GC content of transposons and of their (animal) host genomes

*Radka Symonova1, 2,\** , *Jan Kubečka2* and *Dominik Matoulek3*

1Technical University of Munich

2Institute of Hydrobiology, Biology Centre CAS, České Budějovice

3University of Hradec Kralove, Czech Republic

\*Presenting author

Fish genomes are AT/GC homogenous in comparison with heterogeneous genomes of mammals and birds. The heterogeneity in mammals is apparent on DNA sequences as well as on chromosomes upon a suitable staining. With our finding that both extant genera of ancient ray-finned fish gars possess AT/GC heterogenous genomes similarly as mammals do, we initiated a search for possible reasons of this compositional heterogeneity. In this effort, we developed a Python tool segmenting genomes assembled to the chromosome level according to their GC% and repetitive content (soft-masked DNA) and plotting these values in a single plot. The plot is represented by a colored profile of GC% values along the chromosomes, where the color represents the gradient between the fully repetitive DNA (green) and fully non-repetitive DNA (red) with the sliding window of the size 1 kb as the default but freely adjustable. Our survey across numerous freshwater fish assemblies with this tool showed the repetitive fraction as the potentially AT/GC homogenizing factor in lower vertebrates. At this stage, the quality of the soft-masking procedure is the most crucial step. Otherwise, this tool can be used in any other group with genomes assembled to the chromosome level, e.g. plants and fungi, to explore the role of the repetitive fraction in other major lineages beside vertebrates.

Acknowledgement

The study was supported by the European Union within ESIF in frame of Operational Programe Research, Development and Education (project no. CZ.02.1.01/0.0/0.0/16\_025/0007417 administrated by the Ministry of Education, Youth and Sports of the Czech Republic)