

Genomics of *Blastocrithidia*, the trypanosomatid with all three stop codons reassigned

Anna Nenarokova^{1,2}, Kristína Záhonová^{1,3,4}, Serafim Nenarokov¹, Vyacheslav Yurchenko^{3,4}, and Julius Lukeš^{1,2}

1. Institute of Parasitology, Biology Centre, České Budějovice (Budweis), Czech Republic

2. Faculty of Science, University of South Bohemia, České Budějovice (Budweis), Czech Republic

3. Life Science Research Centre, Department of Biology and Ecology, Ostrava, Czech Republic

4. Institute of Environmental Technologies, Faculty of Science, University of Ostrava, Ostrava, Czech Republic

Recently, two groups of protists bewildered molecular biologists: trypanosomatid *Blastocrithidia* and several ciliate species were shown to reassign all three stop codons for encoding amino acids. In these organisms at least one stop codon has ambiguous meaning: it acts as a sense codon in some cases and as a termination codon in the other. This finding challenges the established view of protein synthesis termination, one of the most basic cellular processes. However, the mechanisms of this reassignment and translation termination in such systems still remain speculative. *Blastocrithidia* represents an ideal model system for studying this phenomenon. It belongs to kinetoplastids, a well-studied protist group, which include model objects, such as *Trypanosoma* and *Leishmania*, with available complete genomes and established laboratory methods and techniques. Unlike ciliates, which are well-known for stop codon reassignment, all known kinetoplastids aside from *Blastocrithidia* genus have the canonical nuclear genetic code. Thus, looking to this lineage, we can trace the main steps leading to the emergence of such system.

Here, we have sequenced and analyzed genomes of two cultivable *Blastocrithidia* species and *Leptomonas jaculum*, the closest relative of *Blastocrithidia* with the canonical genetic code. We have created a new software for annotation of *Blastocrithidia* genome, as existing annotation programs are not able to deal with ambiguous stop codons. This allowed us to look at the reassigned stop codons from a wider perspective to see the general trends in their features and distribution. The ultimate goal of our study is to address the following intriguing questions: How does translation termination function without defined stop codons? How do numerous reassigned stop codons influence translation? What are the prerequisites of this lineage that made the reassignment possible? What could be the intermediate steps between a system with a standard genetic code and one with all three stop codons reassigned?