DOI 10.18699/SBB-2022-045

New insights on genome organization, gene expression regulation and evolution of the anhydrobiotic midge *Polypedilum vanderplanki*

Shagimardanova E.[1]*, Kozlova O.[1], Nesmelov A.[1], Cherkasov A.[1], Deviatiiarov R.[1], Shaikhutdinov N.[1], Gazizova G.[1], Ryabova A.V.[1], Cornette R.[2], Kikawada T.[2], Gusev O.[1, 3]

[1] *Extreme biology laboratory, Regulatory Genomics Research Center, Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia*

[2] *Institute of Agrobiological Sciences, National Agriculture and Food Research Organization (NARO), Tsukuba, Ibaraki, Japan*

[3] *Department of Regulatory Transcriptomics for Medical Genetic Diagnostics, Graduate School of Medicine, Juntendo University, Tokyo, Japan*

** ryukula@gmail.com*

Key words: anhydrobiosis, *Polypedilum vanderplanki*, stress response, genome assembly, gene expression

Motivation and Aim: Extreme environmental conditions, including physical (temperature, high radiation), chemical (salinity, toxins), and biological (infection agents) factors, are the major engine of evolutionary changes. Organisms adapted to tolerate conditions in which others would not survive are referred to as extremophiles. A remarkable example of an extremophile is the sleeping chironomid midge *Polypedilum vanderplanki*. The midge tolerates complete desiccation by entering an ametabolic state, anhydrobiosis. The phenomenon of anhydrobiosis in *P. vanderplanki* was first discovered in 1951 by Howard Hinton [1]. Further studies in the early 2000s revealed basic biochemical mechanisms underlying tolerance to complete desiccation [2–4]. The whole genome of *P.vanderplaki* was decoded in 2014 [5]. The comparative genome study with sister species *P.nubifer* sensitive to desiccation showed that the genome of *P. vanderplanki* contains multicopy clusters of genes contributing to the formation of molecular shield. Such clusters include genes encoding antioxidants, chaperone-like Lea proteins, DNA and protein repair molecules, trehalose synthase and trehalose transporter. *P.vanderplanki* and recently discovered *P. pembae* are the only known species of chironomids capable of inducing anhydrobiosis, suggesting the recent origin and rapid adaptive evolution of this extreme adaptation in insects. *Methods and Algorithms*: Libraries of Illumina short pair-end reads, PacBio long reads, and Hi-C reads were used for genome assembly RNA-seq, Cage-seq, and CTR-seq for gene prediction and differential gene expression, as well as for promoter identification. Metabolites were measured using flight mass spectrometry (CE-TOFMS; Agilent Technologies). Single cell libraries were obtained using the Chromium platform (10xGenomics). For intraspecies comparative study genomes of other organisms were obtained from ENSEMBL, NCBI, and other databases. *Results*: To further investigate genetic mechanisms of adaptations to complete desiccation, we obtained a high-quality chromosome-level assembly of the *P.vanderplanki* genome. RNA-seq and CTR-seq data allowed the annotation of additional genes so that their number increased to 18990 compared to 17852 in the previous genome assembly. Anhydrobiosis-associated genes were found to preferably be localized in Chromosome IV, and most of the genes localized on this chromosome are highly specific for *P. vanderplanki* and do not have orthologs in other Dipteran insects [6]. Such genes have been extensively multiplicated, forming specific loci, anhydrobiosis-related islands (ARIDs). One of the most extensively studied ARIDs is a

group of 27 Lea genes that encode intrinsically disordered proteins. It has been suggested that the evolution of Lea went toward subfunctionalization: The paralogs differ in both expression patterns in response to stress and cellular localization [7]. The Lea Island-Located (LIL) orphan genes form another important ARID. 14 Lil protein-coding genes were shown to be upregulated during desiccation. Lils do not possess any known protein domains, so their function is not fully clear. Lils are localized in the membrane, which may suggest their role in the transport of molecules during water loss [8]. Both the Lea and Lil genes are highly up-regulated during desiccation. We showed that the transcription of these genes is regulated by the heat shock factor (Hsf). Hsf controls many other protective genes related to anhydrobiosis, as they harbor Hsf binding sites in their promoter regions. Thus, Hsf is the main regulator for controlling and coordinating drastic transcriptional changes during anhydrobiosis [9]. The main hypothesis of multiple gene duplications in the *P. vanderplanki* genome is the specialization of the paralogs. To verify this hypothesis, we performed a CAGE-seq transcriptome analysis of several tissues of larvae during anhydrobiosis. This method allows to estimate both gene expression and reveal promoter switches suggesting gene regulation mechanisms. We found that some protective genes, including Lea, are expressed in a tissue-specific manner under desiccation. Most of these genes are located on Chromosome IV. Chromosome IV has a lower GC ratio and highly biased nucleotide composition and accumulation of variants compared to other chromosomes, suggesting an increase in the genetic diversity of Chromosome IV, possibly caused by relaxed selective pressure. Furthermore, the synteny block assay detected the absence of collinear blocks in Chromosome IV with any other species of dipteran insect except *P. pembae.* On the other hand, genes located in Chromosome IV of *P. vanderplanki* have some orthologous genes in other dipteran genomes. Interestingly, seven out of nine ARIDs are located in Chromosome IV. We further performed metabolome profiling of the larvae during the dehydration-rehydration cycle to find that, in addition to the well-known protective role, trehalose acts as a major source of energy for the initial steps of rehydration in the larvae. Citrate and adenosine monophosphate (AMP), accumulated in the dry state, allow rapid resumption of metabolism during the recovery phase [10]. Finally, we ask if all protective molecules provide protection to all cell populations in the larvae body or if there are specific cell types that survive and then proliferate during recovery. We conducted a pilot single-cell gene expression profiling of the brain of the larvae during the desiccation-rehydration cycle. Preliminary data showed strong fluctuation associated with anhydrobiosis of certain cell types. We hypothesized that some highly differentiated neural cells do not survive water loss, but stem cells proliferate during rehydration. This suggestion still needs to be validated by further studies. *Conclusions*: Using the multi-omics approach, we made several important steps in the understanding of molecular mechanisms of adaptations to extreme desiccation. Chromosome-level genome assembly allowed us to obtain evidence of the specific role of a single chromosome in the evolution of anhydrobiosis

Acknowledgements: This study was supported by Russian Science Foundation Grant No. 20-44-07002.

References

1. Hinton H.E. *Proc Zool Soc Lond.* 1951:371-380
2. Watanabe M. et al. *J Exp Biol*. 2002;205:2799-2802.
3. Kikawada T. et al. *Integr Comp Biol*. 2005;45(5):710-714.
4. Sakurai M. et al. *Proc Natl Acad Sci U S A*. 2008;105(13):5093-5098.
5. Gusev O. et al. *Nat Commun*. 2014;12(5):4784.
6. Yoshida Y. et al. *NAR Genom Bioinform*. 2022;4(2):lqac029.
7. Kondratyeva S.A. et al. *Biology*. 2022;11(4):487.
8. Voronina T.A. et al. *Sci Rep*. 2020;10(1):11633.
9. Mazin P.V. et al. *Proc Natl Acad Sci U S A*. 2018;115:2477-2486.
10. Ryabova A. et al. *Proc Natl Acad Sci U S A*. 2020;117:19209-19220.