PROCEEDINGS OF THE BALKAN SCIENTIFIC CONFERENCE OF BIOLOGY IN PLOVDIV (BULGARIA) FROM 19TH TILL 21ST OF MAY 2005 (EDS B. GRUEV, M. NIKOLOVA AND A. DONEV), 2005 (P. 79–89)

MOBILE ELEMENTS AND EVOLUTION OF MOLECULAR REGULATORY SYSTEMS

*Evelina Daskalova**

Dept. of Plant Physiology and Molecular Biology
University of Plovdiv, Tzar Assen 24 str, 4000 Plovdiv, Bulgaria
e-mail: eve_das@pu.acad.bg

ABSTRACT. One of the latest developments in the ideas about molecular evolutionary process is that evolution is a natural system engineering process. The natural genetic engineering has the potential to create hierarchical subsystems and complex networks of genome regulation (Shapiro, 2005).

According to most researchers in this field, mobile elements (MEs) are perhaps the most important tool of this natural engineering. They are universal agents of evolutionary change. During evolution, they have created and modified the genome architecture (in global and local scale), the gene structure and regulation, and they do so in genomes of recent organisms. MEs also became the main driving force for cells to develop major epigenetic defense mechanisms. Later in evolution, these defense mechanisms have been utilized as basic cellular regulatory systems. In this article, we review and synthesize the latest ideas about MEs and cellular regulatory systems and propose a hypothesis about MEs as key units, interconnecting global and local scales of genome regulation.

KEY WORDS. mobile elements, evolution, regulatory interactions, microRNA.

INTRODUCTION

What are MEs?

Mobile genetic elements or MEs (transposable elements, transposons) are DNA sequences that can change their genomic location. They move (transpose) through the genome in a process called transposition. The mechanisms of transposition are complex and diverse, but all they result in two basic modes of moving:

– '**cut and paste'** mode of transposition – MEs move from one location (donor site) to a new location (acceptor site), excising themselves from the donor site. The copy

number of MEs remains more or less constant. This mode of transposition is less invasive.

– '**copy and paste'** mode of transposition – MEs transpose to the acceptor site, leaving their copy at the donor site. This mode of transposition results in accumulating great numbers of elements' copies (repeats), occupying in some cases over 2/3 of the genomes.

The origin of MEs is unclear. It is probable that different groups have appeared in genomes in different ways. There are two main groups (classes) of mobile elements (*Figure 1*): retroelements and DNA transposons.

Class I includes all mobile elements that encode or originate from the activity of the enzyme reverse transcriptase (RT). RT is an ancient enzyme, perhaps a relic from the ages of transitions between the old "RNA-world" and recent "DNA world" of genome evolution. Its function is to create a DNA copy on RNA template, the copy then reinserts in the acceptor site becoming a new retroelement copy. Massive presence of RT in all eukaryotic genomes was one of the greatest surprises after completing the first sequencing projects: in the pre-genomic era RT was thought to be an exotic exception of the central dogma, an enzyme restricted mostly to retroviruses. It was a big surprise when it became clear that over 45% of human genome consists of retroelements and related sequences and expresses significant RT activity. Similar is the situation in most eukaryotic genomes.

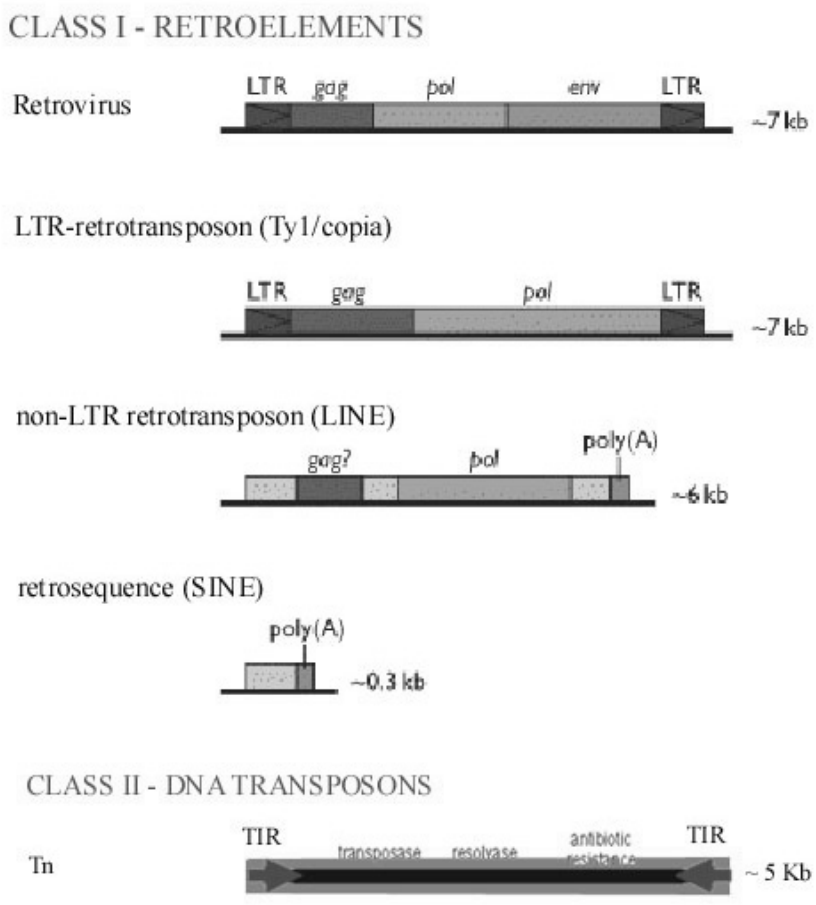

Figure 1. *Basic types of mobile elements.*

Eukaryotic retroelements are divided into three major groups (Kazazian, 2004, Graur&Li, 2001):

- **LTR transposons.** In their structure and mode of transposition they resemble retroviruses, like them, they are flanked by long terminal repeats (LTRs), containing promoters and other important signals for transposition. According to many authors, retroviruses are retroelements that have acquired an *env* gene for building an envelope that allows them to leave the cell and organism and to invade different individuals and even species. All eukaryotic genomes contain also endogenous retroviruses (ERVs) which are retroviruses without an *env* gene, so they cannot cross the cellular barriers and leave the organism.
- **Non-LTR retrotransposons.** A typical group of non-LTR retrotransposons are LINE-1 (long interspersed nuclear elements–1, or L1) elements of mammals. Full length non-LTR retrotransposons are 4 to 6 kb long and usually have two open reading frames (ORFs), one of which encodes the RT. There are hypothesis that the origin of non-LTR retrotransposons is connected to bacterial group II retrointrons
- **Retrosequences.** LTR and non-LTR elements are **autonomous retroelements**, because they encode the activities necessary for their retrotransposition. In contrast, retrosequences are **non-autonomous** – they do not encode RT and other retrotransposition proteins; that's why retrosequences use RT and other enzymes from the autonomous elements for their retrotransposition. Of the most abundant retrosequences are SINEs, typified by *Alu* elements of primates.

Retroelements of prokaryotes are different from eukaryotic elements. There are three main types: retrons, retrointrons and retroplasmids. Of them, **retrointrons** are the most interesting because it is thought that they are connected with the origin of both the sequences preceding recent spliceosomal introns, and of the basic sequences of the five snoRNAs that build the spliceosome.

All retroelements transpose by 'copy and paste' mode of transposition, so they tend to accumulate great number of repeats, especially in eukaryotic genomes. Bacterial and archaeal retroelements are not as invasive as eukaryotic ones; they have much less copy number and do not accumulate in prokaryote genomes. Until recently, it was even thought that prokaryotes do not have any retroelements.

DNA transposons are also classified in several types which we will not discuss here. They also contain autonomous and non-autonomous elements.

What do MEs do in the genomes?

They are extremely powerful agents of genome change. MEs can do practically everything in genomes. The spectrum of ME-induced mutations is broader than that of any other mutator mechanism – the genetic changes induced by MEs range from modifications in the size and arrangement of whole genomes to substitutions, deletions, and insertions of a few or single nucleotide. Translocations and inversions; pseudogene formation; genome duplication and exon shuffling, chromosome inactivation and restructuring; speciation – this is an non-complete list of ME activities. (Kidwell and Lisch, 2001). We can see among these activities the basic

mechanisms of evolving new traits – in fact, the basic mechanisms of molecular evolution. And here comes the paradox:

They are practically invisible. Given their power and abundance, transposable elements would rapidly randomize genome order. But they do not do this – why? In fact, most of the time, most of the elements have no phenotypic expression (this is the main reason for their late discovery in 1940s). There are two complementary explanations of this paradox:

They are effectively self-controlling. All MEs have various mechanisms that restrict their own transposition: nonrandom distribution (site-specificity, tissue and cell type specificity); expression of regulatory proteins (repressors of transposition); increased sequence variation and imprecise excision etc. The last two mechanisms explain the abundance of defective ME copies in genomes that cannot transpose anymore.

They are object of cellular epigenetic regulation. MEs are targets to all epigenetic cellular regulatory systems. Many of them are actively methylated and/or heterochromatinized; or are targeted for RNA interference. Some of their RNA transcripts are intensely edited. But what's more interesting, the cells may owe all these regulatory activities to MEs.

DNA methylation has long been viewed as an essential component of the epigenetic regulation of mammalian genes that evolved to make differentiation in complex organisms possible. However, that has recently been challenged by the view that **methylation evolved as a genomic defense against mobile DNA** (Barlow 1993; Yoder et al. 1997). These authors argue that methylation of host genes in fact is evolved as a system for repression of transposon activity. Similarly, it has been proposed that posttranscriptional silencing in plants (which is often associated with DNA methylation), and RNA interference in animals, are clearly related to the defense of eukaryotic genomes against repetitive or mobile elements (Matzke 1998, Plasterk 1999).

So MEs are controllable, but not completely. As it becomes clear in last years,

They are stress-inducible. Experimentation with a number of different mobile element systems has shown that they can be activated temporarily by response to particular conditions: from blockage of normal chromosome separation during embryonic development to osmotic and other physical stresses, to oxidative starvation stress during adaptive mutation etc. (Shapiro, 2005)

All there characters show that MEs are perfect evolutionary instrument – controllable and inducible. Different groups of organisms, however, have different fate according to the extent and mode of which they make use of MEs' activity.

Synthesis

Two evolutionary strategies

Recent prokaryotes and eukaryotes are examples of two different evolutionary strategies in relation to MEs. Basic evolutionary forces that have shaped prokaryotic and eukaryotic genomes are completely different. As a result we have two

evolutionary and regulatory systems based on different rules. They are summarized in ***Table 1***.

Table 1. The evolutionary strategies of prokaryotic and eukaryotic genomes are different.

Traits	**Prokaryotes**	**Eukaryotes**
Ratio genic/nongenic DNA	Over 90% protein coding genes, little nongenic DNA, no pseudogenes, almost no repeats	Up to 5-10% protein coding genes, abundance of nongenic DNA - pseudogenes and different repeats
Gene structure	Simple, continuous, no or little introns	Modular, almost all genes contain introns and long UTRs
Transcriptome	Simple, mostly mRNAs, little regulatory and other ncRNAs, no data for complex RNA based regulatory networks	Great abundance of noncoding transcripts: regulatory, antisense, mobile etc; mRNA is a small fraction of the transcriptome; complex RNA based regulatory networks
Mobile elements	Little number of copies, DNA transposons and 'cut and paste' transposition prevailing	Great number of copies, many repeats, retroelements and 'copy and paste' transposition prevailing
Basic evolutionary processes	**Mutation based:** rapid proliferation and accumulation of mutations and hypermutations **HGT based:** intensive acquisition of new genes from other species	**Repeat based:** high contents of repeats facilitates recombinations, translocations and shuffling, duplications and rearrangements of different scales **Network based:** complex network of RNA-based and other interaction-based regulatory processes
Main evolutionary tendency	Genome compactness, adaptivity	Genome expansion, complexity

As we see, prokaryotes and eukaryotes are very different in their mode of evolution. We can see also that the main processes of eukaryote genome evolution are repeat-based. The amount and genome distribution of repeats defines the location and frequency of different recombination processes (equal and unequal crossing-over, gene conversion), creating new genetic combinations and proper "formatting" of protein coding sequences which are minor part of the eukaryotic genomes. As Shapiro says, "Since dispersed repeats influence both coding sequence expression and physical organization of genomes, … repeat distribution reflects the establishment of a system architecture required for effectively integrated genome functioning". In addition, it is very important to realize that eukaryote evolution is

network based – eukaryotes are developed complex regulatory networks, based on complex combinations of DNA-protein, DNA-RNA, RNA-RNA, RNA-protein and protein-protein interactions. Certainly, prokaryotic regulation is also based on interaction networks but their interactome is far more simple and "hard-wired" than eukaryotic one.

MEs – the focus of genome evolution and regulation in eukaryotes

As we can see also, the general difference between prokaryotes and eukaryotes is based on the different set and amount of MEs, and accordingly, on the different prevailing mode of transposition. So we can summarize the following basic points:

- MEs are universal DNA/RNA based (genetic) elements; practically all recent cellular regulatory systems are connected to them in one way or another. MEs appear to be a natural focus (point of interconnection) between the systems working at least at two basic regulatory levels – transcriptional and posttranscriptional ***(Figure 2)***.
- Being a focus of interconnection, MEs as genome formatting entities define not only the basic structure of the genome, but also the basic mode of interaction between different cellular regulatory systems.

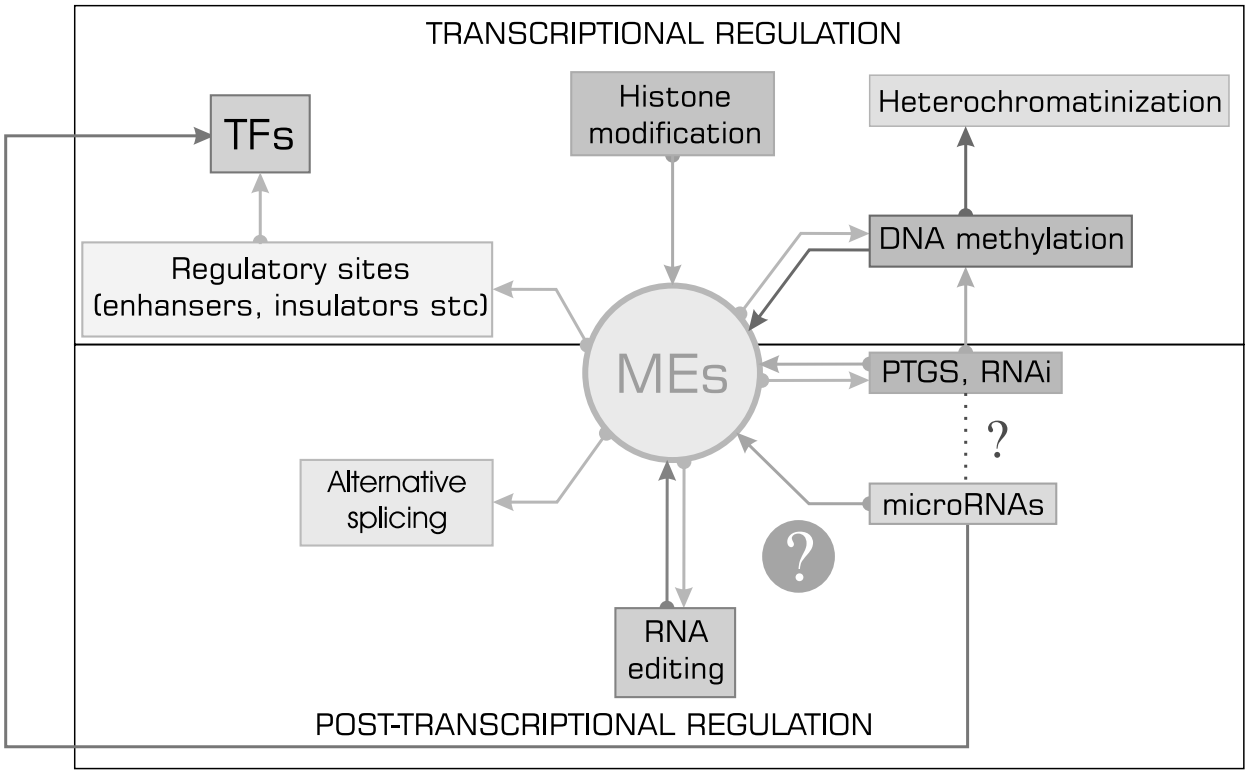

Figure 2. *MEs are focus of interaction of practically all cellular regulatory systems on transcriptional and post-transcriptional levels.*

- they are direct target of regulatory interactions (arrows of relevant color)
- they can induce regulation of genes and genomes through the respective system (violet arrows)

- in evolution, they have been a driving force for the creation of the relevant system (pink-dotted arrows) indicate

Hypothesis

We developed the synthesis of recent ideas and data about MEs and their interaction with the cellular regulatory systems in the following hypothesis:

- **MEs are natural coordinating link between the global and local levels of regulation of the cellular processes.** The activity of MEs itself and the cellular responses to their activity coordinate the global (genome distribution) and local (changes in gene expression of particular genes) modes of regulation.
- **The connection between global and local modes of regulation is realized through the interaction between autonomous and non-autonomous MEs.**
- **In these interacting pairs, autonomous counterparts are responsible for the global regulation, while non-autonomous counterparts regulate mostly the local changes in gene expression.**

As we noticed before, all classes of mobile elements contains autonomous and non-autonomous members, and what is most important, non-autonomous elements rely for their transposition on the activity of autonomous ones. The hypothesis is supported by recent research data:

1. All major classes of mobile elements have particular members that generate and maintain specific classes of non-autonomous elements.

The most famous example are the non-LTR elements (LINEs) which generate and amplify various SINE elements. Other classes however have also such specific pairs: Autonomous LTR retrotransposons (exact type is still unknown) generate so called TRIM elements, and some DNA transposons generate MITE elements (Casacuberta&Santiago, 2003).

2. Autonomous and non-autonomous members of these pairs have radically different behavior to cellular RNA- and protein coding genes.

Autonomous elements generally escape genes and usually insert away from gene-rich areas; nevertheless they prefer to insert their non-autonomous counterparts near and even within genes. L1 elements and *Alu* sequences (primate-specific SINEs they generate) are typical example. L1 element preferentially insert in AT rich, gene poor genome positions; in contrast, they often insert the *Alu*s in close proximity to protein coding genes, in introns, 5' and 3' UTRs, and even in CDSs. Recent data show that up to 75% of human genes have one or more *Alu* inserts in their introns and/or UTRs (Kim et al, 2004). Similar situation is observed with MITE elements. They are short non-autonomous elements found mostly in plants and invertebrates. Their autonomous distributor is still unknown, but, like *Alu*s, they tend to insert near and in genes, with most preferred site the 3'UTRs. LTR-element derived TRIMs show similar behavior.

3. Non-autonomous counterparts are generally more abundant than autonomous ones.

This is most illustrative in the case of MITEs. DNA transposons usually are much less invasive than retroelements, they hale less number of repeats and their accumulation is slow. Despite of that MITES, which are DNA transposon-derived, have amounts and dynamics of accumulation typical for retroelements.

1. The most abundant pairs of interacting elements are taxon- and species-specific and are replaced with other pairs in evolutionary history of the taxon; there changes often coincide with major evolutionary events in the history of the particular taxon (divergence, speciation).

For example, in evolution of primates, the age of appearance of L1-*Alu* pair coincides with the primate divervence.

2. Non-autonomous counterparts are points of interconnection of regulatory systems acting mostly at local scale, while autonomous counterparts participate and induce regulatory interactions at global scale.

For example, *Alu*s inserted in UTRs and CDS in often contain splicing sites, thus mediating the alternative splicing. Their transcripts are often subjects to intensive and specific A to I editing (Kim et al, 2004; Dutko et al, 2005). CpG islands, residing within *Alu*s are often differentially methylated (Jurka, 2004; Xing, 2004).

On their turn, autonomous counterparts more often induce and mediate global epigenetic modifications – methylation, heterochromatinization, RNAi. As it has been shown recently, there is also a specific histone-modifying enzyme, "responsible" for the control of MEs – the lymphoid-specific helicase (LSH) (Huang et al, 2004). L1s could also act as "molecular rheostats" modifying the expression of large gene sets (Han et al, 2005)

PREDICTIONS AND PRIMARY RESULTS

MEs and the miRNA system

Based oh the hypothesis described above, it can be predicted that such pairs of MEs interact with another local systems of regulation of gene expression, and mediate their globalization and network support. In particular, we want to test if there is interaction between the micro RNA (miRNA) regulatory system and mobile elements. At this time, there is still no evidence about such interaction, and if it becomes a fact, it will be a powerful support of the idea that MEs are universal interconnection link between cellular regulatory systems. If it is proved to be true, it will also elucidate the evolutionary connection between miRNAs and RNA interference system. It is possible that **miRNA system originates from siRNA system.** It is possible that in recent organisms MEs' activity is regulated through both processes (by siRNA interference and by miRNA translation repression), and that the cells can switch between siRNA and miRNA regulation of MEs depending on environmental conditions or other factors. It is also possible that retroelements are regulated through some kind of different process having intermediate features of siRNA- and miRNA-regulation.

If this prediction is proved, it means that **cells have achieved an exciting range of regulatory abilities by combining the action of siRNA/miRNA system and the dynamics of MEs.** While the regulatory mode of siRNA system is '1/0 (lack of

repression/complete repression), the regulation through miRNAs probably allows intermediate levels of repression and could maintain low but significant activity of some MEs if this is necessary to the cells. This could be universal, sensitive to environmental changes, and evolutionary productive mechanism for controlling the genome activity. Perhaps this mechanism is a part of a global and complex system of genome regulation on cellular and higher levels of organization. MEs could assure the sensitivity to environmental signals, and siRNA/miRNA system itself could execute the cell's decision whether, to what extent and in what mode the action of retroelements to be allowed in certain conditions.

The examination of above hypothesis is main object of work of our research team.

We are testing the following possibilities of interaction between miRNAs and MEs:

1. miRNAs are coded in the same sequences of MEs they regulate. In this case we expect to find in the sequences of MEs sequences and structures, characteristic for *miRNA precursors* (pre- and pri- miRNAs).

2. miRNAs are coded in other parts of the genome, where are transcribed and processed to target specific MEs. Then sequences of MEs are expected to contain miRNA target sites.

3. There are miRNAs encoded in MEs' sequences **inserted** in other cellular genes – some **non-autonomous MEs could be donors of miRNA regulatory motifs and target sites.** Than we have to test ME-containing cellular genes for such motifs.

4. MiRNA-like sequences encoded by autonomous MEs could also control other target genes. Some of them may be regulators (repressors) of miRNA activity, for example encoding miRNA antisense transcripts (anti-pre-miRNAs).

If one or more of these cases is proved, it may reveal a whole new level of regulatory abilities.

- cells could switch between siRNA and miRNA regulation of MEs depending on environmental conditions or other factors
- **Interaction of MEs and miRNAs could mediate the globalization and network support of the miRNA based system**

Primary results and discission

Our work is at the beginning phase but we have some promising results.

We made an initial collection of about 500 sequences of mobile elements of several types. The collection contains mostly MITE and *Alu* elements. Both types are non-autonomous, probable genetic symbionts often inserting in UTRs, introns and CDS of genes. We have also some L1 elements – the Alu transporters.

Then we performed a series of BLAST homology searches against NCBI/Genbank database. Most interesting results so far we obtained with MITE elements. .

- MITE copies in a single genome are highly similar
- MITE copies in different plant genomes show similar pattern of homology. For example, all "Tourist" type MITE elements from *T. aestivum* and *O. sativa* have 1 to

5 17-30 nt long sequences with 90-100% similarity to parts of different genes and genomic clones from *Arabidopsis, Sorghum, Glycine* and *Zea mays*.

- The most abundant conserved sequences have length in range of 20-25 nt – the length specific for mature micro RNAs.

Then we used these conserved sequences for searching homology to known mature miRNAs against the ***miRNA registry*** database (http://www.sanger.ac.uk/Software/Rfam/ mirna/index.shtml). There were multiple matches to known miRNAs but the degree of similarity was not very high. More promising results were obtained when we used as a query not the conserved sequences but the whole MITE sequences (typically 200-500 nt long). There were again homologies, this time really significant, to hairpin-like mi RNA precursors. The most significant results we obtained for pre-miRNAs for *ath-MIR169c, ath-MIR167d, ath-MIR165a* and *osa-MIR169m*.

This research is still at the initial phase. One of the obstacles is that there are no well annotated, ME-specific databases. **The results obtained so far show that there is obviously a connection between some MEs and miRNAs.** At this time, however, we can not say what exactly that connection is. Initially we thought that MITEs and *Alu*s are very suitable to be donors of regulatory motifs – miRNA target sequences. It could be really the case, but the strong similarity with pre-miRNAs shows that they can also encode for miRNAs themselves. So two or more of the tested possibilities can exist simultaneously**. There may be a complex network of interactions between MEs and miRNAs** in living cells**.**

We will continue our work with *Alu* elements and more MITEs, as well as with L1 elements and some LTR elements. We plan also to use a new tool for finding miRNA target sites, recently created from two of our PhD students.

CONCLUSION

I have always been intuitively attracted by MEs because they give some sense of freedom.

Recent developments of evolutionary thought slowly displace the old, deterministic and gene-centric views with some more integrated more holistic and spontaneous ideas. MEs create some kind of „organized genomic chaos“. Surprisingly it may appear that more complex the system, more it relies on such organized chaos. (Prokaryotes, for example are more „ordered“ and „hard-wired“ than complex multicellular eukaryotes).

The more complex is the organism, the more important the holistic and epigenetic component of its structure, function and behavior. And, in the case of us, humans, there is one more basic component of epigenetic freedom – the consciousness.

REFERENCES

GRAUR, D. and W-X. LI., 2001. Fundamentals of molecular evolution. Elsevier science, New York.

SHAPIRO, J.A., 2005. A 21st century view of evolution: genome system architecture, repetitive DNA, and natural genetic engineering. *Gene 345 91–100*

KIDWELL, M.G., and D.R. LISCH., 2001. Perspective: Transposable elements, parasitic DNA, and genome evolution. *Evolution, 55(1), , pp. 1–24*

HAIG, H., KAZAZIAN, JR., 2004. Mobile Elements: Drivers of Genome Evolution. *Science, vol 303, 12 March, 1626-1632.*

HAN, J.S., S.T. SZAK & J.D. BOEKE., 2004. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature, vol 429, 20 May, 268-274*

CASACUBERTA, J.M., and N. SANTIAGO, 2003.. Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes. *Gene 311 1-11.*

HUANG, J., FAN, T., YAN, Q., ZHU, H., FOX, S., ISSAQ, H.J., BEST, L., GANGI, L., MUNROE, D., MUEGGE, K., 2004. Lsh, an epigenetic guardian of repetitive elements. *Nucleic Acids Res. Sep 24;32(17):5019-28. Print 2004*

KIM, D.D., KIM, T.T., WALSH, T., KOBAYASHI, Y., MATISE, T.C., BUYSKE, S., GABRIEL, A., 2004. Widespread RNA editing of embedded alu elements in the human transcriptome. *Genome Res. Sep;14(9):1719-25*

DUTKO, J.A., SCHAFER, A., KENNY, A.E., CULLEN, B.R., CURCIO, M.J., 2005. Inhibition of a yeast LTR retrotransposon by human APOBEC3 cytidine deaminases. *Curr Biol. Apr 12;15(7):661-6*

JURKA, J., 2004. Evolutionary impact of human Alu repetitive elements. *Curr Opin Genet Dev. 2004 Dec;14(6):603-8.*

XING, J., HEDGES, D.J., HAN, K., WANG, H., CORDAUX, R., BATZER, M.A., 2004. Alu element mutation spectra: molecular clocks and the effect of DNA methylation. J *Mol Biol. Nov 26;344(3):675-82*