DOI 10.18699/SBB-2022-032

Grass frog (*Rana temporaria*) tandem repeats in the genome and oocytes' transcriptome

Dikaya V.[1]*, Travina A.[1], Komissarov A.[2]

[1] *Institute of Cytology, RAS, St. Petersburg, Russia*
[2] *Applied Genomics Laboratory, SCAMT Institute, ITMO University, St. Petersburg, Russia*
* *animeshizik@gmail.com*

Key words: tandem repeats, oogenesis, satellite DNA, amphibian

Motivation and Aim: Tandem repeats (TR) or satellite DNAs' transcripts are of overwhelming majority at lamp-brush stage of oogenesis [2]. The fate of transcripts in late oogenesis stages remains obscure. TR of most organisms are not known because of TR species specificity. *Methods and Algorithms:* We used grass frog genome published (NCBI GCF_905171775.1) and oocyte nuclei transcriptome got *de novo*. The cDNA was synthesized using a Mint-2 kit (Eurogen) and sequenced using Illumina HiSeq 2500 in Saint-Petersburg State University Resource Center. Adapters and optical duplicates were purified using the v2trim and rmdup programs. Ribosomal RNA was filtered by SortMeRNA. The alignment of the obtained message RNA reads was performed by the STAR program. *De novo* assembly of the transcriptome was performed with Trinity, Metabat2 and CheckM were used to detect contamination. The frog transcriptome was aligned to the detected genomes of the found contaminants. Matched regions were excised using the Python3 script from the assembly. The quality of the assembly was assessed by QUAST and BUSCO. Transcriptome annotation was performed using EggNOG, BLAST, Seq2Fun and Trinotate. In the genome, gene prediction was implemented by AUGUSTUS. Genome content of TR and transposable elements (TE) was determined with RepeatModeler2 and RepeatMasker. The search for tandem repeats was done by using the developed De Bruijn graph-based algorithm in Python2.7 and C++ in the genome and transcriptome. These repeats were localized on chromosomes using fluorescence in situ hybridization (FISH).
Results: The transcriptome was collected in 41 121 contigs. However, there were few annotated sequences in it. Trinotate annotated ~30% of the contigs in transcriptome. The BLAST annotation was only ~10%. Transcripts responsible for signaling molecules were most prevalent in the Trinotate annotation. Gene Ontology analysis revealed many transcripts related to nucleic acid binding, nucleus components, ribosome biogenesis, gene expression, and ribonucleoprotein complex formation. Total repeats content in genome is ~70% (~50% unclassified) and ~50% (~41% unclassified) in transcriptome. Tandem repeats are underrepresented in the genome and transcriptome but we found six well comprised TR ranging in length from 5 Mb to 114 Mb. This six TR were mapped on the frog chromosomes and localized to the pericentromeric regions with relative chromosome-specificity.
Conclusion: It was obtained that unknown, apparently species-specific repeats have a large number of representations in the data. The mode of TR transcription is under the data found on birds in similar studies: TR with genus specificity (TR 494A (S1a)) transcribed three orders less than species-specific 47A and 35B [1, 2], which probably reflects the feature common for TR transcription in gametogenesis.
Acknowledgements: This work was supported by Russian Science Foundation (grant No. 19-74-20102).

References

1. Komissarov A.S. et al. New high copy tandem repeat in the content of the chicken W chromosome. *Chromosoma*. 2018;127:73-83.
2. Trofimova I., Krasikova A. Transcription of highly repetitive tandemly organized DNA in amphibians and birds: A historical overview and modern concepts. *RNA Biology*. 2016;12(13):1246-1257.