



Centre of Plant Structural & Functional Genomics

Jan Bartoš, Ph.D.

bartos@ueb.cas.cz
phone: +420 585 238 711

Review of Ph.D. Thesis

Thesis: Transcriptomic analysis of barley (*Hordeum vulgare* L.) and wheat (*Triticum aestivum* L.): tool for crop improvement

Author: Mgr. Filip Zavadil Kokáš
Faculty of Sciences, Palacky University, Olomouc

Reviewer: Mgr. Jan Bartoš, Ph.D.
Institute of Experimental Botany, Olomouc

The aim of the Ph.D. thesis is characterization of transcriptome and analysis of differentially expressed genes with respect to drought stress in two important cereals barley and wheat. Formally, the dissertation meets demands placed on this type of work. The work is written in English, has a total of 168 pages and 21 appendices. Results of the work have been published in three articles accepted in New Biotechnology [IF 3.813]; Data in Brief; and Journal of Computational Biology [IF 1.191]. Filip Zavadil Kokáš is first author of two of them.

Literature overview is well structured. It summarises available information about high-throughput (mainly next-generation) sequencing techniques, design of RNA-Seq experiments and bioinformatics analysis of RNA-Seq data. This part also presents bunch of tools for quality control, short read alignment with emphasis on splicing, quantitative transcriptome analysis and differential expression and finally functional annotation of transcripts. While, as a whole, the overview has a compact impression, some details provide outdated information.

The core part of the thesis is bioinformatics analysis of differentially expressed genes in barley and wheat. In barley, the work was focused on analysis of plants overexpressing cytokinin dehydrogenase 1 (CKX1). In wheat, the work aimed comparison of accessions with long and shorter but more adventitious roots. Filip Zavadil Kokáš have developed pipeline for functional annotation which withstand comparison and even outperform well established tools as Blast2GO. While differential expression analysis in both species reveal long list of affected genes, more detailed look could offer interesting information. For example, in wheat genotypes with short-root architecture several PIN and PILS family proteins are up-regulated compared to long-rooted genotypes. These proteins encode transporters of auxin, which in turn plays role in development of roots.

To the submitted work, I have following comments and questions for discussion:

- 1) There are some minor mistakes in figures and tables in the otherwise well-written work that make it more difficult to understand and reader have to find correct information elsewhere.



E.g. at page 45: Was stress applied for 24 hours (as in figure) or for four days (as in text) in case of plants grown in soil? Tables 16 and 18 obviously show up- and down-regulated GO-terms in the same analysis. However, their labels differ. Moreover, there is incorrectly stated in label of table 18 that it reports up-regulated genes.

- 2) On page 18 states that error rate for Oxford Nanopore sequencing could exceed 90%. This is a bit outdated information. What is the current status of Oxford Nanopore's sequencing quality?
- 3) In the literature overview (p. 21-22) states that different cDNA/RNA fragmentation methods cause different bias. I am eager to know more about it. Could you please explain bias produced by particular methods more in detail?
- 4) While literature overview presents comprehensive range of sequencing platforms and techniques used in RNA/transcriptome analysis, it misses one of the state-of-the-art approaches - Iso-Seq provided by PacBio. For which applications (based on your experience) it is beneficial to use this approach? What are its strengths and weaknesses compared to classical Illumina-based RNA-Seq? Does Iso-Seq data exist for any plant used in your work, barley or wheat?
- 5) There was lower proportion of mapped reads from root tissue than aerial part of plant in experiments with barley (presented in table 2). Do you have any idea what is a reason for such result of experiment? Did you check unmapped reads (its quality and composition) in particular experiment?
- 6) Portion of mapped reads is also significantly reduced depending on removal of "redundant" transcripts during assembly process in experiment with wheat. In contrast to increase in uniquely mapping reads, there is reduction of all mapped reads by 20% during "all-merged-genotypes" assembly. Does it reflect decreasing complexity of assembly in terms of isoforms and splicing variants? Does increase of uniquely mapped reads overweight loss of overall quantity of mapped reads, especially taking in account that methods for assignment of multi-mapping reads in RNA-Seq experiments exist (e.g. in RSEM)? Could this reduction influence results of differential expression analysis?

I conclude that Filip Zavadil Kokáš have generated original and interesting results during his study. To my opinion, his work meets all requirements on Ph.D. thesis. I recommend accepting his dissertation for defense.

May 29, 2019, Olomouc

Jan Bartoš, Ph.D.