Implementation of an interface for exploring molecular diversity by proteogenomics

Yannick Cogne¹, Christine Almunia¹, Olivier Pible¹, Duarte Gouveia¹,², Arnaud Chaumot², Olivier Geffard², Jean Armengaud¹

¹CEA-Marocule, DRF/Joliot-DMTS/SP/Il2D, Laboratory Innovative technologies for Detection and Diagnostics, BP 17171, Bagnols-sur-Cèze
²Irstea, Unité de Recherche MALY, Laboratoire d’écotoxicologie, CS70077, F-69626 Villeurbanne

Abstract
Defining molecular markers of any sentinel organism is an important goal in ecotoxicology. With the long-term goal of testing the biological quality of the water of a river, we developed a test based upon a widespread sentinel organism, namely Gammarus fossarum, a small amphipod. Biomarkers of interest from this organism, which are proteins whose abundance vary depending on the presence of pollutants, can be monitored by proteomics. G. fossarum is a so-called “non-model” species because there is no reference genomic sequence for this genus. A reference protein sequence database can be obtained by de novo assembly of mature transcript sequences with the help of proteogenomics data. To take into account the diversity of Gammarus populations, a combination of bioinformatics tools for assembling RNASeq data and proteomic assignment is required. We choose the Galaxy working environment with an easy portable solution using Docker, a container management program runable under Windows. Galaxy Docker (developed by Björn Grüning) with one container per tool provides version management as well as a high reproducibility of environment variables. The implementation for evaluating strategies for exploring molecular diversity through proteogenomics, including the management of data from new RNA sequencing technologies and their assembly will be detailed.

Exploring molecular diversity of Gammarids

<table>
<thead>
<tr>
<th>5 gammarid species</th>
<th>20 animals (population)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 female</td>
<td>10 male</td>
</tr>
<tr>
<td>10 female</td>
<td></td>
</tr>
</tbody>
</table>

Comparative proteogenomics
Selection of conserved peptides (= core-peptides)

Proteomics
6-frame translation

BIOMARKERS

Docker galaxy, a friendly interface user and OS independant

References
Björn Grüning: https://github.com/bjorgn/galaxy-docker

Acknowledgments
We thank the Commissariat for Atomic Energy and Alternative Energies, IRSTEA, the French national agency ANR (ANR-14- CE12-0006; Proteomics for novel biomarkers relevant in ecotoxicology : challenging the biodiversity and multiplexed immunoassay as diagnostic tool) for their financial support, as well as Guylaine Miotello, Gérard Steinmetz, and Jean-Charles Galliard for their technical support.