Molecular omics resources and tools for amphipod investigation

Yannick Cogne, Duarte Gouveia, Olivier Pible, Arnaud Chaumot, Davide Degli Esposti, Olivier Geffard, Christine Almunia, Jean Armengaud

HAL Id: hal-04488093
https://hal.inrae.fr/hal-04488093

Submitted on 4 Mar 2024

Molecular omics resources and tools for amphipod investigation

Yannick Cogne[1], Duarte Gouveia[1], Olivier Pible[1], Arnaud Chaumot[2], Davide Degli-Esposti[2], Olivier Geffard[2], Christine Almunia[1], **Jean armengaud[1]**

1. CEA-Marcoule, DRF/Joliot/DTMS/SPI/Li2D, Laboratory "Innovative technologies for Detection and Diagnostics", BP 17171, F-30200 Bagnols-sur-Cèze, France.
2. Irstea, UR RIVERLY, Ecotoxicology Group, 5 rue de la Doua, CS 20244, F-69625, Villeurbanne Cedex, France

Corresponding author: Jean.armengaud@cea.fr

Objectives. Gammarids are key animal sentinels for *in situ* ecotoxicological biomonitoring of fresh water. Molecular biomarkers representative of key physiological parameters may be defined for gaining insights into the response of organisms to toxicants and measuring the anthropogenic impact in the environment. Recently, proteogenomics, a novel approach intimately combining next-generation sequencing and proteomic methodologies, has emerged as a straightforward strategy for discovering relevant proteins in non-model organisms. This opens the possibility to analyze the molecular players from any amphipod, and even to investigate its microbiota and parasites.

Methods. We sequenced the transcriptomes of a male and a female for seven different taxonomical groups: *Gammarus fossarum A*, *G. fossarum B*, *G. fossarum C*, *Gammarus wautieri*, *Gammarus pulex*, *Echinogammarus berilloni* and *Echinogammarus marinus*. These taxa were chosen to explore the molecular diversity of transcribed genes of genotyped individuals from these groups. Transcriptomes were *de novo* assembled and annotated. We optimized the *de novo* assembly strategy and constructed an impressive collection of protein sequences for these fourteen gammarids that can be used for interpreting proteomics data. In parallel, we recorded shotgun proteomics data on more than a hundred gammarid individuals to explore several key questions. We also developed several pipelines to investigate their proteogenomes and their microbiota.

Results. For example, we analysed two regional *Gammarus pulex* populations to characterize the potential proteome divergence induced in one site by natural bioavailable Cadmium contamination compared to a non-contaminated site. We have shown that the intra-population proteome variability of long-term exposed *G. pulex* was inflated relatively to the non-contaminated population. While remaining a challenge for such organisms with not yet sequenced genomes, taking into account intra-population variability is important to better define the molecular players induced by toxic stress in a comparative field proteomics approach.

Conclusion. The fourteen RNA-seq derived protein sequence databases proposed here are an important resource for proteogenomics on these non-model organisms. This work illustrates the relevance of omics for development of multiplexed biomarkers. Interestingly, the tools and strategies developed in this project are transposable to any amphipod.