Expression regulation of desiccation-resistance genes in Polypedilum vanderplanki

Pavel Mazin
Skolkovo Institute of Science and Technology, Skolkovo 143025 Russian Federation, iaa.aka@gmail.com

Polypedilum vanderplanki is a striking example of an insect that could survive almost complete water loss. It's genome and series of dehydration-rehydration and development transcriptomes, together with genome of *Polypedilum nubifer* (congeneric desiccation-sensitive midge) were recently released [1]. We used our recently developed package SAJR [2] to predict genes, perform differential gene expression (for statistical tests edgeR [3] was used) and alternative splicing analysis. Results shows that number of differentially (compared with control condition) expressed genes (DEGs) rises gradually during desiccation and drops back under re-hydration *P. vanderplanki*. Interestingly, most of these genes (80% of DEGs after 48 hours of desiccation) are suppressed under desiccation and proportion of protein-coding genes is significantly higher among suppressed genes than among activated (Fisher test, $pv<10^{-24}$, odd ratio > 1.8). Dessication-related gene expression changes exhibit noticeable correlation with gene expression changes occurred under larvae-to-pupae transition (log fold change Pearson correlation is 0.294). Most of this correlation is explained by housekeeping genes that are suppressed under both processes.

Using TSS positions, identified in this work, we have shown that TCTAGAA DNA motif, closely resembled binding site of *D. melanogaster* heat shock transcription activator (HSTF), is significantly enriched in promoter regions of desiccation-induced genes, such as LEA, thioredoxins or trehalose metabolism genes in *P. vanderplanki* but not in *P. nubifer*. Unlike *P. nubifer*, *P. vanderplanki* exhibit doubled TCTAGAA motif in upstream of HSTF, that is likely explanation of much stronger activation of HSTF in *P. vanderplanki* compared to *P. nubifer* under desiccation (3.7 and 1.3 times respectively).

We used *de novo* transcript prediction that allowed as to identify hundreds of new genes, show that up to 53% of genes undergo alternative splicing (AS). These genes are significantly frequently than expected by chance encode kinases, nucleotide binding proteins, components of cytoskeleton and proteins involved in translation. Differential splicing analysis shows that

up to 338 of genes (510 AS events) change it's splicing significantly (GLM, log-likelihood test, BH-correction, qv < 0.05, dPSI (Percent Spliced In change) > 0.1) during desiccation-rehydration cycle. 185 of these events are retained introns, their dPSI after 48 hours of desiccation exhibit low, but significant negative correlation with log fold change of corresponding genes (Pearson correlation coefficient is 0.3, pv < 0.0046). It points to possible role of intron retention in gene expression regulation (presumably through NMD pathway) during desiccation. Interestingly, alternative first exons represent one of popular type of differential isoform usage in *P. vanderplanki* but not in *P. nubifer*. 13 and 1 genes in *P. vanderplanki* and *P. nubifer* respectively exhibit significant shift in first exon preferences after 24 hours of desiccation. In *P. vanderplanki* these genes include ornithine decarboxylase inhibitor, phosphate transporter, several kinases and other genes. One of striking examples of desiccation-induced AS-change is fifth exon of HSTF. This exon encode part of transactivation domain [4] and it's inclusion drops dramatically from 80% to almost zero under dessication in both species, likely altering regulatory properties of encoded protein.

Taking together, our integrative analysis allowed us to greatly extend existing gene annotation and show that AS is widespread in the both species and play a role in desiccation resistance. We have shown dramatic differences in HSTF-regulatory system between *P. vanderplanki* and *P. nubifer.* Our results suggests important role of HTSF in desiccation-induced gene expression activation in *P. vanderplanki* but not in *P. nubifer.*

This is joint work with Mikhail Gelfand and Oleg Gusev.

1. O.Gusev, et al. (2014) Comparative genome sequencing reveals genomic signature of extreme desiccation tolerance in the anhydrobiotic midge, *Nature Communications, **5**:4784*.

2. P. Mazin, et al. (2013) Widespread splicing changes in human brain development and aging, *Mol. Syst. Biol.*, **9**:633

3. M.D.Robinson, et al. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, *Bioinformatics*, **26(1)**:139-40

4. J. Wisniewski, et al (1996) The C-terminal region of Drosophila heat shock factor (HSF) contains a constitutively functional transactivation domain, *Nucleic acids research*, **24(2)**:367–374