



HAL
open science

From SymbAphidBase to SymbAphidCyc: two companion databases to study and compare aphid symbionts from genomes to metabolic pathways

Patrice Baa-Puyoulet, Mathieu Labernardiere, Nicolas Parisot, Jean-Pierre Gauthier, Gérard Febvay, Federica Calevro, Yvan Rahbé, Hubert Charles, Jean-Christophe Simon, Stefano Colella

► To cite this version:

Patrice Baa-Puyoulet, Mathieu Labernardiere, Nicolas Parisot, Jean-Pierre Gauthier, Gérard Febvay, et al.. From SymbAphidBase to SymbAphidCyc: two companion databases to study and compare aphid symbionts from genomes to metabolic pathways. JOBIM 2015 - Journées Ouvertes Biologie Informatique Mathématiques, Université d'Auvergne - Clermont-Ferrand I (UdA). FRA., Jul 2015, Clermont-Ferrand, France. 1 p. hal-01208786

HAL Id: hal-01208786

<https://hal.science/hal-01208786v1>

Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

From SymbAphidBase to SymbAphidCyc: two companion databases to study and compare aphid symbionts from genomes to metabolic pathways

Patrice BAA-PUYOULET¹, Mathieu LABERNARDIERE¹, Nicolas PARISOT¹, Jean-Pierre GAUTHIER², Gérard FEBVAY¹, Federica CALEVRO¹, Yvan RAHBE¹, Hubert CHARLES¹, Jean-Christophe SIMON² and Stefano COLELLA¹

¹ UMR203 BIOLOGIE FONCTIONNELLE INSECTES ET INTERACTIONS (BF2I), INSA LYON, 20 ave Albert Einstein, bât Pasteur, 69621, Villeurbanne, Cedex, France

² UMR 1349 INSTITUTE OF GENETICS ENVIRONMENT AND PLANT PROTECTION (IGEPP), INRA, Domaine de la Motte, BP 35327, 35653, Le Rheu, Cedex, France

Auteur à contacter : Patrice.Baa-Puyoulet@lyon.inra.fr

Résumé

L'évolution des techniques de séquençage donne des opportunités sans précédent pour l'étude des bactéries symbiotiques et comprendre leurs contributions à la physiologie de leur hôte grâce à la génomique comparative et l'analyse des réseaux métaboliques. Les pucerons sont un modèle établi, ils abritent un endosymbiote obligatoire primaire, *Buchnera aphidicola*, et plusieurs symbiotes secondaires facultatifs contribuant à la physiologie et à l'adaptation de l'insecte. Récemment, nous avons développé SymbAphidBase (<http://symbaphidbase.cycadsys.org/>) : une base de données ad hoc pour stocker et analyser les séquences génomiques des symbiotes de pucerons. Actuellement, SymbAphidBase comprend les génomes des 17 souches de *B. aphidicola* de 8 espèces de pucerons disponibles dans GenBank (voir le poster ECCB 2014 <http://f1000.com/posters/browse/summary/1096898>). Dans une deuxième phase, nous avons utilisé CycADS (<http://www.cycadsys.org/>), un système automatisé de gestion d'annotations que nous avons développé pour générer une annotation fonctionnelle enrichie orientée vers la reconstruction métabolique (pour faciliter son utilisation, le pipeline CycADS a récemment été intégré dans la plate-forme web Galaxy) : Les données génomiques ainsi que les annotations fonctionnelles obtenues en utilisant différentes méthodes (telles que KAAS, PRIAM, Blast2GO, InterProScan) ont été collectées pour créer SymbAphidCyc (<http://symbaphidcyc.cycadsys.org/>), la collection de bases de données métaboliques dédiée aux symbiotes du puceron. En conclusion, les 2 bases de données complémentaires, SymbAphidBase et SymbAphidCyc, facilitent le stockage et la comparaison des données génomiques ainsi que l'analyse du métabolisme pour une meilleure compréhension de la symbiose chez les pucerons.

Summary

The next-generation sequencing methods evolution opened the way to unprecedented opportunities to study symbiotic bacteria through genomic comparisons and metabolic networks analyses to understand their contributions to their host biology. Aphids are an established model: they harbour an obligate primary endosymbiont, *Buchnera aphidicola*, and several facultative secondary symbionts contributing to the insect's physiology and adaptation. Recently we developed SymbAphidBase (<http://symbaphidbase.cycadsys.org/>): an ad hoc genome database to store and analyse aphid symbionts' genome sequences. At present SymbAphidBase includes the genomes of 17 strains of *B. aphidicola* from 8 different aphid species available in GenBank (see poster ECCB 2014 <http://f1000.com/posters/browse/summary/1096898>). In a second phase, we used CycADS (<http://www.cycadsys.org/>), an automated annotation management system that we developed during the pea aphid genome project, to generate an enriched functional annotation geared towards metabolic reconstruction. It is worth noting that the CycADS pipeline has been recently integrated into the Galaxy web-based platform to ease its use. The genomic data as well as the functional annotations obtained using different methods (such as KAAS, PRIAM, Blast2GO, InterProScan), were collected to easily build SymbAphidCyc (<http://symbaphidcyc.cycadsys.org/>), the database collection for the aphid symbionts. In conclusion, SymbAphidBase and SymbAphidCyc companion databases facilitate genome data storage and comparisons, as well as metabolism analysis for the better understanding of symbiosis physiology in aphids.