Annotating arthropods genome to study and compare their metabolism: the ArthropodaCyc collection of Cyc databases powered by CycADS

Patrice Baa-Puyoulet, Augusto Vellozo, Jaime Huerta-Cepas, Gérard Febvay, Federica Calevro, Marie-France Sagot, Hubert Charles, Toni Gabaldon, Stefano Colella

HAL Id: hal-02798101

https://hal.inrae.fr/hal-02798101

Submitted on 5 Jun 2020

Annotating arthropods genome to study and compare their metabolism: the ArthropodaCyc collection of Cyc databases powered by CycADS

Patrice Baa-Puyoulet[1], Augusto F. Vellozo[2,3], Jaime Huerta-Cepas[4], Gérard Febvay[1], Federica Calevro[1], Marie-France Sagot[2,3], Hubert Charles[1,3], Toni Gabaldon[5] and Stefano Colella[1,3]

[1] UMR203 BF2I, Biologie Fonctionnelle Insectes et Interaction, INRA, INSA-Lyon, Université de Lyon, F-69621 Villeurbanne, France
```
{Patrice.Baa-Puyoulet, Gerard.Febvay, Yvan.Rahbe, Stefano.Colella}@lyon.inra.fr;
             {Federica Calevro, Hubert.Charles}@insa-lyon.fr
```
[2] Université Lyon 1, CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, Villeurbanne, France
```
                  avellozo@gmail.com; Marie-France.Sagot@inria.fr
```
[3] BAMBOO, INRIA Rhône-Alpes, France
[5] Structural and Computational Biology Unit, EMBL Heidelberg, Heidelberg, Germany
```
                  huerta@embl.de
```
[5] Centre for Genomic Regulation, Barcelona, Spain
```
                  Toni.Gabaldon@crg.es
```

Several arthopods genomes have been and are being sequenced (e.g. the i5K sequencing initiative [1]) and these data open the way to comparative studies of different species to better understand their biology. All comparative genomics studies rely heavily on the quality of genome annotations. To be useful to researchers genomics data have to be collected from various sources, updated regularly and organized in dedicated databases. During the genome annotation for the pea aphid (*Acyrthosiphon pisum*) we developed CycADS [2] (Cyc Annotation Database System), an automated annotation management system that allows the seamless integration of the latest sequence information into metabolic networks reconstruction. Specific genomic data, as well as their functional annotations obtained using different methods (such as KAAS, PRIAM, PhylomeDB, Blast2GO, MetaPhOrs, Interproscan), are collected into a SQL database and later extracted, with the possibility to apply different quality filters. CycADS allows the automatic generation of a complete set of input files to build or update BioCyc databases [3] using the Pathway Tools software. We used CycADS to create "ArthropodaCyc " [4], a collection of arthropods metabolic network Cyc databases, which contains 23 organisms to date. Our collection of databases allows metabolic pathways visualisation, and each protein page includes information about the annotation methods used, as well as hyperlinks to genome specific resources. Comparison using interactive web functionalities, as well as user "omics" data mapping or information extraction, are also available in the BioCyc interface of ArthopodaCyc. Future plans include the addition of other sequenced genomes after their publication and/or in collaborations with arthropods genome sequencing projects. Further developments will also include the implementation of gateways to network analysis tools.

[1] http://arthropodgenomes.org/wiki/i5K; [2] http://www.cycadsys.org/;
[3] http://biocyc.org/; [4] http://arthropodacyc.cycadsys.org/

Keywords: genome annotation; metabolic networks; arthropod genomes; CycADS; BioCyc; ArthopodaCyc