EMBL-EBI

Dr Anton Petrov
RNA Resources Project Leader
Tel.: +44 1223 492550
E-mail: apetrov@ebi.ac.uk

1 September, 2018

Evaluation of a PhD dissertation by Agnieszka Thompson entitled "Biogenesis and function of tRNA-derived fragments"

The goal of Ms Thompson's dissertation was to create an online resource for analysing tRNA-derived fragments (tRFs) and improve our understanding of the tRF biology in the model organism *Arabidopsis thaliana*. tRFs originate from specific cleavage events of tRNA sequences and have been observed in a wide range of organisms. They have been implicated in several human diseases, including cancer and Parkinson's disease. Many laboratories around the world study tRFs in order to identify their functions and regulatory potential. A database dedicated to plant tRFs fills an important niche and is a timely contribution to the field.

Ms Thompson's dissertation provides a comprehensive overview of the current knowledge about tRFs, including existing tRF databases, tRF biogenesis, the role of RNA modifications, and an introduction to the key proteins discussed in the work. The text is clear and easy to follow, and it includes useful figures and relevant literature references.

> **Question 1.** Although the study of tRNA-derived fragments is indeed an established area of research, there are still potential concerns that at least some tRFs may, in fact, represent random degradation products of tRNAs and not a separate non-coding RNA class (for example, see a recent review by Keam and Hutvagner, PMID:26703738). Would it be possible to address such concerns either in the text of the dissertation or during the defense?

The methodology takes full advantage of the existing bioinformatic approaches and is explained with sufficient level of detail. The work is based on an analysis of 300 public and in-house RNA-seq datasets from *Arabidopsis thaliana* that are pre-processed and aligned to the TAIR10 reference genome using Bowtie2. The alignment step allows for multiple mismatches, which helps to retain the reads originating from the modified tRNA sequences that tend to be excluded from the other tRF resources.

> **Question 2.** Given that short reads may align in numerous locations when multiple mismatches are allowed, the post-processing steps have a big impact on the downstream results. Section 4.3 mentions a post-mapping scoring system that is meant to eliminate false positives based on the number of mismatches, insertions, and deletions, however, it is not clear how well it performs in terms of filtering out the spurious hits. Could more details about this step be provided?

One of the main outcomes of this project is the creation of an online resource called tRex. The database contains over 1.4 million tRFs identified in the RNA-seq datasets. The tRFs aligned to the annotated tRNAs were classified using an enhanced version of the previously available tRF classification with the addition of several more specialised classes. The users can browse tRFs by sample and tRNA type as well as explore tRFs in multiple mutant plants and study their response to different environmental conditions.

The tRex website is very well designed and is easy to use. Importantly, the tRFs can be displayed overlayed on top of the tRNA secondary structure in standard orientation, which puts tRFs in the structural context and facilitates the comparison of multiple tRFs. This is a very useful feature that is not available at other resources, such as tRFdb or MINTbase.

The tRex database also contains tRNA modifications predicted using HAMR, as well as protein and transposable element targets that were predicted using a new method based on a modified Blast algorithm and hybridisation energies calculated with RNAfold.

> **Question 3.** What strategy was used to validate the accuracy of the target prediction method and its applicability to tRFs? Are there any experimentally confirmed tRF targets that could be used to validate the target prediction methodology or compare its predictions with the tRF targets from the literature?

The library of RNA-seq data from mutant plants allowed to compare tRFs in samples with mutations in several genes and focus on several proteins, such as RNS2, DCL2/RDR6/RDR3b and HYL1 that are involved in the tRF biogenesis pathway. In addition, the differential expression analysis performed using DESeq2 led to the identification of tRFs that respond to environmental stresses, such as high salinity and drought. These data provide interesting leads for experimental validation in the future.

Despite the questions above, the work undertaken by Ms Thompson is significant and novel, and it represents an important contribution to the field. One of the strong aspects of the dissertation is the combination of bioinformatic analysis and database development with biological applications of the data. I hope that the tRex resource is regularly updated with new RNA-seq datasets and even potentially expanded to other organisms in addition to *Arabidopsis thaliana*, using the analysis framework developed by Ms Thompson.

In conclusion, I am convinced that the dissertation presented by Ms Thompson fulfils the customary and legal criteria required of PhD dissertations in Poland and urge the Council of the Faculty of Biology at Adam Mickiewicz University in Poznan to grant the candidate access to further stages of the procedure.

Sincerely,

Anton Petrov, PhD
RNA Resources Project Leader
EMBL-EBI