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**

D7.5
Third Microme Annotation Jamboree

Due date of deliverable: February 28, 2013 (M39)
Actual submission date: April 30, 2013 (M41)

Start date of project: 1.12.2009 Duration: 48 months

Organisation name of lead contractor for this deliverable: CEA

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

Lead: CEA
Organization: CEA, WUR, CERTH
Coordination and preparation: CEA, WUR, CSIC
Staff initially involved in the DOW: CEA, WUR, SIB
Additional staff involved in the jamboree annotation: EMBL-EBI, CERTH, CNIO, CSIC, AMB

INTRODUCTION

Deliverable reference number: ***D7.5 Third Microme Annotation Jamboree***

This document summarizes the work done during the third Microme Annotation Jamboree that took place at the Centre for Research and Technology-Hellas (CERTH) in Thessaloniki on the 24th of April 2013.
The objective of this Jamboree was to perform horizontal annotation on the genome of *Pseudomonas putida* KT2440 centred on the reconciliation of growth phenotype data generated by DSMZ and the literature mining with in-silico FBA predictions of *P. putida* KT2440 metabolic model.
We focused on the curation of a set of 47 carbon sources for which experimental data report positive growth for *P. putida* KT2440 but for which metabolic models are not able to predict biomass production. Based on the reconstructed metabolic network of *P. putida* KT2440 from re-annotated genome sequence available in the MicroScope platform and the result of Text Mining strategy to search bibliographical information related to enzymatic activities in *P. putida* associated to the compounds of interest, each participants have curated Gene-Reaction associations (approving/rejecting predicted associations or adding missing reactions) involved in degradative pathways of the query compounds.

ORGANISATION

Venue: Centre for Research and Technology Hellas (CERTH). 6th km Charilaou-Thermi Road, 57001 Thermi, Thessalonica, Greece (http://www.certh.gr/root.en.aspx)

Number of participants: 20 (4 from CEA, 1 SIB, 2 EBI, 2 AMB, 3 WUR, 2 CNIO, 1 CSIC, 2 CERTH, 1 MN, 1 ISTHMUS, 1 ULB)

The session started by short presentations to provide all information relevant to the jamboree, i.e. a short overview of the project of genome re-sequencing and re-annotation of *P. putida* KT2440, reconstruction of *P. putida* KT2440 metabolic network from re-annotated genome sequence, and integration of network GPR into existing *P. putida* KT2440 genome-scale metabolic models.

Participants worked in small groups during the curation session. At the end of the day, an overview of the curation results was presented and future steps of the *P. putida* KT2440 genome re-annotation and refinement of the metabolic model have been discussed.

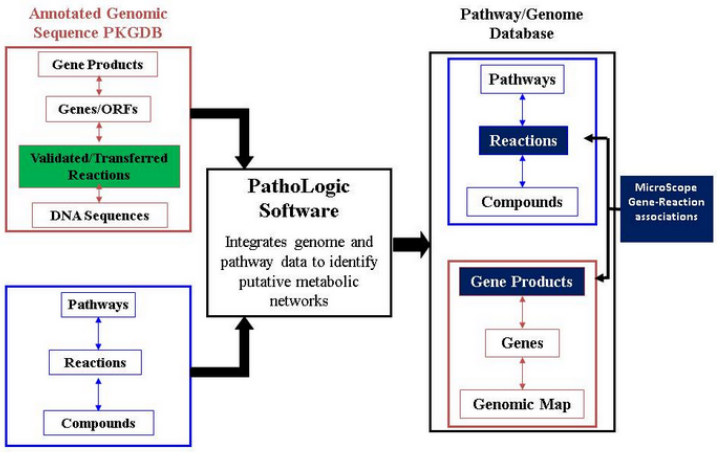
Agenda:
09h00-09h30: Introduction to the work being done on *P. putida* KT2440 prior to the jamboree and purpose of the day
09h30-10h30: Brief summary of the curation tools: MicroScope and Text Mining
10h30-11h00: Finding gene candidates and curation of Gene-Reaction associations in MicroScope: a quick example
11h00-13h00: Free curation on selected compounds
13h00-14h00: Lunch
14h00-18h00: Free curation on selected compounds
18h00-19h00: Global results of the curation, discussion and conclusions

Data and Methods

All information related to the horizontal annotation of *P. putida* KT2440 has been collected in the Microme confluence web site: http://www.ebi.ac.uk/seqdb/confluence/display/Microme/Data+Jamboree+2013

1. *P. putida* KT2440 genome re-annotation and re-sequencing project. Metabolic network reconstruction

The genome of *P. putida* KT2440 and the mutant strain TM407 (deleted strain by using mini-transposon Tn5 system; 198 genes deleted, mostly hypothetical/conserved hypothetical proteins) have been re-sequenced at Genoscope in the context of the Microme project. Genome sequences have been re-annotated using the Microscope annotation pipelines and the corresponding genome-scale metabolic networks has been reconstructed from genome annotations using the Microscope pipeline based on the PathoLogic algorithm of Pathway Tools:



In the context of the *P. putida* KT2440 re-annotation, 3 different sources of Gene-Reaction associations were used by the metabolic reconstruction pipeline in order to improve/extend the content of *P. putida* KT2440 metabolic network:

1. **Gene-Reaction associations automatically transferred from *E. coli* K12:** annotation transfer from *E. coli* K12 to *P. putida* KT2440 using Microscope automatic annotation pipeline based on Bi-directional Best Hit (BBH) relationships and synteny conservation. Annotation transfer includes *E. coli* K12 Gene-Reaction associations from:
    - *E. coli* K12 manual annotations stored in MicroScope (AMB, EBI, LABGEM; MetaCyc-RHEA reaction repository)
    - *E. coli* K12 projected metabolic models (iAF1260, iJO1366) on RHEA (LABGEM, SIB; RHEA reaction repository)
    - EcoCyc
2. **Manually validated gene-reaction associations in Microscope annotation of *P. putida* KT2440 (MetaCyc-RHEA reaction repository)**
3. **Predicted associations using PathoLogic algorithm based on re-annotated *P. putida* KT2440 genome sequence (MetaCyc reactions)**

The result of the Microscope metabolic reconstruction pipeline over re-annotated *P. putida* KT2440 genome sequence is a metabolic network composed by:

- 1477 CDSs
- 1612 reactions (MetaCyc-RHEA reactions)

2. *P. putida* KT2440 metabolic models

The genome-scale metabolic network of *P. putida* KT2440 has been reconstructed and refined by the group of Vitor Martin Dos Santos (WP3):

- Original model reconstruction (**iJP815**; Puchałka J et-al; PLoS Comput Biol. 2008 Oct;4(10):e1000210. doi: 10.1371)
- Refined model reconciled with *P. aeruginosa* metabolic model (**iJP962**;Oberhardt MA et-al; PLoS Comput Biol. 2011 Mar;7(3):e1001116. doi: 10.1371)

Moreover, an additional model of *P. putida* KT2440 (**iJN746**) from Palsson's group that models specifically aromatic compounds degradation was also used in the context of the jamboree (Nogales J et-al;BMC Syst Biol. 2008 Sep 16;2:79. doi: 10.1186).

In order to facilitate the integration of additional GPR from metabolic network reconstruction and FBA simulations from experimental data, the working versions of the metabolic models of *P. putida* KT2440 iJP962 and iJN746 have been defined in the MetaNetX

naming space (www.metanetx.org)

3. **Experimental data of growth capabilities of *P. putida* KT2440**

Two different sources of information were used in the context of the Jamboree:

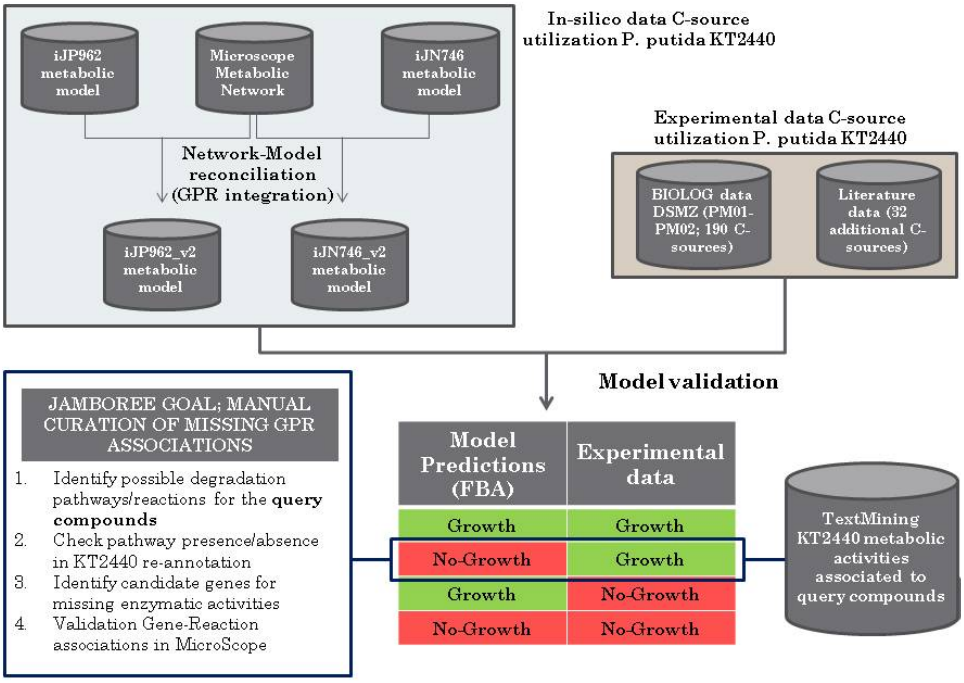
- **Biolog experiments generated by DMSZ:** the composition of Biolog plates PM01, PM02 (C-sources) were mapped in CHEBI-MetaCyc-MetaNetX naming space in order to facilitate FBA simulations.
- **Additional growth phenotype information of *P. putida* coming from the literature:** An additional set of 32 compounds, supported by experimental evidence on *P. putida* KT2440 in the literature, have been included in the analysis. Compounds have been mapped in CHEBI-MetaCyc-MetaNetX naming space.

4. **Integration of metabolic network GPR into *P. putida* KT2440 metabolic models and experimental phenotype validation by Flux Balance Analysis (FBA)**

In order to identify case studies to focus on during the jamboree, Gene-Reaction associations coming from the metabolic network reconstruction of re-annotated *P. putida* KT2440 genome sequence have been integrated into iJP962 and iJN746 metabolic models using MetaNetX naming space. This allows to reconcile resources at reaction level and to extend reaction space of the metabolic models with new reactions coming from the genome re-annotation process.
Following this integration step, original and reconciled models of *P. putida* KT2440 have been validated against experimental growth phenotypes by FBA, and a set of 47 compounds were identified for which experimental data report positive growth on *P. putida* KT2440 but no biomass production was observed in FBA simulations. These compounds have been selected as targets for the curation process during the Jamboree. The curation job includes:

1. Identification of degradation pathways of a set of compounds (from Biolog plates and from compound lists related to aromatic compounds supported by experimental evidence) for which we have positive growth phenotypes on KT2440 (from Biolog plates/from the literature evidences)
2. Identification of candidate genes in KT2440 for missing enzymatic activities associated to degradative pathways identified in the previous step by using tools and resources developed by different Microme partners
3. Validation/rejection of predicted GPRs and the addition of missing reactions by using Microscope curation environment

In order to assist the curation, the Text Mining strategy TeBactEn developed by CNIO has been applied in order to identify relevant literature evidences associated to *P. putida* enzymatic activities related to query compounds.



6. Data curation

The MicroScope platform (https://www.genoscope.cns.fr/agc/microscope) was used to perform gene-reaction association curation. To this purpose, MicroScope accounts were created for each participant. Selected query compounds were compiled in the confluence page, with direct links to reactions and pathways involving these compounds both in *P. putida* KT2440 metabolic network reconstructed from re-annotated genome sequence in MicroScope and in MetaCyc:

Query compounds to focus during the Jamboree

Modifier Partager Ajouter Outils

1 Ajouté par Eugenio Belda, modifié par Eugenio Belda le avr. 22, 2013 (afficher les modifications)

Compounds from Biolog plates PM01-PM02

Compounds selected based on positive growth on Biolog experiments / No prowth on FBA simulations

To access to Putida MicrCyc link:

- Username: guest
- Password: guest

Plate	Well	Name	Putida MicroCyc link	MetaCyc Link	Biocyc ID	Biolog raw measure	Biolog growth phenotype	CHEBI ID	MetaNetx_ID	iJP962 FBA phenotype	iJP962_extendedMicroscope FBA phenotype	iJN746 phenotype
PM01	B05	D-Glucuronic Acid	GLUCURONATE	GLUCURONATE	GLUCURONATE	287	1	4178	MNXM860	-1	-1	-1
PM01	C03	D,L-Malic Acid	MAL	MAL	MAL	261	1	15595	MNXM90553	-1	-1	-1
PM01	C05	Tween 20	CPD-3566	CPD-3566	CPD-3566	177	1	53424	MNXM77769	-1	-1	-1
PM01	C09	a-D-Glucose	ALPHA-GLUCOSE	ALPHA-GLUCOSE	ALPHA-GLUCOSE	310	1	17925	MNXM99	-1	-1	-1
PM01	D05	Tween 40	CPD-3567	CPD-3567	CPD-3567	185	1	53423	MNXM77770	-1	-1	-1
PM01	E05	Tween 80	CPD-3563	CPD-3563	CPD-3563	142	1	53426	MNXM77773	-1	-1	-1
PM01	E06	a-Hydroxy Glutaric Acid-g-Lactone	CPD-13414	CPD-13414	CPD-13414	226	1	-	MNXM89495	-1	-1	-1
PM01	E07	a-Hydroxy-Butyric Acid	CPD-3564	CPD-3564	CPD-3564	83	1	-	MNXM4968	-1	-1	-1
PM01	F06	Bromo-Succinic Acid	CPD-13405	CPD-13405	CPD-13405	86	1	-	MNXM44792	-1	-1	-1
PM01	G01	Glycyl-L-Glutamic Acid	CPD-3569	CPD-3569	CPD-3569	93	1	-	MNXM55454	-1	-1	-1
PM01	G06	L-Alanyl-Glycine	ALA-GLY	ALA-GLY	ALA-GLY	320	1	-	MNXM15783	-1	-1	-1
PM01	G10	Methyl Pyruvate	CPD-3573	CPD-3573	CPD-3573	166	1	51850	MNXM61806	-1	-1	-1
PM01	H01	Glycyl-L-Proline	CPD-10814	CPD-10814	CPD-10814	97	1	70744	MNXM11725	-1	-1	-1
PM01	H10	D-Galacturonic Acid	D-GALACTURONATE	D-GALACTURONATE	D-GALACTURONATE	291	1	18024	MNXM4445	-1	-1	-1
PM02	A12	Pectin	PECTIN	PECTIN	PECTIN	93	1	17309	MNXM2706	-1	-1	-1
PM02	E08	b-Hydroxy-Butyric Acid	CPD-335	CPD-335	CPD-335	322	1	37054	MNXM10022	-1	-1	-1
PM02	E10	a-Keto-Valeric Acid	CPD-3618	CPD-3618	CPD-3618	253	1	28644	MNXM4302	-1	-1	-1
PM02	F01	D-Lactic Acid Methyl Ester	CPD-3621	CPD-3621	CPD-3621	92	1	-	MNXM61755	-1	-1	-1
PM02	F07	D-Ribono-1,4-Lactone	CPD-13413	CPD-13413	CPD-13413	327	1	-	MNXM48597	-1	-1	-1
PM02	F09	Sorbic Acid	CPD-3624	CPD-3624	CPD-3624	131	1	35962	MNXM82480	-1	-1	-1
PM02	G02	L-Alaninamide	CPD-7692	CPD-7692	CPD-7692	250	1	-	MNXM59268	-1	-1	-1
PM02	G08	4-Hydroxy-L-Proline (trans)	L-4-HYDROXY-PROLINE	L-4-HYDROXY-PROLINE	L-4-HYDROXY-PROLINE	327	1	18240	MNXM4365	-1	-1	-1

In order to allow curators to select their compounds during the jamboree and to summarize the results of the curation process, an additional Google spreadsheet was created, that allows dynamic edition by different curators simultaneously during the jamboree.

Each participant was free to use the web tools that he/she was most familiar with. The MicroScope platform implements several functionalities that guide the curation process, such as synteny maps, the phyloprofile tool (which performs searches for co-evolved genes), and the CanOE strategy. A short summary of all these tools has been added to the Microme confluence page.
All the curation work carried out during the jamboree was stored into the Microscope platform. This includes:

- The association of genes to RhEA and MetaCyc reactions
- The update of gene annotations (including the description of gene product, EC number, bibliographical references, etc)

Results

At the end of the Jamboree, 18 compounds were selected by the curators from the query compound list. At the end of the day we obtained:

- 44 genes manually curated in the MicroScope platform
- 27 genes with manually validated Gene-Reaction associations
- 25 MetaCyc reactions manually validated
- 19 RhEA reactions manually validated
- Degradation pathways detected for 5 of these compounds, connecting them with compounds of the central metabolism

Conclusions and Perspectives

The main objective of this Jamboree was to reconcile experimental data of *P. putida* KT2440 growth phenotypes with metabolic model predictions of biomass production for a set of 47 carbon sources where positive growth in experiments and no biomass production was observed in FBA model simulations. After discussion with the Jamboree participants, WP2 will finish the curation of the 47 carbon sources initiated during the jamboree, and curated GPRs will be handled by WP3 for their progressive integration and validation into metabolic models of *P. putida* KT2440.

This work will be included into the project of genome re-sequencing and re-annotation of *P. putida* KT2440 that is taking place in Microme. General discussion also focused on the definition of the next steps that have to be completed for the final publication, which is planned by the end of the year 2013. The following timeline has been adopted:

Realease	Date	Leader	Impact Layer	Update Description
R1	June 13	CEA	Network	Query compound jamboree
R2	Sept 13	WU	Model	Integration of R1
R3	Sept 13	CEA	Network Transport	Non carbon source Biolog Integration of EBI predictions
R4	Oct 13	WU	Model Model	Integration of R3 Detection of conflict on NonC
R5	Oct 13	CEA	Network Network	putative enzyme curation Tag GPR with evidence level
R6	Nov 13	WU	Model	Integration of R5
R7	Nov 13	CEA	Network Pathway	Core model compound Definition specific pathways
R8	Dec 13	WU	Model	Integration of R7