

MrBac: A web server for draft metabolic network reconstructions for bacteria

Yu-Chieh Liao, Ph.D.
Assistant Investigator
National Health Research Institutes

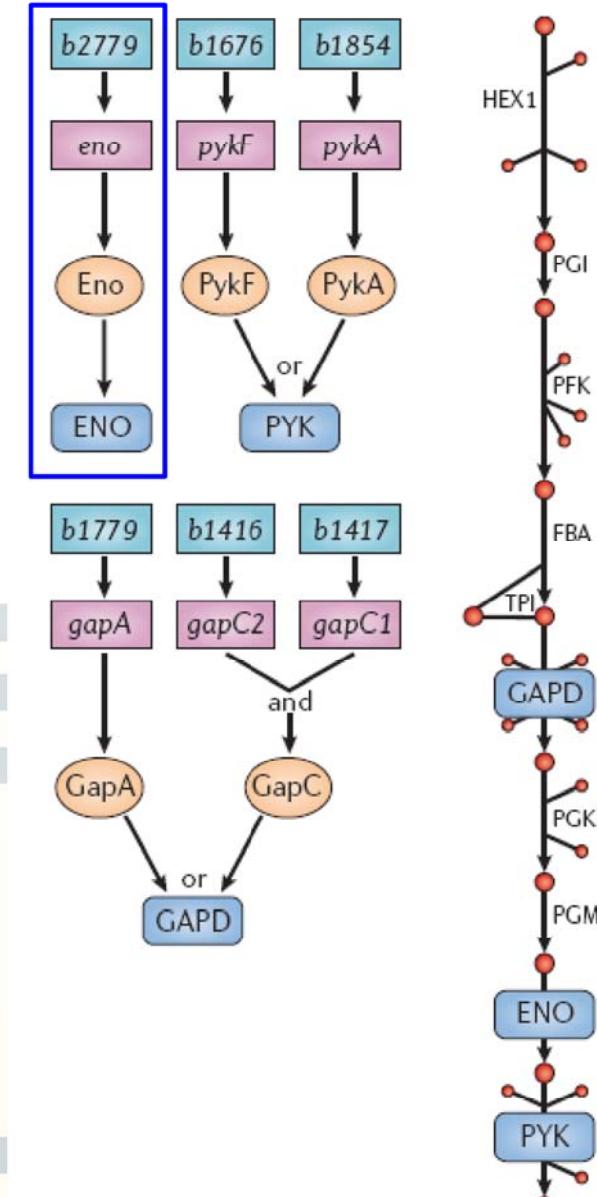
<http://sb.nhri.org.tw/MrBac>

Box 3 | Assembly and representation

A list of charge and elementally balanced metabolic reactions can be represented in a stoichiometric matrix (S), where rows and columns correspond to metabolites and reactions and the elements are the stoichiometric coefficients. In genome-scale metabolic networks these stoichiometric matrices contain few non-zero elements, as relatively few metabolites participate in a given reaction. Connections between genes and reactions can be represented as gene–protein–reaction (GPR) associations by using Boolean rules or visualized using graphic images. In the GPR scheme, the first level (teal) corresponds to genetic loci, the second level (pink) to transcripts, the third level (orange) to functional proteins, and the fourth level (blue) to reactions. [c], cytoplasmic reactions.

Abbreviation	Glycolytic reactions	Genes
HEX1	[c]GLC + ATP → G6P + ADP + H	<i>glk</i>
PGI	[c]G6P ↔ F6P	<i>pgi</i>
PFK	[c]ATP + F6P → ADP + FDP + H	<i>pfkA, pfkB</i>
FBA	[c]FDP ↔ DHAP + G3P	<i>fbaA, fbaB</i>
TPI	[c]DHAP ↔ G3P	<i>tpiA</i>
GAPD	[c]G3P + NAD + PI ↔ 13DPG + H + NADH	<i>gapA, gapC1, gapC2</i>
PGK	[c]13DPG + ADP ↔ 3PG + ATP	<i>pgk</i>
PGM	[c]3PG ↔ 2PG	<i>gpmA, gpmB</i>
ENO	[c]2PG ↔ H ₂ O + PEP	<i>eno</i>
PYK	[c]ADP + H + PEP → ATP + PYR	<i>pykA, pykF</i>

	HEX1	PGI	PFK	FBA	TPI	GAPD	PGK	PGM	ENO	PYK
ATP	-1	0	-1	0	0	0	1	0	0	1
GLC	-1	0	0	0	0	0	0	0	0	0
ADP	1	0	1	0	0	0	-1	0	0	-1
G6P	1	-1	0	0	0	0	0	0	0	0
H	1	0	1	0	0	1	0	0	0	-1
F6P	0	1	-1	0	0	0	0	0	0	0
FDP	0	0	1	-1	0	0	0	0	0	0
DHAP	0	0	0	1	-1	0	0	0	0	0
G3P	0	0	0	1	1	-1	0	0	0	0
NAD	0	0	0	0	0	-1	0	0	0	0
PI	0	0	0	0	0	-1	0	0	0	0
13DPG	0	0	0	0	0	1	-1	0	0	0
NADH	0	0	0	0	0	1	0	0	0	0
3PG	0	0	0	0	0	0	1	-1	0	0
2PG	0	0	0	0	0	0	0	1	-1	0
PEP	0	0	0	0	0	0	0	0	1	-1
H ₂ O	0	0	0	0	0	0	0	0	1	0
PYR	0	0	0	0	0	0	0	0	0	1



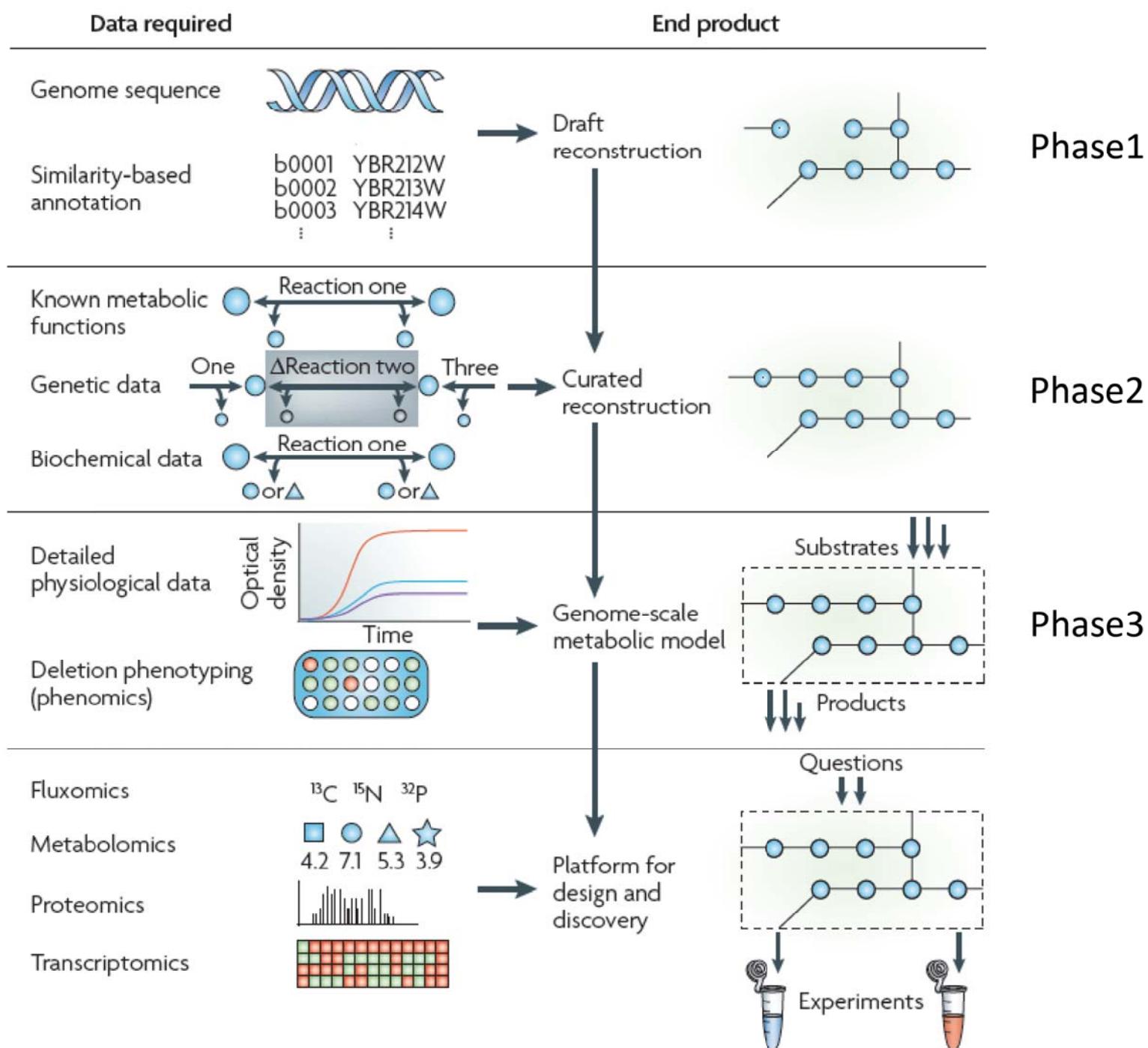


Figure 1 | Phases and data used to generate a metabolic reconstruction.

MrBac (<http://sb.nhri.org.tw/MrBac>)

A web server to draft



| Home | Draft reconstruction | About | How-to | Contact | Supplementary |

*Select a query and a reference species: [[Example](#)]

Query Species:	Reference Species:
Acaryochloris marina MBIC11017	Acaryochloris marina MBIC11017
Acholeplasma laidlawii PG-8A	Acholeplasma laidlawii PG-8A
Acidiphilium cryptum JF-5	Acidiphilium cryptum JF-5
Acidithiobacillus ferrooxidans ATCC 23270	Acidithiobacillus ferrooxidans ATCC 23270
Acidithiobacillus ferrooxidans ATCC 53993	Acidithiobacillus ferrooxidans ATCC 53993
Acidobacteria bacterium Ellin345	Acidobacteria bacterium Ellin345
Acidobacterium capsulatum ATCC 51196	Acidobacterium capsulatum ATCC 51196
Acidothermus cellulolyticus 11B	Acidothermus cellulolyticus 11B
Acidovorax citrulli AAC00-1	Acidovorax citrulli AAC00-1
Acidovorax sp. JS42	Acidovorax sp. JS42

*Choose Blast Settings: [[Default](#) / [Select All](#)]

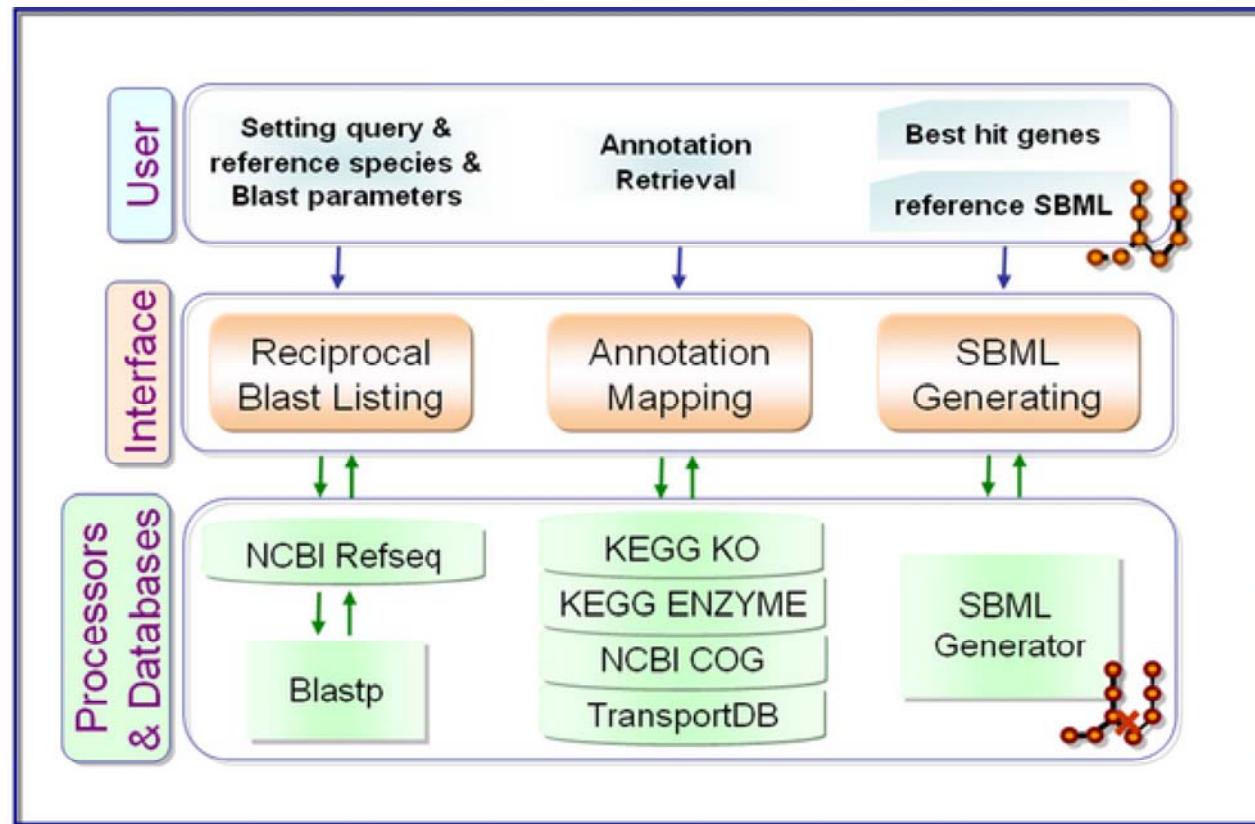
E value: & Percent Identity: %

Choose the blast columns to be shown: [?](#)

Ref Query Perclden MLen mismatches gap openings q.start q.end r.start r.end EVal BitScore

* indicates required input.

Flow Chart



Reciprocal Blast

BLAST® Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

NCBI/ BLAST/ blast suite Align Sequences Protein BLAST

blastn blastp blastx tblastn tblastx

Enter Query Sequence
Enter accession number(s), gi(s), or FASTA sequence(s) Clear Query subrange
From To

Or, upload file 未選擇檔案

Job Title Enter a descriptive title for your BLAST search

Align two or more sequences

Enter Subject Sequence
Enter accession number, gi, or FASTA sequence Clear Subject subrange
From To

Or, upload file 未選擇檔案

Program Selection
Algorithm blastp (protein-protein BLAST)
Choose a BLAST algorithm

BLAST Search protein sequence using Blastp (protein-protein BLAST)
 Show results in a new window

+Algorithm parameters

<http://blast.ncbi.nlm.nih.gov/>

http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=Download

Query	Ref	Percliden	MLen	mismatches	gap openings	q.start	q.end	r.start	r.end	EVal	Percliden2	MLen2	mismatches2	gap openings2	q.start2	q.end2	r.start2	r.end2	EVal2
STM0002	b0002	94.51	820	45	0	1	820	1	820	0.0	94.51	820	45	0	1	820	1	820	0.0
STM0003	b0003	93.51	308	20	0	1	308	1	308	2e-165	93.51	308	20	0	1	308	1	308	6e-154
STM0004	b0004	93.46	428	28	0	1	428	1	428	0.0	93.46	428	28	0	1	428	1	428	0.0
STM0005	b0006	86.38	257	35	0	1	257	1	257	6e-132	86.38	257	35	0	1	257	1	257	6e-132
STM0006	b0007	76.26	476	113	0	1	476	1	476	0.0	76.26	476	113	0	1	476	1	476	0.0
STM0007	b0008	94.64	317	17	0	1	317	1	317	0.0	94.64	317	17	0	1	317	1	317	0.0

Salmonella enterica subsp. enterica serovar Typhimurium str. LT2 versus Escherichia coli str. K-12 substr. MG1655

Your current blast settings are

Evalue: 1e-5

PercIden: 50.0

Show all Query Genes: Yes No

*Choose the annotation columns to be shown for both query and reference: [?](#)

<input checked="" type="checkbox"/> QueryID	<input checked="" type="checkbox"/> gene			
<input type="checkbox"/> 5'Coordinate	<input type="checkbox"/> 3'Coordinate	<input type="checkbox"/> Product	<input type="checkbox"/> ProteinLength	
<input type="checkbox"/> KO:	<input type="checkbox"/> KONum	<input type="checkbox"/> KODEFIN	<input type="checkbox"/> KOCLASS	
<input type="checkbox"/> EC:	<input type="checkbox"/> ECNum	<input type="checkbox"/> ECNAME	<input type="checkbox"/> ECCLASS	<input type="checkbox"/> ECSYSNAME
<input type="checkbox"/> COG:	<input type="checkbox"/> FunID	<input type="checkbox"/> FUNCcategory	<input type="checkbox"/> FUNDesc	
<input type="checkbox"/> MT:	<input type="checkbox"/> TransporterType	<input type="checkbox"/> FamilyName	<input type="checkbox"/> Substrate	

<input checked="" type="checkbox"/> ReferenceID	<input checked="" type="checkbox"/> gene			
<input type="checkbox"/> KO:	<input type="checkbox"/> KONum	<input type="checkbox"/> KODEFIN	<input type="checkbox"/> KOCLASS	
<input type="checkbox"/> EC:	<input type="checkbox"/> ECNum	<input type="checkbox"/> ECNAME	<input type="checkbox"/> ECCLASS	<input type="checkbox"/> ECSYSNAME
<input type="checkbox"/> COG:	<input type="checkbox"/> FunID	<input type="checkbox"/> FUNCcategory	<input type="checkbox"/> FUNDesc	
<input type="checkbox"/> MT:	<input type="checkbox"/> TransporterType	<input type="checkbox"/> FamilyName	<input type="checkbox"/> Substrate	

[Back to blast filtering](#)

[Proceed to annotation matching...](#)

Annotation retrieval

Salmonella enterica subsp. enterica serovar Typhimurium str. LT2 versus Escherichia coli str. K-12 substr. MG1655

Evalue: 1e-5 Percent Identity: 50.0

MrBac

[<< Back to blast filtering](#)
[< Back to column selection](#)

[Save as .xls file](#)

[Save Gene Pair File](#)
Metabolic Network Generator>

Query	gene	5'Coordinate	3'Coordinate	Product	ProteinLength	KONum	KODEFIN	ECCLASS	ECNum	ECN
PSLT067	--	54974	55216	putative cytoplasmic protein	80	--	--	--	--	--
STM0002	thrA	337	2799	bifunctional aspartokinase I/homoserine dehydrogenase I	820	K00003;;K00928	homoserine dehydrogenase [EC:1.1.1.3];;aspartate kinase [EC:2.7.2.4]	Metabolism; Amino Acid Metabolism; Glycine, serine and threonine;metabolism [PATH:ko00260];Metabolism; Amino Acid Metabolism; Cysteine and methionine;metabolism [PATH:ko00270];Metabolism; Amino Acid Metabolism; Lysine biosynthesis;[PATH:ko00300]	1.1.1.13;;1.1.1.3;;2.7.2.4	L-arabinitol 2-dehydrogenase; L-a forming); L-arabinitol (ribulose-for homoserine dehydrogenase; HSDI aspartokinase; AK; beta-aspartoki
STM0003	thrB	2801	3730	homoserine kinase	309	K00872	homoserine kinase [EC:2.7.1.39]	Metabolism; Amino Acid Metabolism; Glycine, serine and threonine;metabolism [PATH:ko00260]	2.7.1.39	homoserine kinase; homoserine ki
STM0004	thrC	3734	5020	threonine synthase	428	K01733	threonine synthase [EC:4.2.3.1]	Metabolism; Amino Acid Metabolism; Glycine, serine and threonine;metabolism [PATH:ko00260];Metabolism; Metabolism of Cofactors and Vitamins; Vitamin B6;metabolism [PATH:ko00750]	4.2.3.1	threonine synthase; threonine synt phospho-lyase (adding water)
STM0005	yaaA	5887	5114	hypothetical protein	257	K09861	hypothetical protein	Unclassified; Poorly Characterized; Function unknown	--	--
STM0006	yaaJ	7396	5966	alanine/glycine transport protein	476	K03310	alanine or glycine:cation symporter, AGCS family	Environmental Information Processing; Membrane Transport; Other;ion-coupled transporters	--	--

Metabolic Network Draft Generation

Draft metabolic network generation requires a reference SBML file and the list of reciprocal blast best hits as inputs to proceed.

By taking these two inputs, we are assuming that query ORFs in the best hit list have the same functions as their corresponding reference ORFs. The generator only keeps reactions, in which all associated reference ORFs under "AND" condition and partial reference ORFs under "OR" condition have best hit matches to query ORFs. Only when those two criteria are met, the reference ORF id are replaced by corresponding query id, and the metabolites involved in those reactions are kept in the draft SBML file.

The draft in SBML format can be saved as the final output after query is submitted and can be further curated and edited manually in flux balance analysis softwares such as [COBRA](#).

Upload a reference SBML file ([sample_iAF1260](#)) [?](#)

SBML File:

Upload a one to one gene pair file ([sample_SLT2vsECO](#)) [?](#)

Gene Pair File:

Type in a model name

New Model Name:

```
<?xml version="1.0" encoding="UTF-8"?>
<sbml level="2" version="1" >
  <model id="Ec_iAF1260" name="Ec_iAF1260" >
    <listOfUnitDefinitions>
      <unitDefinition id="mmol_per_gDW_per_hr" >
        <listOfUnits>
          <unit kind="mole" scale="-3" />
          <unit exponent="-1" kind="gram" />
          <unit exponent="-1" kind="second" multiplier=".00027777" />
        </listOfUnits>
      </unitDefinition>
    </listOfUnitDefinitions>
    <listOfCompartments>
      <compartment id="Extra_organism" />
      <compartment id="Periplasm" outside="Extra_organism" />
      <compartment id="Cytosol" outside="Periplasm" />
    </listOfCompartments>
    <listOfSpecies>
      <species boundaryCondition="false" charge="-2" compartment="Cytosol" id="M_10fthf_c" name="M_10_Formyltetrahydrofolate_C2">
      <species boundaryCondition="false" charge="0" compartment="Cytosol" id="M_12ppd_R_c" name="M_R_Propane_1_2_diol_C3H8O2">
      <species boundaryCondition="false" charge="0" compartment="Cytosol" id="M_12ppd_S_c" name="M_S_Propane_1_2_diol_C3H8O2">
      <species boundaryCondition="false" charge="-4" compartment="Cytosol" id="M_13dpq_c" name="M_3_Phospho_D_glyceroyl_phospho">
      <species boundaryCondition="false" charge="0" compartment="Cytosol" id="M_14glucan_c" name="M_1_4_alpha_D_glucan_C36H62O3">
      <species boundaryCondition="false" charge="0" compartment="Periplasm" id="M_14glucan_p" name="M_1_4_alpha_D_glucan_C36H62O3">
      <species boundaryCondition="false" charge="2" compartment="Cytosol" id="M_15dap_c" name="M_1_5_Diaminopentane_C5H16N2" />
      <species boundaryCondition="false" charge="2" compartment="Periplasm" id="M_15dap_p" name="M_1_5_Diaminopentane_C5H16N2" />
      <species boundaryCondition="false" charge="0" compartment="Periplasm" id="M_lagepe120_p" name="M_1_Acyl_sn_glycero_3_phospho">
      <species boundaryCondition="false" charge="0" compartment="Periplasm" id="M_lagepe140_p" name="M_1_Acyl_sn_glycero_3_phospho">
      <species boundaryCondition="false" charge="0" compartment="Periplasm" id="M_lagepe141_p" name="M_1_Acyl_sn_glycero_3_phospho">
      <species boundaryCondition="false" charge="0" compartment="Periplasm" id="M_lagepe160_p" name="M_1_Acyl_sn_glycero_3_phospho">
      <species boundaryCondition="false" charge="0" compartment="Periplasm" id="M_lagepe161_p" name="M_1_Acyl_sn_glycero_3_phospho">
      <species boundaryCondition="false" charge="0" compartment="Periplasm" id="M_lagepe180_p" name="M_1_Acyl_sn_glycero_3_phospho">
      <species boundaryCondition="false" charge="0" compartment="Periplasm" id="M_lagepe181_p" name="M_1_Acyl_sn_glycero_3_phospho">
      <species boundaryCondition="false" charge="-1" compartment="Periplasm" id="M_lagpg120_p" name="M_1_Acyl_sn_glycero_3_phospho">
      <species boundaryCondition="false" charge="-1" compartment="Periplasm" id="M_lagpg140_p" name="M_1_Acyl_sn_glycero_3_phospho">
    </listOfSpecies>
  </model>
</sbml>
```

Starting materials for manual curation

3	STM1097--	1184453	1184863	hypothetical protein	136	K07127	--	Unclassifi	--	--	--	--	R	POORLY	General ft	--	--	b1970	hiuH	
4	STM1098 hpaC	1185708	1185196	4-hydroxyphenylacetate catabolism	170	K00484	4-hydroxy Metabolis	1.14.13.3	4-hydroxy Oxidoreduc	4-hydroxy R					POORLY	General ft	--	--	--	--
5	STM1099 hpaB	1187288	1185726	4-hydroxyphenylacetate catabolism	520	K00483	4-hydroxy Metabolis	1.14.13.3	4-hydroxy Oxidoreduc	4-hydroxy Q					METABC	Secondary	--	--	--	--
5	STM1100 hpaR	1187946	1187506	4-hydroxyphenylacetate catabolism	146	--	--	--	--	--	--	--	K	INFORM.	Transcript	--	--	--	--	
7	STM1101 hpaG	1188221	1189510	4-hydroxyphenylacetate catabolism	429	K05921	5-oxopent Metabolis	4.1.1.68;	5-oxopent Lyases; C	5-oxopent Q					METABC	Secondary	--	--	--	--
3	STM1102 hpaE	1189507	1190973	4-hydroxyphenylacetate catabolism	488	K00151	5-carboxy Metabolis	1.2.1.60	5-carboxy Oxidoreduc	5-carboxy C					METABC	Energy pr	--	--	--	--
9	STM1103 hpaD	1190975	1191826	4-hydroxyphenylacetate catabolism	283	K00455	3,4-dihyd	Metabolis	1.13.11.1	3,4-dihyd	Oxidoreduc	3,4-dihyd S		POORLY	Function t	--	--	--	--	
0	STM1104 hpaF	1191836	1192216	4-hydroxyphenylacetate catabolism	126	K01826	5-carboxy Metabolis	5.3.3.10	5-carboxy Isomerase	5-carboxy E					METABC	Amino aci	--	--	--	--
1	STM1105 hpaH	1192359	1193162	4-hydroxyphenylacetate catabolism	267	K02509	2-oxo-hep Metabolis	4.2.1.-	--	Lyases; C	--	--	Q		METABC	Secondary	--	--	--	--
2	STM1106 hpaI	1193173	1193964	4-hydroxyphenylacetate catabolism	263	K02510	2,4-dihyd	Metabolis	4.1.2.-	--	Lyases; C	--	--	G	METABC	Carbohydr	--	--	--	--
3	STM1107 hpaX	1194036	1195412	4-hydroxyphenylacetate catabolism	458	K02511	MFS trans	Protein F	--	--	--	--		G;; E;; P;; METABC	Carbohydr	Secondary	The Majo	4-hydroxy	--	--
4	STM1108 hpaA	1195422	1196318	4-hydroxyphenylacetate catabolism	298	K02508	AraC fam:	Protein F	--	--	--	--	K	INFORM.	Transcript	--	--	--	--	--
5	STM1109--	1196332	1197270	putative periplasmic protein	312	--	--	--	--	--	--	--	R	POORLY	General ft	--	--	--	--	--

- We can identify, for instance, the metabolic genes that are present in SLT2 but absent in ECO.
- Examples include the cob (STM2016-STM2035), hpa (STM1098-STM1108) and pdu (STM2037-STM2058).
- These genes have been reported to confer SLT2 the abilities to synthesize cobalamin and to utilize 4-hydroxylphenylacetate and propanediol.

A quick look at GEMSiRV

<http://sb.nhri.org.tw/GEMSiRV/en/GEMSiRV>

The screenshot shows the GEMSiRV software interface. At the top, there is a large watermark-like graphic of a metabolic pathway map with various enzymes like FUM, SUCDI, MALS, ME1, ME2, CS, ACONTe, ACONTb, PGM, ENO, PPS, PFL, and PFK. To the right of the map, the word "GEMSiRV" is written in large, bold, black letters, with a blue and white striped swoosh underneath. The NHRI logo is in the top right corner. Below the main title, there is a navigation bar with links for "文章" (Articles), "註解" (Annotations), "歷史" (History), "連結" (Links), and "列印" (Print). On the far right of the navigation bar are "登入" (Login) and "註冊" (Register) buttons. The main content area has a sidebar on the left with sections for "HOME", "Guide", "Repository", and "Toolbox". The "Guide" section contains links for "Download", "Installation", "Manual", and "Users' Comments". The "Repository" section contains links for "Metabolic Models", "Reference Databases", and "Metabolic Maps". The "Toolbox" section contains links for "TextReplacer", "GBKParser", and "MrBac". The main content area has a heading "GEMSiRV" and a sub-heading "A software platform for GEome-scale Metabolic models Simulation, Reconstruction and Visualization." It also mentions that the work has been published in Bioinformatics (Pubmed). There are several paragraphs of text describing the software's features and benefits.

文章 註解 歷史 連結 列印

登入 註冊

HOME

Guide

Download
Installation
Manual
Users' Comments

Repository

Metabolic Models
Reference Databases
Metabolic Maps

Toolbox

TextReplacer
GBKParser
MrBac

GEMSiRV

A software platform for GEome-scale Metabolic models Simulation, Reconstruction and Visualization.

This work has been published in Bioinformatics (Pubmed). If you have any question, please do not hesitate to contact the author by emailing to sysbio@nhri.org.tw.

Start GEMSiRV! An example model (*E. coli* textbook model) and its map can be directly imported by click on the Help to Import an Example.

Genome-scale metabolic network models have become an indispensable part of the increasingly important field of systems biology.

Metabolic systems biology studies usually include three major components – network model construction, objective- and experiment-guided model editing and visualization, and simulation studies based mainly on flux balance analyses.

Here we present a software platform, GEMSiRV, to provide functionalities of easy metabolic network drafting and editing, amenable network visualization for experimental data integration, and flux balance analysis tools for simulation studies.

GEMSiRV is an open-source software for building users' metabolic systems biology project, it provides interactive features in model management, simulation, visualization and integration of omics data.

Furthermore, all of the GEMSiRV-generated metabolic models and analysis results, including projects in progress, can be easily exchanged in the research community. GEMSiRV is a powerful integrative resource that may facilitate the development of systems biology studies.

A schematic overview of the GEMSiRV outlines the key features implemented in the GEMSiRV.

Acknowledgements

- Julie Chih-Yu Chen
 - Ming-Hsin Tsai
 - Yueh-Hsia Tang
-
- Dr. Feng-Chi Chen
 - Dr. Chao A. Hsiung