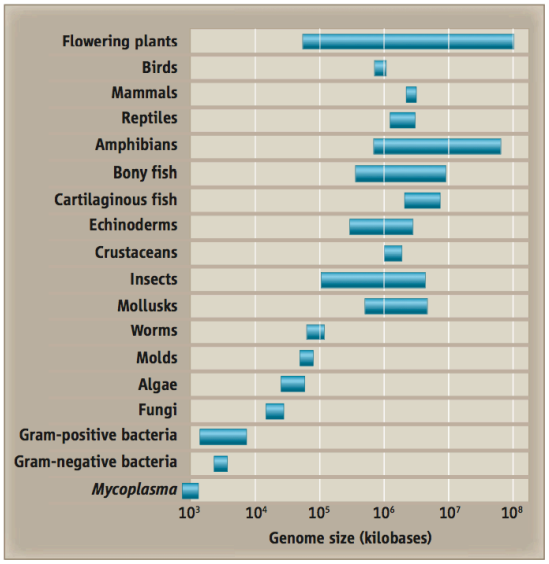

Figure 4. The C-value paradox. The range of haploid genome sizes in dependence of groups of organisms on the left (from [23])

The question that still remains is how could genomes accumulate such vast amounts of repetitive sequences? Transposons were around long before the eukaryotic lifestyle and they coevolved with all the rest of the eukaryotic genome's inhabitants. Also DNA that has little or no phenotypic effect is not under the selective pressure but still can multiply within the genome. Transposons are subject to horizontal transfer in

eukaryotes and lately has been provided evidence that a DNA transposon called SPIN has colonized the genome of nearly every major linage of reptiles [32]. But still the frequency and contribution of this phenomenon to genome evolution in eukaryotes remain poorly understood.

4.2. Driving evolution

Over years and years of evolution, mobile elements have achieved a balance between harmful effect on individual and long-term beneficial effect on a species [33] with constant push to complexity [34]. Eukaryotic genome accumulated transposable elements and grew through time with parallel evolution of prokaryotic epigenetic mechanisms, which firstly limiting recombination among horizontally exchanged sequences and afterwards regulate homologous recombination. The ability to suppress homologous recombination might be what pushed the balance between duplication and deletion in favor of sequence endo-reduplication and transposon proliferation [23]. Differential gene expression headed by DNA and histone modification, small RNA-mediated and transcriptional mechanisms enabled organisms through evolution to retain duplicated sequences.

Transposases are transposon-encoded enzymes that cleave transposon ends and attach them to new sequences. Excision usually leaves behind target site duplication, which generates sequence diversity. TEs mechanism guarantees genome variations, rearrangements and ensures that evolution is pushed forward.

5. CONCLUSION

Phenotype-altering mutations caused by transposon insertion are not so frequent, much less than are point mutation in most organisms. But nevertheless, transposition is disruptive in nature and it is imperative for host genomes to evolve mechanisms that suppress the activity of transposable elements. Two main mechanisms are used: (i) methylation and (ii) cosupression usually mediated by small interfering RNA (siRNA). There are two opinions on how epigenetic silencing evolved. One older and till recently prevailing, considered that these mechanism arose to control invading, parasitic transposons. On the other hand, there is more and more evidence that epigenetic silencing underlies both the genome expansion and proliferation of TEs.

All properties that lead TEs to be labeled as 'junk DNA' for many years might have enabled TEs to provide genomes with evolutionary potential to evolve new tools for generating diversity. Indeed, their ability to move has structured and restructured genes and their regulatory sequences influenced eukaryotic evolvability. Many questions still remain but one thing is certain; the more we understand epigenetic regulations of TEs, many other linked epigenetic phenomena are likely to be elucidated.

6. LITERATURE

1. Goldberg, A.D., C.D. Allis, and E. Bernstein, *Epigenetics: a landscape takes shape.* Cell, 2007. **128**(4): p. 635-8.
2. Hurd, P.J., *The era of epigenetics.* Briefings in Functional Genomics, 2010. **9**(5-6): p. 425-428.
3. Fingerman, I.M., et al., *NCBI Epigenomics: what's new for 2013.* Nucleic Acids Res, 2013. **41**(Database issue): p. D221-5.
4. Seffer, I., et al., *Unexplored Potentials of Epigenetic Mechanisms of Plants and Animals-Theoretical Considerations.* Genetics & Epigenetics, 2013. **5**(3751): p. 23-41.
5. Mohtat, D. and K. Susztak, *Fine tuning gene expression: the epigenome.* Semin Nephrol, 2010. **30**(5): p. 468-76.
6. Nuthikattu, S., et al., *The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs.* Plant Physiol, 2013. **162**(1): p. 116-31.
7. Mc, C.B., *The origin and behavior of mutable loci in maize.* Proc Natl Acad Sci U S A, 1950. **36**(6): p. 344-55.
8. McClintock, B., *The significance of responses of the genome to challenge.* Science, 1984. **226**(4676): p. 792-801.
9. Cowley, M. and R.J. Oakey, *Transposable elements re-wire and fine-tune the transcriptome.* PLoS Genet, 2013. **9**(1): p. e1003234.
10. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.
11. de Koning, A.P., et al., *Repetitive elements may comprise over two-thirds of the human genome.* PLoS Genet, 2011. **7**(12): p. e1002384.
12. Kidwell, M.G. and D.R. Lisch, *Transposable elements and host genome evolution.* Trends Ecol Evol, 2000. **15**(3): p. 95-99.
13. Biemont, C., *A brief history of the status of transposable elements: from junk DNA to major players in evolution.* Genetics, 2010. **186**(4): p. 1085-93.
14. Orgel, L.E. and F.H. Crick, *Selfish DNA: the ultimate parasite.* Nature, 1980. **284**(5757): p. 604-7.
15. Slotkin, R.K. and R. Martienssen, *Transposable elements and the epigenetic regulation of the genome.* Nat Rev Genet, 2007. **8**(4): p. 272-85.
16. Miller, W.J. and P. Capy, *Applying mobile genetic elements for genome analysis and evolution.* Mol Biotechnol, 2006. **33**(2): p. 161-74.
17. Girard, L. and M. Freeling, *Regulatory changes as a consequence of transposon insertion.* Dev Genet, 1999. **25**(4): p. 291-6.
18. Davis, M.B., et al., *Transposable element insertions respecify alternative exon splicing in three Drosophila myosin heavy chain mutants.* Genetics, 1998. **150**(3): p. 1105-14.
19. Hall, B.G., *Transposable elements as activators of cryptic genes in E. coli.* Genetica, 1999. **107**(1-3): p. 181-7.
20. Lippman, Z., et al., *Role of transposable elements in heterochromatin and epigenetic control.* Nature, 2004. **430**(6998): p. 471-6.
21. Holliday, R. and J.E. Pugh, *DNA modification mechanisms and gene activity during development.* Science, 1975. **187**(4173): p. 226-32.

22. Schotta, G., et al., *Position-effect variegation and the genetic dissection of chromatin regulation in Drosophila.* Semin Cell Dev Biol, 2003. **14**(1): p. 67-75.
23. Fedoroff, N.V., *Presidential address. Transposable elements, epigenetics, and genome evolution.* Science, 2012. **338**(6108): p. 758-67.
24. Huda, A., L. Marino-Ramirez, and I.K. Jordan, *Epigenetic histone modifications of human transposable elements: genome defense versus exaptation.* Mob DNA, 2010. **1**(1): p. 2.
25. Hollister, J.D. and B.S. Gaut, *Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression.* Genome Res, 2009. **19**(8): p. 1419-28.
26. Warnefors, M., V. Pereira, and A. Eyre-Walker, *Transposable elements: insertion pattern and impact on gene expression evolution in hominids.* Mol Biol Evol, 2010. **27**(8): p. 1955-62.
27. Gendrel, A.V., et al., *Dependence of heterochromatic histone H3 methylation patterns on the Arabidopsis gene DDM1.* Science, 2002. **297**(5588): p. 1871-3.
28. Woodcock, D.M., et al., *Asymmetric methylation in the hypermethylated CpG promoter region of the human L1 retrotransposon.* J Biol Chem, 1997. **272**(12): p. 7810-6.
29. Robertson, H.M. and W.R. Engels, *Modified P elements that mimic the P cytotype in Drosophila melanogaster.* Genetics, 1989. **123**(4): p. 815-24.
30. McDonald, J.F., *Evolution and consequences of transposable elements.* Curr Opin Genet Dev, 1993. **3**(6): p. 855-64.
31. Gall, J.G., *Chromosome structure and the C-value paradox.* J Cell Biol, 1981. **91**(3 Pt 2): p. 3s-14s.
32. Gilbert, C., et al., *Rampant horizontal transfer of SPIN transposons in squamate reptiles.* Mol Biol Evol, 2012. **29**(2): p. 503-15.
33. Kazazian, H.H., Jr., *Mobile elements: drivers of genome evolution.* Science, 2004. **303**(5664): p. 1626-32.
34. Bowen, N.J. and I.K. Jordan, *Transposable elements and the evolution of eukaryotic complexity.* Curr Issues Mol Biol, 2002. **4**(3): p. 65-76.

7. SUMMARY

Transposable elements are mobile DNA sequences that are widely distributed in bacteria, plants and animals. They are a common component in many epigenetic mechanisms, as they generally exist in inactive ‘epigenetically silenced’ form.

New insight challenges the view on TEs as genomic parasites and gives them profoundly generative role in genome evolution. TEs accumulate because of, not despite, epigenetic mechanisms. Although silencing slows the pace of genome restructuring to an evolutionary time scale, mobile elements within genome have indubitably driven genome evolution in diverse ways.

8. SAŽETAK

Pokretni genetički elementi su mobilne DNA sekvence koje su široko rasprostranjene među bakterijama, biljkama i životinjama. Oni su česta komponenta u mnogim epigenetičkim mehanizmima i uglavnom postoje u inaktivnom ‘epigenetički utišanom’ obliku.

Nova saznanja dovode u pitanje pogled na TE kao genomske parazite i daje im temeljitu generativnu ulogu u evoluciji genoma. Iako utišavanje usporava stopu restrukturiranja genoma na razini evolucijske vremenske skale, pokretni elementi u genomu su nesumnjivo poveli evoluciju genoma u različitim pravcima.