Project No. *222886-2*

MICROME

The Microme Project:
A Knowledge-Based Bioinformatics Framework for Microbial Pathway Genomics

Instrument: **Collaborative project**

Thematic Priority: **KBBE-2007-3-2-08: BIO-INFORMATICS - Microbial genomics and bio-informatics**

D7.2 First MICROME training course

Due date of deliverable: M18
Actual submission date: M21

Start date of project: 1.12.2009 Duration: 48 months

Organisation name of lead contractor for this deliverable: ULB

Project co-funded by the European Commission within the Seventh Framework Programme (2009-2013)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Contributors

Responsible Beneficiary: ULB
Organization of the training sessions: ULB, CEA, SIB, EMBL-EBI.
Additional partners having followed the training course: CERTH, CNIO, WTSI, MN, KO and TAU.

INTRODUCTION

D7.2: First MICROME training course.

1. A general introductory course to present the goals, approaches and components of the MICROME project.
2. Training in metabolism-related bioinformatics with specific sessions for curators and bioinformaticians.
3. Improvement of the teaching material that remains from course to course.

Methods

This first training course was primarily addressed to MICROME participant, in order to stimulate mutual understanding of the bioinformatics approaches, resources and curation tools to be used in the MICROME project. The course also served to prepare training material that will be upgraded in order to open future MICROME training sessions to external researchers.

The training course was organized on June 27-28, 2011 at the CEA (Evry, France) jointly with the first MICROME jamboree (deliverable D7.1). This joint organization facilitated the logistics and transportation, and allowed to directly connect the training to a concrete session of MICROME data curation.

Results (if applicable, interactions with other workpackages)

1. General introductory course

The training course started with general presentations of the resources used to annotate different components of the MICROME database (compounds, reactions, pathways).

1. MICROME annotation meeting June 2011 (Avazeh Ghanbarian, Paul Kersey and Alessandro Vullo, EMBL-EBI).
    - The pre-launch Microme data set was introduced; and set in the context of the user interface that is being developed to make it available to the scientific user base. The user interface reflects the internal structure of the data and preparing data correct and complete for dissemination through this interface (and through associated web services and file dumps) is the goal of the curation process. The presentation explained how the existing Reactome interface fits to bacterial data, and identified priorities for work to finalise the initial data set for launch.
2. Microme-Reactome DB content (Paul Kersey and Alessandro Vullo, EMBL-EBI)
    - A detailed presentation was given about the preparation of the pre-launch Microme data sets; where the data had come from, how data had been reconciled from various sources, how different source data had been fitted to the data model and outstanding problems that remained in reconciling inconsistent sources. Several versions of the pre-launch data set had been prepared for potential use at the meeting, and these were presented and discussed.
3. MICROME data (Anne Morgat, SIB)
    - Anne Morgat summarized the different data types and resources that the project partners are working with. She presented the current status of the curation, in terms of chemicals, biochemical reactions, and the mapping of public metabolic networks onto Microme reference resources (ChEBI for chemical entities and Rhea for reactions) as well as maping to the EcoCyc/MetaCyc pathway resource. This data constitutes the reference data used to build the Microme DB.
4. MicroScope functionalities to support pathways curation (David Vallenet, Eugeni Belda, Damien Mornico, François Le Fèvre and Claudine Médigue, CEA)
    - Claudine Médigue detailed the Microscope web platform (http://www.genoscope.cns.fr/agc/microscope) dedicated to microbial comparative genome analysis and manual functional annotation. Its relational database schema (PkGDB, for «Prokaryotic Genome DataBase») stores computed results of syntactic and functional annotation pipelines as well as Pathway Tools metabolic networks. PkGDB is also linked to these metabolic pathway databases (MicroCyc). The MicroScope web interface (MaGe, for «Magnifying Genomes») has been specifically designed to

assist curators in the evaluation of all available sequence-based (e.g., InterPro domain predictions), context-based (i.e., synteny results and metabolic network predictions) and experimental data (e.g., growth experiments, if any), with the aim of assigning the best possible annotation to a given gene product. Moreover, MicroScope offers many useful data-exploring functionalities, such as allowing users to perform (complex) queries, comparative genomic studies, and metabolic analyses. It also provides summary views of statistics and information for each genome as a whole. Finally, MicroScope can be used both as:

i. a community resource, for comparative analysis and annotation of publicly available genomes

ii. a private resource, as access rights to private projects can be restricted to a limited group of annotators defined by the project leader.

2. Training in metabolism-related bioinformatics with specific sessions for curators and bioinformaticians

The presentations and demos were followed by practical sessions, where trainees received exercises to familiarize themselves with the REACTOME and MICROSCOPE annotation interfaces, as well as the tools for analyzing bacterial regulation (Regulatory Sequence Analysis Tools) and metabolic pathways (Network Analysis Tools).

1. **Demo/Training of Reactome.** Bijay Jassal (EBI) gave a demonstration of the way to use the Reactome interface for entering new pathways in the MICROME database.

- Bijay Jassal (EBI, Reactome group) gave a short presentation on Reactome (a database of human biological pathways and processes), it's rationale and data model and then demonstrated the graphical user interface (called the Curator Tool). Mr Jassal started by making sure all attendees had downloaded the curator tool from Reactome's download page and applied the correct settings to access the Microme database. Mr Jassal then used the tool to create microbial metabolic reactions, adding the required properties of a reaction Microme wish to use and showing the attendees how this new information could be synchronised with Microme's central repository. Once reactions were created, Mr Jassal then showed the attendees how to create reaction diagrams (called Entity Level Views in Reactome) using the tool and once again checking them into Microme's central repository. Overall the tutorial was well followed by all the attendees.
- David Vallenet made a demonstration focused on a curation example: the annotation of the L-carnitine dehydrogenase (PP0302 coding sequence) of *Pseudomonas putida*. This demonstration gave the opportunity to the curators to follow a complete case study from Biolog growth experiments, pathway projection predictions and gene-annotation using the MicroScope platform tools. Finally, a new feature of the Microscope platform dedicated to the curation of the gene-reaction associations was used to annotate the L-carnitine dehydrogenase enzymatic activity in the genome of *P. putida*.
- The training session was made on a preselected set of *Pseudomonas putida* incomplete pathways. This session was focused on pathway holes extracted from MicroCyc metabolic profile tool. Pathway validation was asserted in agreement with Biolog assays provided by DSMZ. Several tools were used to find candidate genes for missing enzymatic reactions, in particular the new phylogenetic profile module integrated in MicroScope (Engelen, et al. 2011).

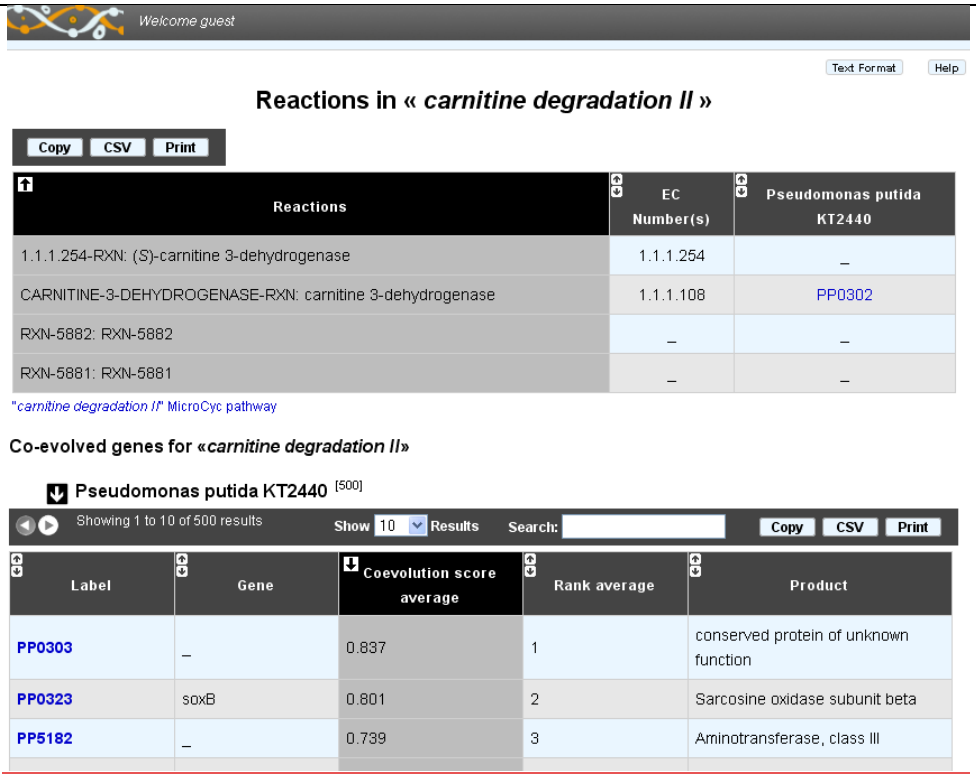

Figure 1: Screenshot of the carnitine degradation pathway completion in MicroScope in regards of the phylogenetic tool results listing the candidate genes for missing enzymatic reactions.

2. **Demo/Training of Regulatory Sequence Analysis Tools and Network Analysis Tools**.

Jacques van Helden (ULB) showed a live demo of selected tools of the Regulatory Sequence Analysis Tools (RSAT, http://rsat.ulb.ac.be/rsat/) and Network Analysis Tools (NeAT, http://rsat.ulb.ac.be/neat/), showing how those tools could be used to automatically predict putative pathways from Bacterial genomes. Those predicted pathways could be considered as hypotheses to be validated by manual curation (analysis of consistency, literature search, etc).

- The RSAT tool ***infer-operons*** predicts the grouping of bacterial genes in operons, using a distance-based method.
- The RSAT tool ***footprint-discovery*** detects phylogenetically conserved cis-regulatory elements, and infers networks of co-regulation.
- ***Pathway discovery.*** Groups of enzyme-coding genes regrouped in operons or regulated by a common transcription factors are likely to participate in a common metabolic pathways. The NeAT tool ***pathway-extraction*** allows predicting metabolic pathways from such sets of "seed" genes, by connecting the reactions catalyzed by their gene products, and inferring the missing steps based on a weighted path finding method (Faust et al., 2011a; Faust and van Helden, 2011b).
- The NeAT tool ***compare-classes*** performs comparisons between sets of objects of any type (e.g. genes, compounds, reactions). The demonstration showed an application of this tool can serve as pathway projection tool (Deliverable D1.4) as well as to compare predicted pathways (resulting from the pathway discovery or from pathway projection) with databases of annotated pathways (Deliverable D1.6).

3. Improvement of the teaching material that remains from course to course.

All the didactic material used for the course (slide presentation and material used for the exercises) is available on the MICROME Web site (http://www.microme.eu/documentation/documents/microme-lectures-d72).

As improved material, Faust and van Helden published a chapter in Methods in Molecular Biology (Faust and van Helden, 2011b) that provides a step-by-step description of the protocol for predicting metabolic pathways from bacterial operons. This chapter will be used as teaching material for future MICROME training sessions. ULB is currently extending this didactic material by preparing a tutorial addressed to bioinformaticians, explaining them the command-line usage of the pathway discovery tools in order to predict metabolic pathways from all the operons and co-regulation groups identified in a bacterial genome.

Perspectives

The next MICROME training session will be organized by ULB partner in the beginning of 2012. It will be addressed to MICROME curators but also open to external participants. This second course will rely on an extended training material, in the form of tutorials and protocols documented with carefully chosen study cases, in order to illustrate the different components of the MICROME annotation pipeline (compounds, reactions, GPR, pathway projection, pathway discovery, genome-scale metabolic models). The tutorials will be formatted according to pre-defined standards and, if possible, will be published in specialized journals (Methods in Molecular Biology, Nature Protocols) or Web sites (Protocol Exchange).

Publications

1. Faust, K., Croes, D. and van Helden, J. (2011a). Prediction of metabolic pathways from genome-scale metabolic networks. Biosystems.
    - This article was published in a special issue of Biosystems, dedicated to the analysis of metabolic network. We review the path finding and subgraph extraction methods developed since 10 years by our group (ULB), and which are now applied in the MICROME project to predict metabolic pathways in Bacterial genomes.
2. Faust, K. and van Helden, J. (2011b). Predicting metabolic pathways by sub-network extraction. Methods in Molecular Biology in press, 15.
    - This chapter will appear in November 2011 in a volume of the series "*Methods in Molecular Biology*" dedicated to "*Bacterial Molecular Networks*" (ed. Jacques van Helden, Ariane Toussaint and Denis Thieffry). The chapter provides a detailed protocol explaining how to use the Web interface of the pathway extraction tool (http://rsat.bigre.ulb.ac.be/neat/) in order to predict metabolic pathways from operons (one of the pathway discovery strategies for the ULB contribution to MICROME). This protocol will serve as didactic material for future MICROME training sessions.
3. Engelen S., Vallenet D., Médigue M. and Danchin Antoine. Distinct co-evolution patterns of genes associated to DNA polymerase III DnaE and PolC (submitted)
4. Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, Caudy M, Garapati P, Gopinath G, Jassal B, Jupe S, Kalatskaya I, Mahajan S, May B, Ndegwa N, Schmidt E, Shamovsky V, Yung C, Birney E, Hermjakob H, D'Eustachio P, Stein L. (2011) Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res. 2011 Jan;39(Database issue):D691-7.