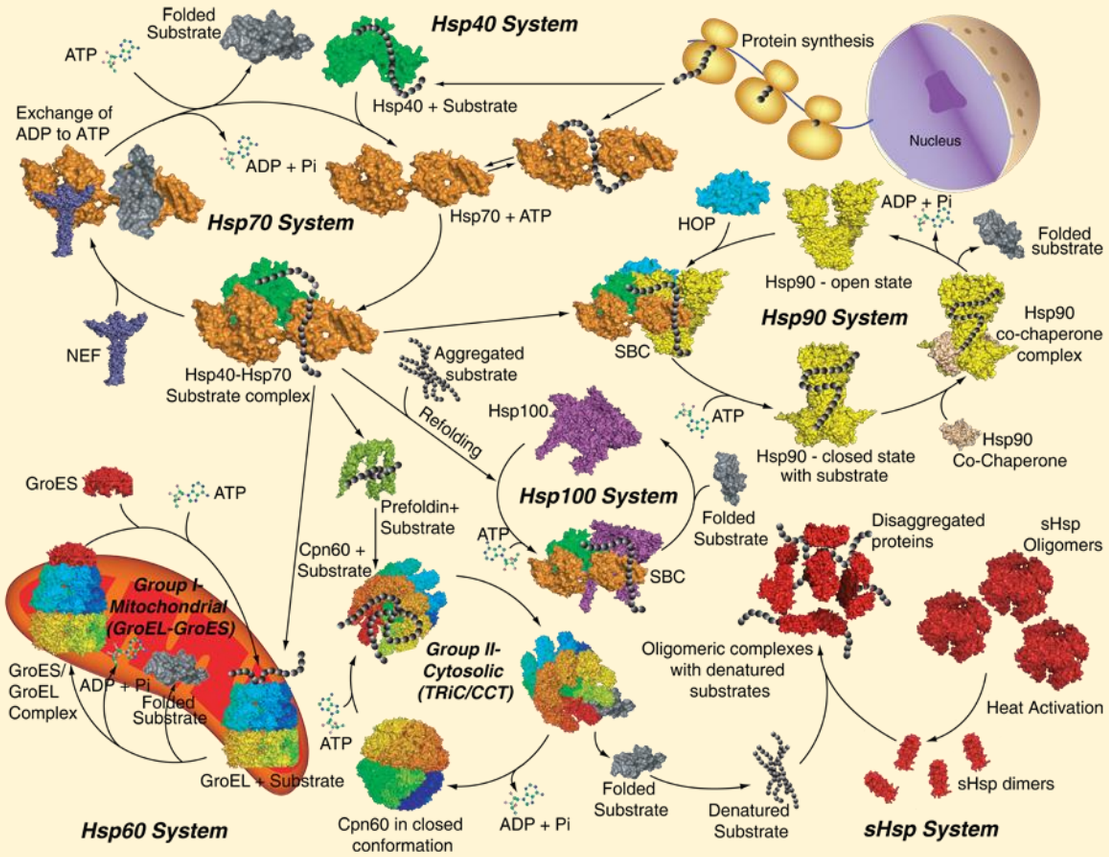
Comparative genomics of heat shock proteins system in extremophile nonbiting midges

Olga Kozlova, Guzel Gazizova, Elena Shagimardanova, Oleg Gusev

Kazan Federal University

HSP – universal chaperone system

Most abiotic stresses – such as elevated temperature, desiccation or chemical stress – lead to denaturation and aggregation of wrongly folded proteins. In order to protect their proteins from aggregation and non-native folding, all living systems use wide range of molecular chaperones, including heat shock proteins (HSPs) which present the largest and the most evolutionary conservative class of such chaperones.

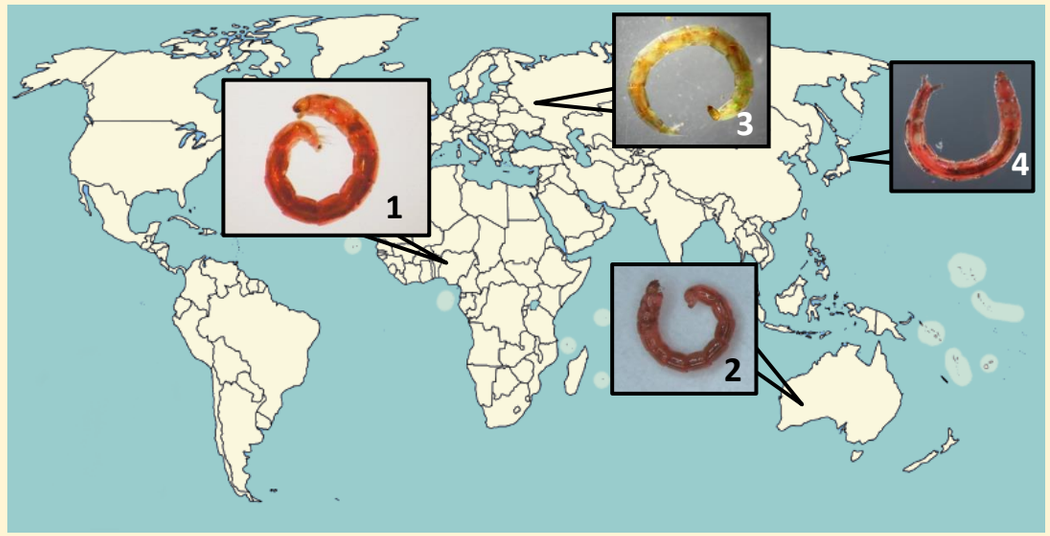


Taken from HSPIR: http://pdslab.biochem.iisc.ernet.in/hspir/

The regulation of HSPs-coding genes in eukaryotes is performed by specific family of transcription factors – heat shock factors (HSFs), which bind to special regulatory sequence – heat shock element (HSE) in the promoter region of a target gene.

Chironomids and adaptations

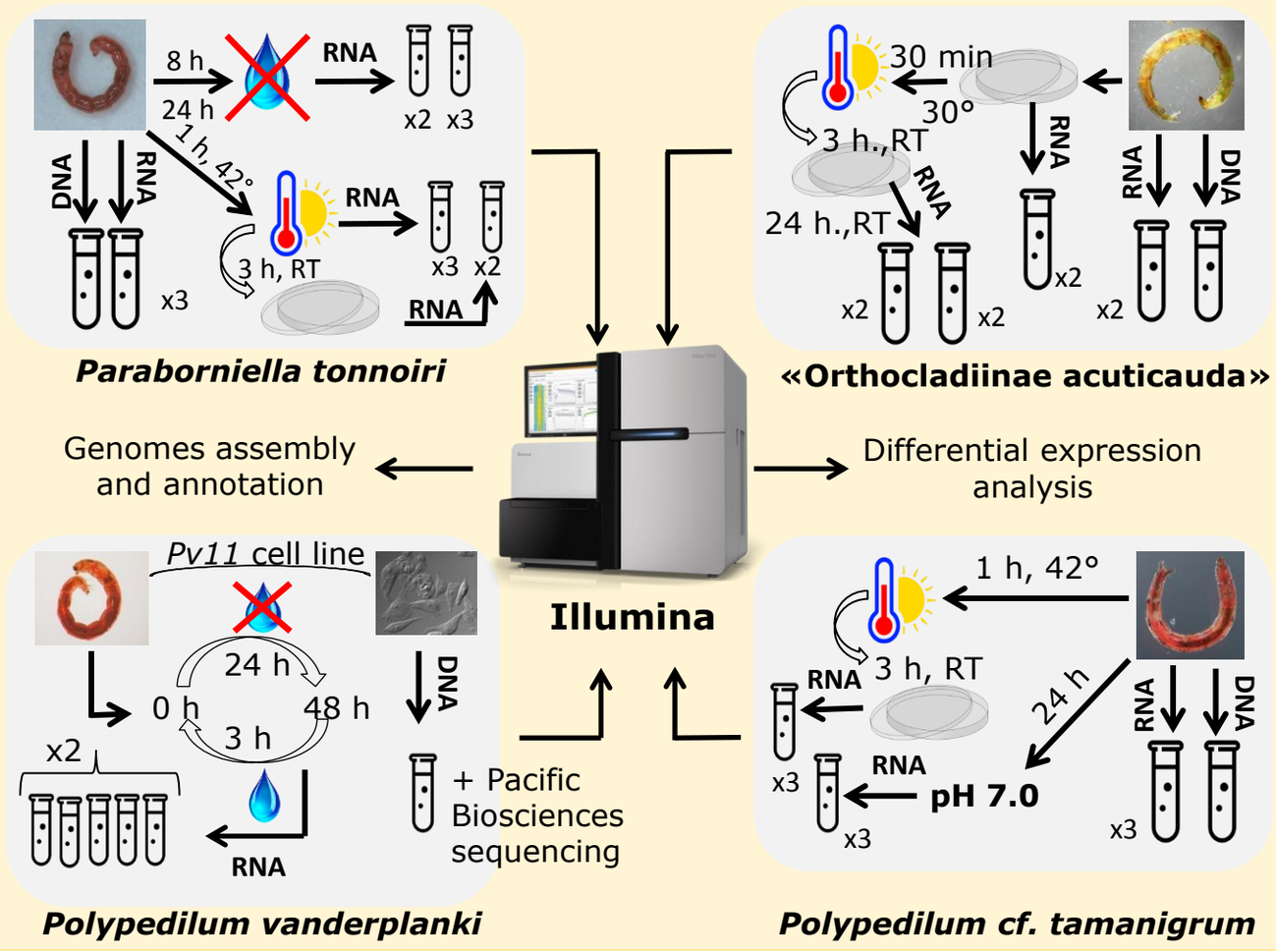
It's well known that HSPs take part in the reaction towards wide range of abiotic stresses – not only to heat shock itself. Taken this into account, it could be interesting to compare features of HSPs-coding genes in species, which are remote enough from each other to the evolutionary and geographic point of view, but united by habitation in ecological niches, linked to different extreme impacts. To this point of view the **Chironomidae (non-biting midges)** is considered to be one of the most perspective family of Diptera, because many of its species during their larval stage of development face need to vitality preservation in different extreme and unstable conditions, for what they use specific behavioral, physiologic and molecular-genetic adaptations. Here we study comparative genomics of HSP-related system in larvae of four extremophile insects from Chironomidae family.



1. *Polypedilum vanderplanki* – anhydrobiotic midge from Africa (desiccation-tolerant)
2. *Paraborniella tonnoiri* – desiccation-resistant Australian midge
3. «Orthocladiinae acuticauda» - psammorheophilous midge, heat shock-resistant (Russia)
4. *Polypedilum cf. tamanigrum* – acid-tolerant midge from Japan

Methods

In the given research three draft genome assemblies of extremophile Chironomidae species were performed. In order to investigate changes in expression levels of HSPs-coding genes we also sequenced RNA of larvae in different stress conditions – heat shock and desiccation for Australian larvae of *Paraborniella tonnoiri*, heat shock for Russian larvae of «Orthocladiinae acuticauda» and heat shock and ion stress (exposition in fresh water) for acid-tolerant larvae of *Polypedilum cf. tamanigrum* from Japan. We also performed brand new variant of genome assembly for unique anhydrobiotic African species *Polypedilum vanderplanki* by sequencing DNA of Pv11 cell line, derived from embryonic mass of the insect. For differential expression analysis of genes of *P. vanderplanki* we used earlier sequenced libraries, which reflect desiccation-rehydration cycle of its larvae. For orthogroups searching we used all predicted aminoacid sequences of species under consideration as well as proteins for 6 other Diptera.



Genomes assembly and annotation

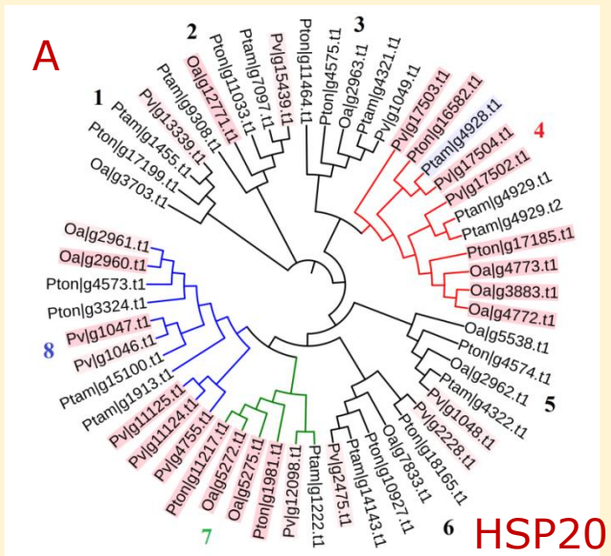
	P. vanderplanki	*P. tonnoiri*	«O. acuticauda»	*P. cf. tamanigrum*
Genome size, Mb	120.4	**177.4**	**100.6**	105.5
% of genome size, covered by long (≥10 000 bp) scaffolds	99.7	61.5	79.7	86.1
N50 (scaffolds ≥ 500 bp), bp	1001272	15447	40474	76058
BUSCO (Diptera), %	C: 96.8; M: 1.5	C: 91.6; M: 2.9	C: 91.7; M: 2.9	C: 95.1: M: 1.7
% of repeats	11.54	**33.37**	8.94	**4.96**
Total number of «genes»	17 993	18 583	**15 112**	**18 864**
Total number of «transcripts»	19 279	19 849	**16 304**	**20 037**
Mean intron length, bp	545	**771**	524	**487**
GC-content, %	28.1	29.9	31.9	40.1

- **The variance** of genome sizes is achieved by **changes** of **repeats percentage** and **introns length**
- *P. cf. tamanigrum* has the most **«compact»** genome

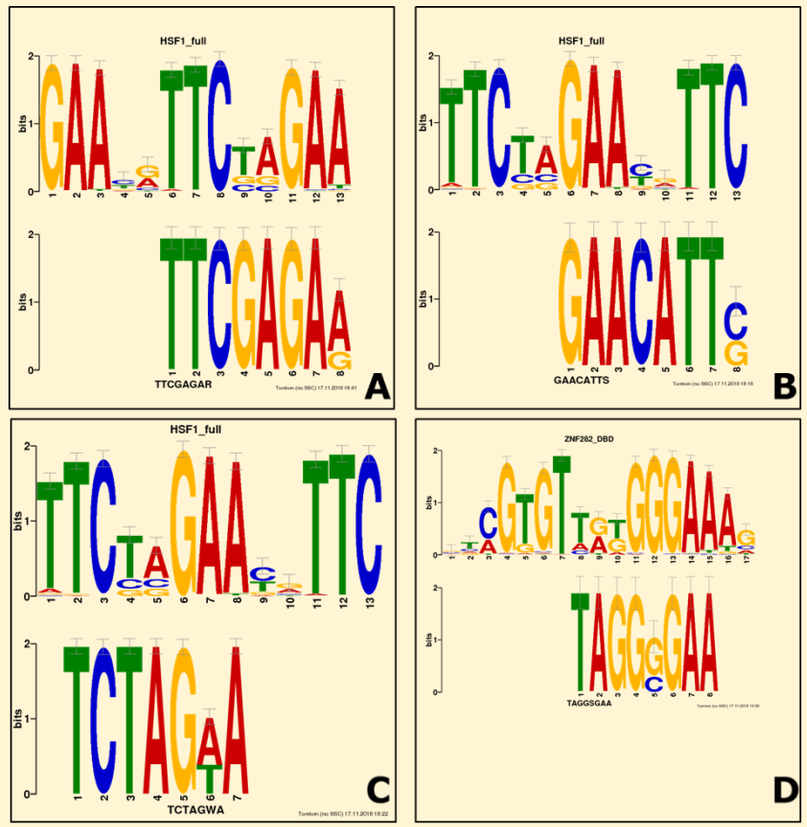
Heat Shock Proteins

In contrast to high-molecular HSPs (B, C, D), the number of genes coding for small HSPs, differs between species: from 10 copies for *P. cf. tamanigrum* till 15 copies for *P. vanderplanki* (A). It was shown that each extremophile species (except for *P. cf. tamanigrum*) has its own specific cluster with expansion of gene copies (blue and green clusters; red cluster includes universally up-regulated genes), and exactly these species-specific genes show sharp increase of expression after stress (A). This allows suggesting that the adaptation to extreme environments by using small HSPs walked the path of gene expansion as well as regulation. Among all HSPs-coding genes those corresponding to low-molecular HSP and HSP70 families showed the most dramatic and universal up-regulation of expression in response to abiotic stresses. In contrast to other extremophile species, in case of desiccation-tolerant midge *P. vanderplanki* we noted up-regulation of almost each HSP-coding gene (except for some low-molecular HSPs). Chaperonins (HSP60 family) were specifically up-regulated in case of desiccation, but not after heat shock. But the most surprising observation was linked to acid-tolerant species *P. cf. tamanigrum* which showed significant up-regulation of HSPs-coding genes neither after heat shock nor after ion stress. In order to find a possible reason of this notion we performed motif enrichment analysis in supposed promoter regions of HSPs-coding genes, taking 500 random genes of each genome as a control group. As a result we found statistically significant enrichment of motifs, similar to Drosophila's HSE for genomes of «Orthocladiinae acuticauda», *Paraborniella tonnoiri* and *Polypedilum vanderplanki*, but no enrichment for *Polypedilum cf. tamanigrum*. Thus we could make a conclusion that all HSP system is not involved in reaction to heat shock in *P. cf. tamanigrum*, though for the present it is not clear, how such conservative mechanism of regulation could have been broken.

HSPs cladograms



Motif enrichment analysis



Enriched motifs in promoter regions of HSP genes (A – «O. acuticauda», B – *P. tonnoiri*, C – *P. vanderplanki*, D – *P. cf. tamanigrum*) (up) compared to known motifs (down)