

## Tandem repeats in mammalian genomes

D. Ostromyshenskii<sup>1\*</sup>, O. Podgornaya<sup>1,2</sup>

<sup>1</sup>*Institute of Cytology RAS, Saint Petersburg, Russia*

<sup>2</sup>*Saint Petersburg State University, Saint Petersburg, Russia*

\* e-mail: [necroforus@gmail.com](mailto:necroforus@gmail.com)

**Key words:** tandem repeat, satellite DNA, heterochromatin

*Motivation and Aim:* Heterochromatin definitely plays an important role in chromatin 3D structure, but the detailed mechanisms not determined. Tandemly repeated sequences are the DNA class, which is absent in prokaryote but appear in eukaryotes. Tandem repeats (TR, or satDNA – satellite DNA) is the mostly fast evolving genome component. It is impossible to determine TR role and functions in the genome housekeeping without their classification and annotation. We are going to classify TR in the mammalian genomes with similar degree of assembly available in databases and to check the pattern of TR evolution *in silico*.

*Methods and Algorithms:* The modified pipeline of the one used previously in our Lab [1] applied to the several mammalian genomes. For TR searching we used mammalian genomes assembly from [www.ncbi.nlm.nih.gov/assembly](http://www.ncbi.nlm.nih.gov/assembly) with N50 metrics > 40 Kb. Totally we searched TR in assembled genomes of 57 species of 11 mammalian orders. For each genomes we accounted consensus sequence of each TR family, monomers length, GC-content and variability within the TR arrays. For *Mus musculus*, *Cricetulus griseus*, *Mesocricetus auratus* and *Sus scrofa* we confirmed *in silico* prediction by molecular biology methods such as FISH and PCR.

*Results:* We found full sets of TR in 57 mammalian genomes assembly. Comparison of sets of TR in the different genomes showed that despite different primary sequences of TR of different genomes, the distribution of TR within each genome according to GC-content, monomer length and variability within the TR arrays is similar. Almost in all genomes, the family of major TR has been found. In different groups of mammals, it can be formed as AT-rich TR (for example, rodents and primates) or GC-rich TR (for example, carnivorans and artiodactyls), but always it is characterized by organization in high order repeats, long monomers, centromeric-pericentromeric localization. The second family of TRs by representation in genome is often formed by TR with a relatively short monomer, more complex arrays and has chromosome-specific variants (for example, the HS1-4 human satellite or the mouse TRPC-21A-MM family).

*Conclusion:* By comparison of the TR set from the genomes we show that the TR fields distribution is similar in the genomes. Such a set reflects the “bar-code”, which predetermine the hierarchy of chromosome domains associations. We expect that heterochromatic “bar-code” made up of the TR could be the base for the hypothetical General Morphogenetic program attributed to the heterochromatin.

*Acknowledgements:* This work was supported by the granting program of Russian Foundation for Basic Research (No. 18-34-00238) and the granting program ‘Molecular and cell biology’ of the Russian Academy of Sciences (No. 01.2.01457147).

### References

1. Komissarov A.S., Gavrilova E.V., Demin S. et al. (2011) Tandemly repeated DNA families in the mouse genome. *BMC Genomics*. 12(1):531-552.