

Laboratoř proteinů s biotechnologickým potenciálem

Laboratory of protein biotechnology

Ústav biochemie a mikrobiologie, Vysoká škola chemicko-technologická v Praze, Technická 3, Praha 6, 166 28

Department of Biochemistry and Microbiology, Institute of Chemical Technology, Prague, Technická 3, Prague 6, 166 28, Czech Republic

<http://biomikro.vscht.cz/cz/research/groups/protbiotech/> • Tel.: +420 22044 3028 • +420 22044 5109 • Fax: +420 22044 3010

Běstvína, 13 June 2019

Assoc. Prof. Vojtěch Spiwok

Department of Biochemistry and Microbiology

University of Chemistry and Technology, Prague

Technická 3, Prague 6, 166 28, Czech Republic

E-mail: spiwokv@vscht.cz

Evaluation report on doctoral thesis

**Transcriptomics analysis of barley (*Hordeum vulgare* L.)
and wheat (*Triticum aestivum* L.): tool for crop improvement**
submitted by **Filip Zavadil Kokáš**

The thesis of Filip Zavadil Kokáš presents three plant transcriptomics projects. Two projects directly analyze data from RNA sequencing experiments. One project focuses on tool development. The thesis is written in a classical way (not as introduction + attached publications). The results of the thesis are published in two scientific articles in respected journals (*New Biotechnology*, *Journal of Computational Biology*). I very much appreciate publication of a data set (*Data in Brief*), which supports openness and interoperability of the results, but is currently not rewarded by funding agencies. At the Metacentrum web interface I found that the applicant has been hard working, using 498.7 CPU days.

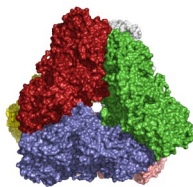
From the thesis it is not clear what is the contribution of the applicant and what is the contribution of other co-authors of studies discussed in the thesis. I believe this will be clearly explained in the thesis defense.

The introduction presents current state of the art in high-throughput nucleic acid sequencing. Tools for analysis of sequencing data are introduced with special focus on RNA sequencing (RNAseq) experiments. I liked very much the chapter describing designs of RNA sequencing experiments. Importance of this step of transcriptomic experiments is often underestimated and many projects fail due to poor design.

The first project presents analysis of RNAseq data to find the molecular rationale for drought tolerance in studied variants of barley. High number of differentially expressed genes was observed. The applicant analyzed these genes in the formalism of gene ontology (GO) terms. There are other approaches that can be used to analyze and interpret transcriptomics experiments. My question is for what type of phenomena (in terms of number of differentially expressed genes, complexity etc.) is gene ontology analysis suitable, according to applicant's experience.

My second series of questions is related to the tool SATrans, which is presented as the second project. I have downloaded the program from the web site of the university. I did not have any data in hand to test its functionality, however, as far as I can judge the code looks fully functional. I would like to ask on the future of the code. Does the applicant plan to maintain and further develop the code himself? Or does he plan to attract some potential co-developers to contribute to the code and support its continuity? Author may consider using some code sharing and collaboration platform such as GitHub or Bitbucket. I am not familiar with development of Perl programs, so I would like to ask whether Perl programs can be installed by some package management tools (something similar to Pip or Conda for programs written in Python). If yes, does the applicant plan to provide the tool this way? Finally, bioinformatics frameworks such as Galaxy or Chipster are becoming very popular. This growth in popularity is also driven by need of reproducibility of genomics data analysis because these frameworks record program settings, versions etc. How difficult can be integration of SATrans into such framework?

The third project studies molecular mechanism drought tolerance in wheat variants differing in root morphology. I understood that hexaploidy nature of wheat was a great challenge. However, there are other organisms for which genomics and transcriptomics research is highly challenging due to large genome size, large number of repetitive sequences, AT or GC enrichment and other problems. Can the applicant somehow put the barley transcriptomics experiment on some scale (for example from 0 to 10) and compare with other



Laboratoř proteinů s biotechnologickým potenciálem

Laboratory of protein biotechnology

Ústav biochemie a mikrobiologie, Vysoká škola chemicko-technologická v Praze, Technická 3, Praha 6, 166 28

Department of Biochemistry and Microbiology, Institute of Chemical Technology, Prague, Technická 3, Prague 6, 166 28, Czech Republic

<http://biomikro.vscht.cz/cz/research/groups/protbiotech/> • Tel.: +420 22044 3028 • +420 22044 5109 • Fax: +420 22044 3010

transcriptomics experiments? Finally, is there any idea how important are non-coding RNAs in the phenomena studied in the first and the third project?

Overall, the thesis presents a combination of state-of-the-art analysis of data from RNA sequencing experiments together with development of a computational tool for this purpose. The results were published in respected scientific journals or provided to scientific community as a software tool. The thesis is a valuable contribution to the field of plant genomics and it clearly demonstrates ability of the applicant to conduct an independent research. I therefore **strongly recommend the Department of Biochemistry, Faculty of Science of Palacký University to accept this Ph.D. thesis.**

Vojtěch Spiwok