

How to use modelBorgifier

Author: John T. Sauls, UCSD

Reviewer: Almut Heinken, Luxembourg Centre for Systems Biomedicine

INTRODUCTION

modelBorgifier is a package that allows users to compare and combine COBRA Toolbox ("Toolbox") style metabolic reconstructions ("models"). It is explicitly designed with the notion that models from different sources use disparate naming and annotation schemes. It uses greedy string comparisons as well as network topology to identify reactions and metabolites shared and unique between models. The procedure is GUI based, and uses manual matches to train learning methods that facilitate auto-matching.

Please read the publication (1) and accompanying manual for more information. If you find this package helpful for your work please cite:

Sauls, J. T., & Buescher, J. M. (2014). Assimilating genome-scale metabolic reconstructions with modelBorgifier. *Bioinformatics* (Oxford, England), 30(7), 1036–8. <http://doi.org/10.1093/bioinformatics/btt747>

Correspondance: johntsauls@gmail.com

Developed at: BRAIN Aktiengesellschaft, Microbial Production Technologies Unit, Quantitative Biology and Sequencing Platform, Darmstaeter Str. 34-36, 64673 Zwingenberg, Germany. www.brain-biotech.de

PROCEDURE

In this tutorial we will compare the E. coli core metabolism model Ecoli_core to the Helicobacter pylori model iT341 (2). An outline of the procedure is as follows:

1. Installation and set-up

Assuming you have successfully installed and tested the COBRA Toolbox for Matlab no additional configuration should be necessary.

2. Load and verify the comparison model (Cmodel)

The comparison model (Cmodel) is any model that can be read into the COBRA Toolbox. Cmodel is "compared to" the template model (see next step). Our Cmodel will be the Ecoli_core model.

3. Load and verify the template model (Tmodel)

The template model (Tmodel) can be simply any model, or it can be an amalgamation of models that have already been combined via modelBorgifier. Our Tmodel will be iT341.

4. Compare models

Every reaction in Cmodel is compared against every reaction in Tmodel and given a similarity score based on 40 parameters. This computationally expensive step is done before user-guided matching.

5. Match models

Matching models is done with command `reactionCompare`. `reactionCompare` calls a GUI that allows the user to choose a match for a given reaction in Cmodel, and also to match the metabolites for that reaction. Comparison is facilitated by automation and proper weighting of the scoring parameters.

6. Merge models

Once all reactions and metabolites have been reviewed, Cmodel and Tmodel can be merged into a composite model. The composite model is the most direct way to return statistics on the similarity between two models.

7. Extract a model

Models merged together or into an existing composite model can be later retrieved with `readCbTmodel.m`. This reproduces the initial model with additional annotation information.

8. Save work

As the composite model can be used as a template for future comparisons, save it.

2. Load and verify the comparison model (Cmodel)

We first load the model we wish to compare, "Cmodel." `modelBorgifier` requires this model to be in a format readable by COBRA Toolbox. The verification step is specific to this package to ensure that the model has all necessary information arrays for comparison.

i. Load the E. coli core model

We are going to use the E. coli core model located in the `test/models/` directory of the Toolbox. We use this model for the tutorial because it is small and will require less time to match.

```
% Load the model using the Toolbox function
Cmodel = readCbModel('ecoli_core_model.mat', ...
                    'modelDescription', 'Ecoli_core')
```

```
Cmodel =
    rxns: {95x1 cell}
    mets: {72x1 cell}
    S: [72x95 double]
    lb: [95x1 double]
    ub: [95x1 double]
    c: [95x1 double]
    rxnNames: {95x1 cell}
    subSystems: {95x1 cell}
    rxnECNumbers: {95x1 cell}
    rxnReferences: {95x1 cell}
    rxnNotes: {95x1 cell}
    grRules: {95x1 cell}
    metNames: {72x1 cell}
    metFormulas: {72x1 cell}
    metKEGGID: {72x1 cell}
    metChEBIID: {72x1 cell}
    metPubChemID: {72x1 cell}
    metInChIString: {72x1 cell}
    genes: {137x1 cell}
    description: 'Ecoli_core'
    rules: {95x1 cell}
    rxnGeneMat: [95x137 double]
    b: [72x1 double]
```

```
rxnConfidenceScores: [95x1 double]
    metCharges: [72x1 int32]
        osense: -1
        csense: [72x1 char]
```

ii. Verify the model

modelBorgifier requires that both the comparison and template model have the proper data arrays before comparison. This function creates those arrays if they are absent and populates them when possible. You will also be prompted to keep or edit the model name ("description"). Simply press 'y' in for this tutorial.

Note that this verify function (verifyModelBorg) is different from the Toolbox function verifyModel.

```
% Verify model has the necessary fields required for later processing
Cmodel = verifyModelBorg(Cmodel, 'keepName', 'Verbose');
```

```
Array .rxnID not in Model. Adding.
Array .rxnKEGGID not in Model. Adding.
Array .rxnSEEDID not in Model. Adding.
Array .rxnEquations not in Model. Adding.
Array .metID not in Model. Adding.
Array .metSEEDID not in Model. Adding.
Array .metCharge not in Model. Adding.
Making sure reactions are all forwards
Fixing names of metabolites and reaction
Checking if reaction IDs (.rxns) are unique.
Checking if metabolite IDs (.mets) are unique.
All metabolites have compartment designation.
```

3. Load and verify the template model (Tmodel)

We now load the template model ("Tmodel"), to which Cmodel will be compared. If you are simply comparing two models, it is arbitrary which model is the Cmodel and which is the Tmodel. However, after comparison, the two models can be merged into a composite model (from which either of the original models can be retrieved). This composite model can be used as the Tmodel for future comparisons. This will make future comparisons easier, as there will be more annotations information available, and allows for multi-way model comparisons.

i. Load the model iT341

We will be using the *Helicobacter pylori* model packaged with the Toolbox as our template model. Load it the same way as any model. If you had previously combined two models using modelBorgifier, you could simply load that composite model as your Tmodel.

```
global CBTDIR
pth=which('initCobraToolbox.m');
CBTDIR = pth(1:end-(length('initCobraToolbox.m')+1));
Tmodel = readCbModel([CBTDIR filesep 'test' filesep 'models' filesep
'iT341.xml'], ...
    'modelDescription', 'iT341')
```

```
The model contains 0 errors and 1 warnings.
Tmodel =
```

```
rxns: {554x1 cell}
mets: {485x1 cell}
S: [485x554 double]
```

```

lb: [554x1 double]
ub: [554x1 double]
c: [554x1 double]
rxnNames: {554x1 cell}
metNames: {485x1 cell}
metFormulas: {485x1 cell}
metKEGGID: {485x1 cell}
metChEBIID: {485x1 cell}
genes: {339x1 cell}
description: 'iIT341'
modelVersion: [1x1 struct]
comps: {2x1 cell}
compNames: {2x1 cell}
compisbigg_46__compartmentID: {2x1 cell}
methMDBID: {485x1 cell}
metMetaNetXID: {485x1 cell}
metisbigg_46__metaboliteID: {485x1 cell}
metisbiocycID: {485x1 cell}
metisec_45__codeID: {485x1 cell}
metiskegg_46__reactionID: {485x1 cell}
metislipidmapsID: {485x1 cell}
metismetanetx_46__reactionID: {485x1 cell}
metisncbigiID: {485x1 cell}
metisrheaID: {485x1 cell}
metisseed_46__compoundID: {485x1 cell}
metisumbbd_46__compoundID: {485x1 cell}
metisunipathway_46__compoundID: {485x1 cell}
metisunipathway_46__reactionID: {485x1 cell}
b: [485x1 double]
csense: [485x1 char]
proteins: {339x1 cell}
geneNames: {339x1 cell}
rxnisbigg_46__reactionID: {554x1 cell}
rules: {554x1 cell}
osense: -1

```

ii. Verify and convert Tmodel

Because we are just using an arbitrary model as our template model, we need to verify it and convert it to a proper template model. You will be asked to confirm the name. Note that the final Tmodel, 'lb', 'ub', and 'Models' are structures containing information specific to each model.

```

% If Tmodel is just another model, verify it as well and convert it to a
% proper format for comparison. Also make sure it carries flux.
Tmodel = verifyModelBorg(Tmodel, 'keepName', 'Verbose');

```

```

Array .rxnID not in Model. Adding.
Array .subSystems not in Model. Adding.
Array .rxnECNumbers not in Model. Adding.
Array .rxnKEGGID not in Model. Adding.
Array .rxnSEEDID not in Model. Adding.
Array .rxnEquations not in Model. Adding.
Array .rxnReferences not in Model. Adding.
Array .rxnNotes not in Model. Adding.
Array .grRules not in Model. Adding.
Array .metID not in Model. Adding.
Array .metSEEDID not in Model. Adding.
Array .metPubChemID not in Model. Adding.
Array .metInChIString not in Model. Adding.
Array .metCharge not in Model. Adding.
Making sure reactions are all forwards

```

```
Fixing names of metabolites and reaction
Checking if reaction IDs (.rxns) are unique.
Checking if metabolite IDs (.mets) are unique.
All metabolites have compartment designation.
```

```
Tmodel = buildTmodel(Tmodel);
```

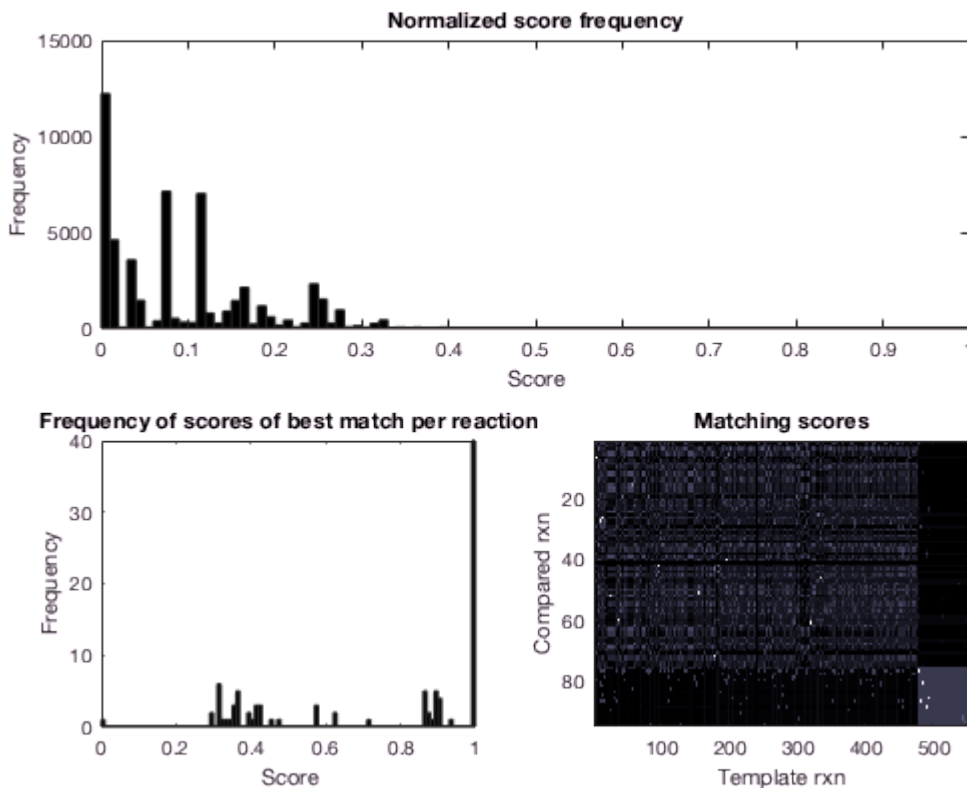
4. Compare models

compareCbModels scores all reactions in Cmodel against all reactions in Tmodel. It returns Score, a 3D matrix with size (# of reactions in Cmodel, # of reactions in Tmodel, # of scoring parameters per reaction). There are ~40 scoring parameters, such as name, E.C. number, metabolite similarity, and network topology. The returned Cmodel and Tmodel have some appendend information, but are functionally the same as the inputs. The structure Stats contains information about the best matches per each reaction.

Additionally, the function outputs some graphs describing the reaction scores. In particular the bottom right graph shows a reaction by reaction matrix of the scores. Lighter colors indicate a higher matching score between any two reactions. Note the transport reactions along the bottom and right of the graph.

```
[Cmodel, Tmodel, score, Stats] = compareCbModels(Cmodel, Tmodel, 'Verbose');
```

```
Adding comparison information time = 5.095442e-01.
Name match time = 2.733851e-01.
EC match time = 2.516028e-02.
Reaction KEGG ID match time = 7.361790e-04.
Reaction SEED ID match time = 5.262290e-04.
Subsystem match time = 7.466888e-02.
Metabolte number and stoich match time = 2.876957e-02.
Reaction compartment match time = 4.991152e-03.
Network topology match time = 3.724050e-01.
Met name match time = 4.133636e+01.
```



5. Match models

reactionCompare is the major step in the comparison process. It will launch a GUI that facilitates reaction-by-reaction comparison between Cmodel and Tmodel. This section will outline the different functions of the GUI.

Note you must run reactionCompare in the Command Window, as GUIs are not properly rendered within the Matlab Live script.

```
if ~exist('rxnList', 'var') || ~exist('metList', 'var') || ~exist('Stats', 'var')
    rxnList = [];
    metList = [];
    Stats = [];
end

% Initial comparison and matching.
% [rxnList, metList, Stats] = reactionCompare(Cmodel, Tmodel, score);

% Subsequent comparisons and matching.
% [rxnList, metList, Stats] = reactionCompare(Cmodel, Tmodel, score, rxnList, metList, Stats);
```

i. Comparing similarity of reactions. Reactions from Cmodel (Ecoli_core) are displayed 1-by-1 along with the best matches from Tmodel. Information about the current reaction (gapd, reaction #46) can be seen in the red box labeled 1. Information about the best match from Tmodel (gapd, reaction #335) is the blue box labeled 2.

The score of this reaction is indicated by the blue arrow. The subsequent best reactions are to the right (Match B).

reactionCompare
Ecoli_core to lIT341

Automatch Parameters: Low Margin High
Rxn 0.01 0.1 0.99
Mets 0.01 0.1 0.99

Weighting: Default Weights
linearOpt
expOpt
SVM
Random Forest

Minimum score to review: 0.9031

Matched Rxns: 1
New Rxns: 0
Need Review: 94
Reaction Number: 46
Current Match: None
Number of Matches: 2
C balance
Compartment
Stoichiometry
Choose Match
New Reaction

1%
99%

Populate Table
Next Undeclared Reaction

	new Rxn	Match A	Match B
Score; Rxn #	46	0.90323;335	0.35484;141
Reaction ID	gapd	gapd	ipdps
Reaction Name	glyceraldehyde-3-phosphate dehydrogenase	Glyceraldehyde-3-phosphate dehydrogenase	1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase
Equation	g3p[c] + nad[c] + pi[c] <=> 13dp[c] + h[c] + nadh[c]	nad[c] + pi[c] + g3p[c] <=> nadh[c] + h[c] + 13dp[c]	nadh[c] + h[c] + h2mb4p[c] --> h2o[c] + nad[c] + ipdp[c]
EC Number			
KEGG ID			
SEED ID			
Subsystem	Glycolysis/Gluconeogenesis		

	Metabolites of new Rxn	Match A	Match B
Reactant IDs	nad[c];pi[c];g3p[c]	pi[c];nad[c];g3p[c]	h[c];nadh[c];h2mb4p[c]
Reactant Names	nicotinamide_adenine_dinucleotide;phosphate;glyceraldehyde-3-phosphate;nicotinamide_adenine_dinucleotide;glyceraldehyde-3-phosphate	phosphate;nicotinamide_adenine_dinucleotide;glyceraldehyde-3-phosphate	h+;nicotinamide_adenine_dinucleotide_reduced;1_hydroxy_2_methyl_2_oxo_3_oxo_4_oxo_5_oxo_6_oxo_7_oxo_8_oxo_9_oxo_10_oxo_11_oxo_12_oxo_13_oxo_14_oxo_15_oxo_16_oxo_17_oxo_18_oxo_19_oxo_20_oxo_21_oxo_22_oxo_23_oxo_24_oxo_25_oxo_26_oxo_27_oxo_28_oxo_29_oxo_30_oxo_31_oxo_32_oxo_33_oxo_34_oxo_35_oxo_36_oxo_37_oxo_38_oxo_39_oxo_40_oxo_41_oxo_42_oxo_43_oxo_44_oxo_45_oxo_46_oxo_47_oxo_48_oxo_49_oxo_50_oxo_51_oxo_52_oxo_53_oxo_54_oxo_55_oxo_56_oxo_57_oxo_58_oxo_59_oxo_60_oxo_61_oxo_62_oxo_63_oxo_64_oxo_65_oxo_66_oxo_67_oxo_68_oxo_69_oxo_70_oxo_71_oxo_72_oxo_73_oxo_74_oxo_75_oxo_76_oxo_77_oxo_78_oxo_79_oxo_80_oxo_81_oxo_82_oxo_83_oxo_84_oxo_85_oxo_86_oxo_87_oxo_88_oxo_89_oxo_90_oxo_91_oxo_92_oxo_93_oxo_94_oxo_95_oxo_96_oxo_97_oxo_98_oxo_99_oxo_100_oxo
Reactant Formulas	C21H26N7O14P2;HO4P;C3H5O6P	HO4P;C21H26N7O14P2;C3H5O6P	H;C21H27N7O14P2;C5H9O8P2
Reactant Charges	0;0;0	0;0;0	0;0;0
Reactant KEGG IDs	::	D05467;D00002;C00661	C00080;C00004;C11811
Product IDs	h[c];nadh[c];13dp[c]	h[c];nadh[c];13dp[c]	h2o[c];nad[c];ipdp[c]
Product Names	h+;nicotinamide_adenine_dinucleotide_reduced;3_phosphoglycerate	h+;nicotinamide_adenine_dinucleotide_reduced;3_phosphoglycerate	h2o;nicotinamide_adenine_dinucleotide;isopentenyl_diphosphate
Product Formulas	H;C21H27N7O14P2;C3H4O10P2	H;C21H27N7O14P2;C3H4O10P2	H2O;C21H26N7O14P2;C5H9O7P2
Product Charges	0;0;0	0;0;0	0;0;0
Product KEGG IDs	::	C00080;C00004;C00236	D06322;D00002;C00129

All clear.

Review Metabolites
Finish Comparison

The development of this program was kindly sponsored by BRAIN Aktiengesellschaft.

ii. Choose a matching reaction or declare a new reaction. To pair a reaction from Cmodel to a reaction in Tmodel, put the reaction number of the match into the box and press "Choose Match" (indicated by the arrow and the blue box labeled 1). If there is no appropriate match then click "New Reaction". By clicking on any of the information in the match table (such as the highlighted reaction equation under Match A), the blocks in the red box labeled 2 will indicate if this reaction matches the current reaction from Cmodel in terms of carbon balance, compartment, and metabolite stoichiometry.

reactionCompare
Ecoli_core to iIT341

Automatch Parameters: Low Margin High Weighting: Default Weights
 Rns 0.01 0.1 0.99 linearOpt
 Mets 0.01 0.1 0.99 expOpt
 Automatch SVM Random Forest

Matched Rns: 1
 New Rns: 0
 Need Review: 94
 Reaction Number: 46
 Current Match: None
 Number of Matches: 2
 Match R: 335
 99%

Minimum score to review: 0.9031

2
 1
 Choose Match
 New Reaction

	new Rxn	Match A	Match B
Score: Rxn #	46	0.90323;335	0.35484;141
Reaction ID	gapd	gapd	ipdps
Reaction Name	glyceraldehyde-3-phosphate dehydrogenase	Glyceraldehyde-3-phosphate dehydrogenase	1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase
Equation	$g3p[c] + nad[c] + pi[c] \rightleftharpoons 13dp[g] + h[c] + nadh[c]$	$nad[c] + pi[c] + g3p[c] \rightleftharpoons nadh[c] + h[c] + 13dp[g]$	$nadh[c] + h[c] + h2mb4p[c] \rightarrow h2o[c] + nad[c] + ipdp[c]$
EC Number			
KEGG ID			
SEED ID			
Subsystem	Glycolysis/Gluconeogenesis		

	Metabolites of new Rxn	Match A	Match B
Reactant IDs	nad[c];pi[c];g3p[c]	pi[c];nad[c];g3p[c]	h[c];nadh[c];h2mb4p[c]
Reactant Names	nicotinamide_adenine_dinucleotide;phosphate;glyceraldehyde	phosphate;nicotinamide_adenine_dinucleotide;glyceraldehyde...	h+;nicotinamide_adenine_dinucleotide_reduced;1_hydroxy_2...
Reactant Formulas	C21H26N7O14P2;HO4P;C3H5O6P	HO4P;C21H26N7O14P2;C3H5O6P	H;C21H27N7O14P2;C5H9O8P2
Reactant Charges	0;0;0	0;0;0	0;0;0
Reactant KEGG IDs	::	D05467;D00002;C00661	C00080;C00004;C11811
Product IDs	h[c];nadh[c];13dp[g]	h[c];nadh[c];13dp[g]	h2o[c];nad[c];ipdp[c]
Product Names	h;nicotinamide_adenine_dinucleotide_reduced;3_phosph...	h+;nicotinamide_adenine_dinucleotide_reduced;3_phospho_d...	h2o;nicotinamide_adenine_dinucleotide;isopentenyl_diphosph...
Product Formulas	H;C21H27N7O14P2;C3H4O10P2	H;C21H27N7O14P2;C3H4O10P2	H2O;C21H26N7O14P2;C5H9O7P2
Product Charges	0;0;0	0;0;0	0;0;0
Product KEGG IDs	::	C00080;C00004;C00236	D06322;D00002;C00129

All clear. Review Metabolites Finish Comparison

The development of this program was kindly sponsored by BRAIN Aktiengesellschaft.

iii. Compare metabolites. When a reaction from Cmodel has been matched or declared as new, its metabolites are then reviewed in an analogous GUI. Choose the best matching metabolite from the table (Match A, Match B, ...), with the radio buttons in the red box labeled 1. You are only allowed to declare a new metabolite if the reaction itself was declared new. After all metabolites have been reviewed (in the blue box labeled 2), you can press the button "Add Metabolite(s)" (blue arrow), and resume comparing reactions. Alternatively, you can postpone matching metabolites by pressing "Skip Matching." However, matching metabolites facilitates determining the fate of as of yet unreviewed reactions.

metCompare

Ecoil_core to IIT341

Reactions Declared : 2 / 95
Metabolites Declared : 2 / 72
Reaction Number : 46
Reaction Name : gapd:glyceraldehyde-3-phosphate dehydrogenase
Reaction Equation : $g3p[c] + nad[c] + p[c] \rightleftharpoons 13dpg[c] + h[c] + nadh[c]$

Display **5** Matches

	New Metabolite	Match A	Match B	Match C
Score		0.9033	0.024176	0.024176
ShortName, ID #	h[c], 1	h[c], 1	p[c], 7	nad[c], 18
Compartment	c	c	c	c
LongName	h	h+	phosphate	nicotinamide_adenine_dinucleotide
Formula	H	H	HO4P	C21H26N7O14P2
Charge	0	0	0	0
KEGG ID		C00080	D05467	D00002
SEED ID				
Model ID	h[c]	IIT341:h[c]	IIT341:p[c]	IIT341:nad[c]

add info h[c], 1

Match Number : 335

New Equation : $1 nad[c] + 1 p[c] + 1 g3p[c] \rightarrow 1 h[c] + 1 nadh[c] + 1 13dpg[c]$

Unseen Mets : h[c], 1; nad[c], 7; nadh[c], 8; p[c], 9; g3p[c], 21; 13dpg[c], 51;

Review required. Skip Matching Add Metabolite(s)

Choose Best Match

- ☒ Match A
- ☐ Match B
- ☐ Match C
- ☐ Match D
- ☐ Match E
- ☐ Create New Met
- ☐ Other Met Met #

Choose Match

iv. Finish/pause comparison. You can quit comparison and save your work at any time by pressing "Finish Comparison" in the red box labeled 1. The number of reactions which have been reviewed, matched, or declared new is located in blue box labeled 2. Finishing comparison produces two arrays, "rxnList" and "metList," which indicated to which reaction or metabolite in Tmodel matches a reaction or metabolite in Cmodel, respectively. New reactions are given a designation "-1," while new metabolites are given new metabolite numbers immediately (such numbers will be higher than the total number of metabolites in Tmodel). Unreviewed reactions and metabolites have the designation "0."

When you resume comparison, give rxnList and metList as arguments to reactionCompare (see above).

reactionCompare
Ecoli_core to iIT341

Automatch Parameters: Low Margin High
Rxn 0.01 0.1 0.99
Mets 0.01 0.1 0.99
Automatch

Weighting: Default Weights
linearOpt
expOpt
SVM
Random Forest

Minimum score to review: 0.89147

Matched Rxns: 39
New Rxns: 5
Need Review: 51
Reaction Number: 53

Current Match: None
Number of Matches: 2
C balance
Compartment
Stoichiometry
Match Rxn #:
Choose Match
New Reaction

Populate Table
Next Undeclared Reaction

	new Rxn	Match A	Match B
Score; Rxn #	53	0.32258;59	0.32258;166
Reaction ID	acontb	fum	h2co3d
Reaction Name	aconitase (half-reaction B, Isocitrate hydro-lyase)	Fumarase	H2CO3D
Equation	acon_c[c] + h2o[c] <=> icit[c]	h2o[c] + fum[c] <=> mal_l[c]	h2o[c] + co2[c] <=> h2co3[c]
EC Number			
KEGG ID			
SEED ID			
Subsystem	Citric Acid Cycle		

	Metabolites of new Rxn	Match A	Match B
Reactant IDs	h2o[c];acon_c[c]	h2o[c];fum[c]	h2o[c];co2[c]
Reactant Names	h2o;cis_aconitate	h2o;fumarate	h2o;co2
Reactant Formulas	H2O;C6H3O6	H2O;C4H2O4	H2O;CO2
Reactant Charges	0;0	0;0	0;0
Reactant KEGG IDs	:	D06322;D02308	D06322;D00004
Product IDs	icit[c]	mal_l[c]	h2co3[c]
Product Names	isocitrate	L_malate	carbonic_acid
Product Formulas	C6H5O7	C4H4O5	H2CO3
Product Charges	0	0	0
Product KEGG IDs		C00149	

All Clear Review Metabolites **1** Finish Comparison

The development of this program was kindly sponsored by BRAIN Aktiengesellschaft.

v. Review additional reactitons. You can control which reactions you review in two ways. You can review any arbitrary reaction from Cmodel by putting in the reaction number and pressing "Populate Table," indicated by the red box labeled 1. Alternatively, you can press the button "Next Undeclared Reaction," which will go to the next undeclared reaction in Cmodel with the lowest reaction number (blue box labeled 2). You can require that reactions presented by "Next Undeclared Reaction" have at least one match above a give score by moving the slider indicated by the arrow.

reactionCompare
Ecoli_core to IIT341

Automatch Parameters:

Low Margin High
Rxn 0.01 0.1 0.99
Mets 0.01 0.1 0.99

Weighting:

Default Weights
linearOpt
expOpt
SVM
Random Forest

Minimum score to review: 0.90698

Matched Rxns: 39
New Rxns: 5
Need Review: 51

Reaction Number: 19
Current Match: None
Number of Matches: 2

Populate Table
Next Undeclared Reaction

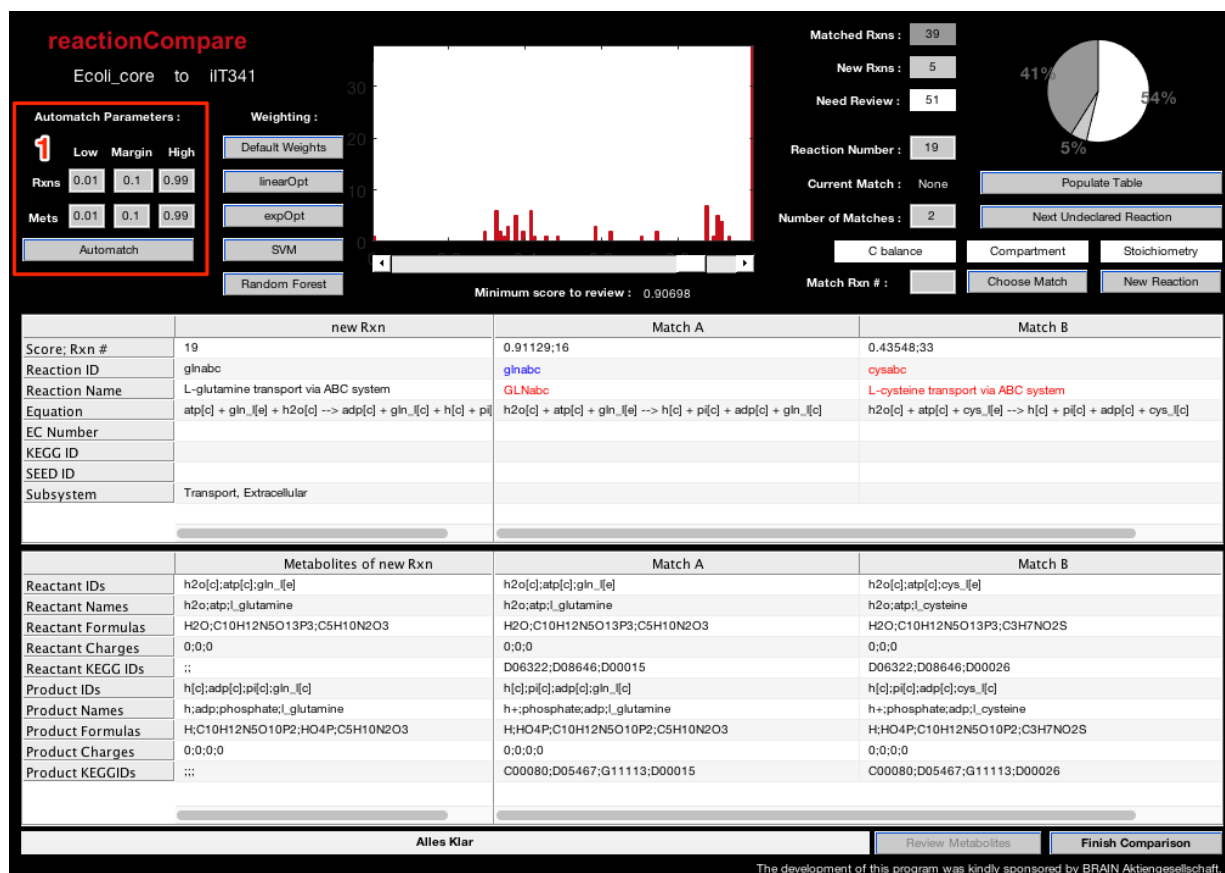
	new Rxn	Match A	Match B
Score: Rxn #	19	0.91129;16	0.43548;33
Reaction ID	glnabc	glnabc	cysabc
Reaction Name	L-glutamine transport via ABC system	GLNabc	L-cysteine transport via ABC system
Equation	atp[c] + gln_l[e] + h2o[c] --> adp[c] + gln_l[c] + h[c] + pi[c]	h2o[c] + atp[c] + gln_l[e] --> h[c] + pi[c] + adp[c] + gln_l[c]	h2o[c] + atp[c] + cys_l[e] --> h[c] + pi[c] + adp[c] + cys_l[c]
EC Number			
KEGG ID			
SEED ID			
Subsystem	Transport, Extracellular		

	Metabolites of new Rxn	Match A	Match B
Reactant IDs	h2o[c];atp[c];gln_l[e]	h2o[c];atp[c];gln_l[e]	h2o[c];atp[c];cys_l[e]
Reactant Names	h2o;atp;l glutamine	h2o;atp;l glutamine	h2o;atp;l cysteine
Reactant Formulas	H2O;C10H12N5O13P3;C5H10N2O3	H2O;C10H12N5O13P3;C5H10N2O3	H2O;C10H12N5O13P3;C3H7NO2S
Reactant Charges	0;0;0	0;0;0	0;0;0
Reactant KEGG IDs	::	D06322;D08646;D00015	D06322;D08646;D00026
Product IDs	h[c];adp[c];pi[c];gln_l[c]	h[c];pi[c];adp[c];gln_l[c]	h[c];pi[c];adp[c];cys_l[c]
Product Names	h;adp;phosphate;l glutamine	h+;phosphate;adp;l glutamine	h+;phosphate;adp;l cysteine
Product Formulas	H;C10H12N5O10P2;HO4P;C5H10N2O3	H;HO4P;C10H12N5O10P2;C5H10N2O3	H;HO4P;C10H12N5O10P2;C3H7NO2S
Product Charges	0;0;0;0	0;0;0;0	0;0;0;0
Product KEGG IDs	:::	C00080;D05467;G11113;D00015	C00080;D05467;G11113;D00026

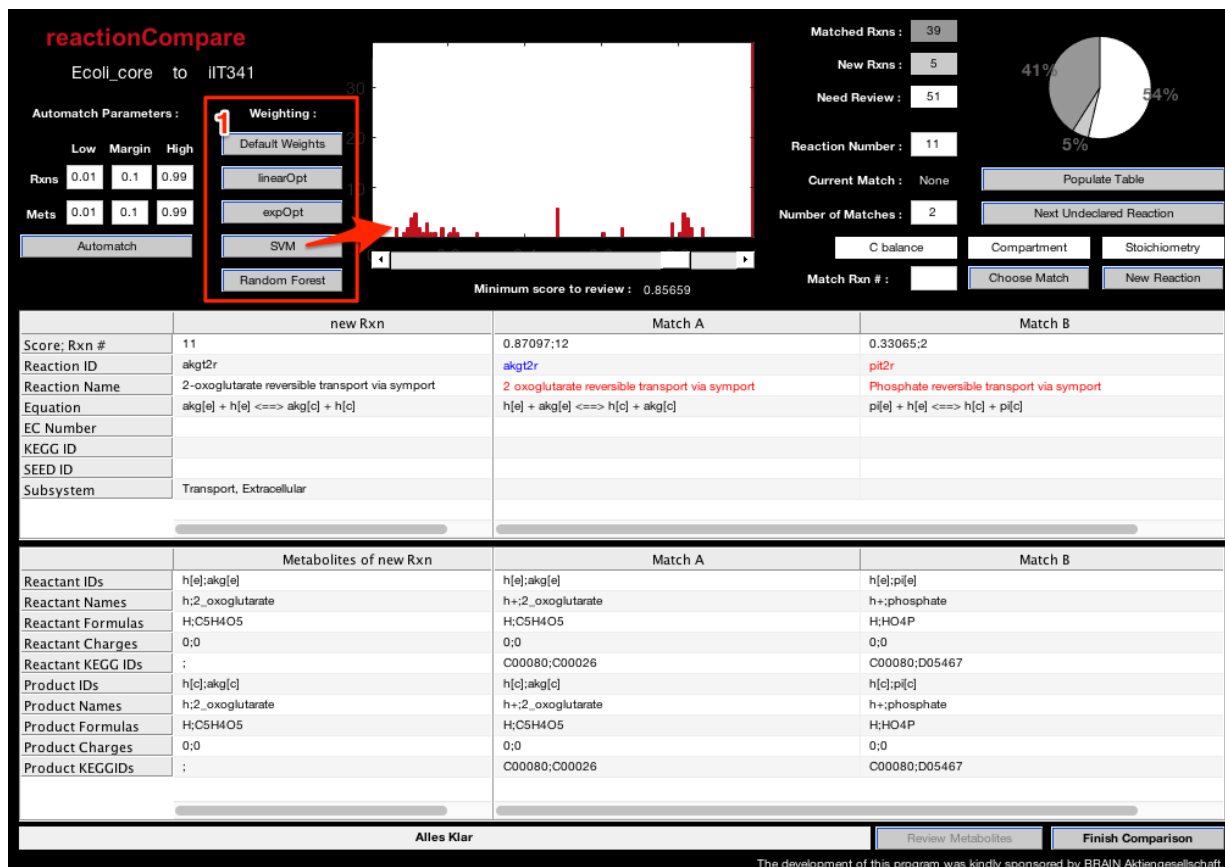
Alles Klar Review Metabolites Finish Comparison

The development of this program was kindly sponsored by BRAIN Aktiengesellschaft.

vi. Automatch reactions and metabolites. High and low scoring reactions may be safely matched or declared new, respectively. This is done with the options in the red box labeled 1. Reactions or metabolites above the score in the box "High" will be matched with their best match from Tmodel, as long as the score of the best match from Tmodel is better than the second best match by the value in "Margin." Reactions and metabolites whose matches are not above the score in "Low" will be declared as new. When a metabolite is declared as new, then all reactions in Cmodel containing that metabolite are also declared new.



vii. **Score weighting.** Once you have manually compared some reactions, this information can be used to determine which scoring parameters are most informative and can be weighted accordingly. Four weighting functions/algorithms are provided in addition to the default weighting scheme (red box labeled 1). There is a linear optimization, an exponential optimization, a support vector machine (SVM) learning method, and a random forest algorithm. See the script "optimalScores.m" for more information. The image below shows the scores under SVM weighing. Automatching should be used in conjunction with score weighting to reduce manual intervention.



6. Merge models

mergeModelsBorg will combine Cmodel into Tmodel and into a composite model and return it as TmodelC. It will iteratively check the fidelity of the merging and will prompt the user if errors are found. It will also produce a copy of Cmodel which has been extracted from TmodelC (see next step).

Merge models and test results.

```
if ~isempty(rxnList) && ~isempty(metList) && ~isempty(Stats)
    [TmodelC, Cspawn, Stats] = mergeModelsBorg(Cmodel, Tmodel, rxnList,
metList, Stats, 'Verbose');
end
```

Problems within metList, resolve with GUI.
 Skipped resolving, will not check fidelity of matrices.
 Extracting Ecoli_core from Tmodel
 Removing empty cell arrays:
 rxnKEGGID
 rxnSEEDID
 rxnReferences
 metSEEDID
 metPubChemID
 metInChIString
 merging pyrt2(iIT341) and pyrt2r(Ecoli_core)
 merging ex_gal_e(iIT341) and ex_fru_e(Ecoli_core)
 merging ex_lac_l_e(iIT341) and ex_lac_d_e(Ecoli_core)

```

Checking if reaction IDs (.rxns) are unique.
Checking if metabolite IDs (.mets) are unique.
Extracting Ecoli_core from Tmodel
Removing empty cell arrays:
metSEEDID
metPubChemID
metInChIString

```

The structure Stats contains information about the number of unique and shared metabolites between the models, as well as the completeness of annotations.

```

if ~isempty(rxnList) && ~isempty(metList) && ~isempty(Stats)
    % Shared reaction between the models. Values along the diagonal how many
    % reactions in the model are unique.
    Stats.sharedRxns
    % Shared metabolites between models,
    Stats.sharedMets
    Stats.sharedMetsNoComp % does not consider differences in compartment.

    modelNames = fieldnames(TmodelC.Models) ;
    for iM = 2:length(modelNames)+1
        fprintf([modelNames{iM-1}, ' has ', num2str(Stats.sharedRxns{iM,iM}
(1)), ' unique reactions.', ...
                ' (', num2str(Stats.sharedRxns{iM,iM}(2)), ' percent
of ', ...
                num2str(sum(TmodelC.Models.(modelNames{iM-1}).rxns)),
' reactions).\n'])
        fprintf([modelNames{iM-1}, ' has ', num2str(Stats.sharedMets{iM,iM}
(1)), ' unique metabolites.', ...
                ' (', num2str(Stats.sharedMets{iM,iM}(2)), ' percent of ', ...
                num2str(sum(TmodelC.Models.(modelNames{iM-1}).mets)), '
metabolites).\n'])
        fprintf([modelNames{iM-1}, ' has ',
num2str(Stats.sharedMetsNoComp{iM,iM}(1)), ...
                ' unique metabolites when not considering compartment.\n'])
    end
    fprintf([modelNames{1}, ' shares ', num2str(Stats.sharedRxns{2, 3}(1)),
' reactions with ', ...
            modelNames{2}, '.\n'])
    fprintf([modelNames{1}, ' shares ', num2str(Stats.sharedMets{2, 3}(1)),
' metabolites with ', ...
            modelNames{2}, '.\n'])
    fprintf([modelNames{1}, ' shares ', num2str(Stats.sharedMetsNoComp{2, 3}
(1)), ' metabolites with ', ...
            modelNames{2}, ' when not considering compartment.\n'])
end

```

```
ans =
```

```

    '[Count, %]'    'in iIT341'    'in Ecoli_core'
    'iIT341'        [1x2 double]    [1x2 double]
    'Ecoli_core'    [1x2 double]    [1x2 double]
ans =
    '[Count, %]'    'in iIT341'    'in Ecoli_core'
    'iIT341'        [1x2 double]    [1x2 double]
    'Ecoli_core'    [1x2 double]    [1x2 double]
ans =
    '[Count, %]'    'in iIT341'    'in Ecoli_core'
    'iIT341'        [1x2 double]    [1x2 double]
    'Ecoli_core'    [1x2 double]    [1x2 double]
iIT341 has 493 unique reactions. (0.88989 percent of 554 reactions).
iIT341 has 418 unique metabolites. (0.86186 percent of 485 metabolites).
iIT341 has 362 unique metabolites when not considering compartment.
Ecoli_core has 34 unique reactions. (0.35789 percent of 95 reactions).
Ecoli_core has 5 unique metabolites. (0.069444 percent of 72 metabolites).
Ecoli_core has 4 unique metabolites when not considering compartment.
iIT341 shares 61 reactions with Ecoli_core.
iIT341 shares 67 metabolites with Ecoli_core.
iIT341 shares 50 metabolites with Ecoli_core when not considering compartment.

```

7. Extract a model

A model can be extracted from the combined model with the function `readCbTmodel` and referencing its name. Extracted models should be mathematically identical to the model that went in, but will contain additional annotation information garnered from the comparison. For example, the extracted `Ecoli_core` model now contains KEGG IDs for its metabolites.

```

%% Extract both models
if ~isempty(rxnList) && ~isempty(metList) && ~isempty(Stats)
    Ecoli_core = readCbTmodel('Ecoli_core', TmodelC, 'Verbose');
    iIT341 = readCbTmodel('iIT341', TmodelC, 'Verbose');
end

```

```

Extracting Ecoli_core from Tmodel
Removing empty cell arrays:
rxnKEGGID
rxnSEEDID
rxnReferences
metSEEDID
metPubChemID
metInChIString
Extracting iIT341 from Tmodel
Removing empty cell arrays:
rxnKEGGID
rxnSEEDID
rxnReferences
grRules
metSEEDID
metPubChemID
metInChIString

```

8. Save work

Finally, you should save your combined model to be used for future comparison. Subsequent comparisons become easier as a `Tmodel` gains information.

```
% save([filesep 'Tmodel_' datestr(now,'yyyy.mm.dd') '.mat'], 'TmodelC')
```

REFERENCES

1. Sauls, J. T., & Buescher, J. M. (2014). Assimilating genome-scale metabolic reconstructions with modelBorgifier. *Bioinformatics* (Oxford, England), 30(7), 1036–8. <http://doi.org/10.1093/bioinformatics/btt747>
2. Thiele, I., Vo, T. D., Price, N. D., & Palsson, B. Ø. (2005). Expanded metabolic reconstruction of *Helicobacter pylori* (J1T341 GSM/GPR): an *in silico* genome-scale characterization of single- and double-deletion mutants. *Journal of Bacteriology*, 187(16), 5818–5830. <http://doi.org/10.1128/JB.187.16.5818>