Supplementary material:

An environmental perspective on metabolism

October 25, 2007

1 Data import

We have extracted the metabolic networks for 447 organisms the KEGG database (as of Feb 13, 2007). The organisms have been selected in the following way: It has been verified that the number of reactions was realistic compared to similar organisms, if available. Otherwise, when the number of reactions seemed abnormally low, the original genome sequence paper was checked to verify that the low number is in line with biological knowledge, for example in case of a low number of genes, a metabolic deficiency and/or parasitism.

The corresponding metabolic networks have been extracted as follows. First, from the LIGAND subdivision (plain text file), the complete list of 6825 reactions has been imported. The reactions have been checked for consistency. We rejected 290 reactions because they showed an erroneous stoichiometry, by which we mean that some atomic species occurred in different numbers on both sides of the reaction. Further, we did not include 342 reactions involved in glycan synthesis because the focus of our investigation lies on the metabolism of small chemical species and does not include macromolecular syntheses.

Information on the reversibility of reactions has been extracted from the KGML files which specify the pathways for all organisms included in KEGG. In general, a particular reaction is listed in several KGML files and the information on its reversibility may be ambiguous. In fact, we identified 136 reactions for which this is the case. For the present

calculations, we consider a reaction to be irreversible only if it is defined as irreversible in all corresponding occurrences in the KGML files. This is the case for 2622 reactions.

The organism specific networks were determined using the 'reaction' and 'enzyme' files from the KEGG/LIGAND database. In a first step, for all reactions the EC numbers of the catalyzing enzymes were retrieved from their corresponding entries in the 'reaction' file (section ENZYME). Subsequently, from the 'enzyme' file, for each enzyme a list of organisms is obtained in which there exists a corresponding gene (section GENES). Thus, for each organism the metabolic network is defined by all those reactions for which a catalyzing enzyme is encoded in its genome. In all cases where an enzyme is not fully classified (e.g. EC1.3.1.-), the corresponding entry in the 'enzyme' file contains no GENES section. As a consequence, no such reactions are included in organism specific networks.

Further, the KO section of the database is inspected. Reactions specified in the DBLINKS/RN section of a KO entry are also assigned to the set of reactions of the organisms listed in the GENES section of this entry.

2 Target metabolites

We define a set of metabolites, called the target set T, that are required for the synthesis of compounds necessary to sustain cellular maintainance or growth. In our analysis this set contains all chemical compounds that are present in at least 90 percent of the organism specific networks. The set contains amino acids, nucleotides, cofactors, organic acids and phosphorylated sugars. From this list, the following metabolites have been removed because they usually require themselves or a related compound in the seed: IDP, ITP and dCMP due to missing phosphorylation or dephosphorylation reactions; 'Glutamate', as it is a generic version of L-Glutamate; Mercaptopyruvate, as it requires toxic compounds for it synthesis; 4-Trimethylammoniobutanal and 3-Hydroxy-N6,N6,N6-trimethyl-L-Lysine as they are part of the degradation of protein bound Lysine that cannot be produced with the present reactions; 3-Phospho-D-erythronate and Phosphoenol-4-deoxy-3-tetrulosonate, as they are separated from the remaining network; Selenomethionine and Se-Adenosylselenomethionine as they are isolated in most cases; 2-Deoxy-5-keto-D-gluconic acid 6-phosphate which is part of the poorly annotated inos-

itol degradation pathway. Further we removed from the list all compounds containing variable elements such as residues (indicated by an "R" in the sum formula) or multiple chain elements because it is not assured whether these can be synthesized from regular compounds by the defined reaction set. The complete list of compounds within the target set is given in Table 1.

3 Preferred seed compounds

For each organism, we determine minimal sets of seed compounds from which all metabolites of the target set T can be produced. The algorithm allows to provide a set of compounds that are preferably used as seed compounds. In this work we assumed that small organic and inorganic precursors can be taken up by the cell and used as nutrients. Further, we assumed that compounds transported by ABC transporters or the Phosphotransferase system, as defined by the KEGG pathways KO02010 and KO02060, respectively, can be utilized by metabolism. Table 2 shows a list of the preferred seed compounds.

4 Compound clusters and resource types

The algorithm predicts clusters of exchangeable seed compounds for each organism. These clusters are provided as supplementary files for a few example organisms listed in Table 3. Organism specific clusters are compiled into general clusters, containing seed compounds that are exchangeable in most organisms. In general these clusters contain classes of metabolites such as sugars, amino acids or vitamins, clustered with their respective precursors or derivatives. The supplementary file "global_clusters.eps" shows these automatically determined clusters. The nutrients sets are further processed manually. If compounds in a cluster are obviously direct precursors of compounds in another cluster these two are united. The resulting sets of seed compounds are called 'resource types' and labelled according to their most familiar members. Table 4 shows a list of the defined nutrient types and their member metabolites.

Acetate	Acetyl-CoA
Adenine	Adenosine
ADP	alpha-D-Glucose 6-phosphate
AMP	ATP
beta-D-Fructose 1,6-bisphosphate	beta-D-Fructose 6-phosphate
beta-D-Glucose 6-phosphate	Biotin
Carbamoyl phosphate	CDP
CMP	CO2
CoA	CTP
D-Erythrose 4-phosphate	D-Fructose 1,6-bisphosphate
D-Fructose 1-phosphate	D-Fructose 6-phosphate
D-Glucosamine 6-phosphate	D-Glucose 1-phosphate
D-Glucose 6-phosphate	D-Glyceraldehyde
D-Mannose 6-phosphate	D-Ribose 5-phosphate
D-Ribulose 5-phosphate	D-Sedoheptulose 7-phosphate
D-Xylulose 5-phosphate	dADP
dAMP	dATP
dCDP	dCTP
Deamino-NAD+	Deoxyadenosine
Deoxyguanosine	Dephospho-CoA
dGDP	dGMP
dGTP	Dihydrofolate
Dihydropteroate	Dimethylallyl diphosphate
dTDP	dTMP
dTTP	dUDP
dUMP	dUTP
FAD	FADH2
Formate	Fumarate
GDP	Glycerone phosphate
Glycine	GMP
GTP	Guanine
H+	H2O
H2O2	HCO3-
IMP	Inosine
Isopentenyl diphosphate	L-Alanine
L-Arginine	L-Asparagine
L-Aspartate	L-Cysteine
L-Glutamate	L-Glutamine
L-Histidine	L-Isoleucine
L-Leucine	L-Lysine
L-Methionine	L-Ornithine
L-Phenylalanine	L-Proline
L-Serine	L-Threonine
L-Tryptophan	L-Tyrosine
L-Valine	N6-(1,2-Dicarboxyethyl)-AMP
NAD+	NADH
NADP+	NADPH
NH3	Nicotinamide D-ribonucleotide
Nicotinate D-ribonucleotide	Orthophosphate
Oxaloacetate	Oxygen
Phenylpyruvate	Phosphoenolpyruvate
Propanoyl-CoA	Pyrophosphate
Pyruvate	S-Adenosyl-L-homocysteine
S-Adenosyl-L-methionine	Sedoheptulose 1,7-bisphosphate
Sedoheptulose 7-phosphate	sn-Glycerol 3-phosphate
Succinate	Tetrahydrofolate
Thiamin diphosphate	UDP
UDP-N-acetyl-D-glucosamine	UMP
UTP	Xanthosine 5'-phosphate
(2R)-2-Hydroxy-3-(phosphonooxy)-propanal	(S)-Malate
1-(5'-Phosphoribosyl)-5-amino-4-(N-succinocarboxamide)-imidazole	1-(5'-Phosphoribosyl)-5-amino-4-imidazolecarboxamide
10-Formyltetrahydrofolate	2,3-Bisphospho-D-glycerate
2-(alpha-Hydroxyethyl)thiamine diphosphate	2-Oxobutanoate
2-Oxoglutarate	2-Phospho-D-glycerate
3-(4-Hydroxyphenyl)pyruvate	3-Dehydroquinate
3-Dehydroshikimate	3-Methyl-2-oxobutanoic acid
3-Oxopropanoate	3-Phospho-D-glycerate
3-Phospho-D-glyceroyl phosphate	5,10-Methenyltetrahydrofolate
5,10-Methylenetetrahydrofolate	5-Phospho-alpha-D-ribose 1-diphosphate

Table 1: List of compounds in the target set, alphabetically sorted.

ABC transported	PTS transported	Small metabolites
Betaine	alpha,alpha-Trehalose	Acetamide
Butyro-betaine	alpha-D-Glucose	Acetate
Carnitine	Arbutin	Allyl alcohol
Choline	Ascorbate	Carbamate
Choline sulfate	beta-D-Glucose	Carbonic acid
Cobalt	beta-D-Glucoside	Chloride
Crotono-betaine	Cellobiose	Cl-
Cyclomaltodextrin	D-Fructose	CO2
D-Allose	D-Glucosamine	Cobalt
D-Aspartate	D-Glucose	Dimethylamine
D-Galactose	D-Sorbitol	Ethanol
D-Glucose	Galactitol	Ethanolamine
D-Methionine	Glucose	Ethylamine
D-Ribose	Lactose	Fe2+
D-Xylose	Maltose	Fe3+
Fe(III)dicitrate	Mannitol	Formate
Fe(III)hydroxamate	N-Acetyl-D-glucosamine	Glycine
Fe-enterobactin	N-Acetylgalactosamine	Glycolate
Fe2+	Nitrogen	H+
Fe3+	Salicin	H2O
Ferrichrome	Sorbose	HCO3-
Heme	Sucrose	HO-
Hemine		Imidazole
Iron chelate		Iron
L-Arabinose		Magnesium
L-Arginine		Manganese
L-Aspartate		Methane
L-Glutamate		Methanol
L-Glutamine		Methylguanidine
L-Histidine		NH3
L-Isoleucine		Nitrate
L-Leucine		Nitric oxide
L-Lysine		Nitrite
L-Methionine		Nitrogen
L-Ornithine		Nitrous oxide
L-Proline		Oxygen
L-Threonine		Propan-2-ol
L-Valine		Propane-1-ol
Maltose		Propanoate
Manganese		Sulfur
Molybdate		Trimethylamine N-oxide
Nickel		Urea
Nitrate		(R)-1-Aminopropan-2-ol
Orthophosphate		1,3-Diaminopropane
Putrescine		1-Aminopropan-2-ol
sn-Glycerol 3-phosphate		1-Butanol
Sodium		
Spermidine		
Sulfate		
Taurine		
Teichoic acid		
Tetrabenazine		
Thiamin		
Thiosulfate		
Tungsten		
Urea		
Vitamin B12		
Zinc		
2,6-Dimethoxybenzoquinone		
2-(beta-D-Glucosyl)-sn-glycerol		

Table 2: List of compounds preferably used as seed compounds, sorted alphabetically

Organism	file name
Buchnera aphidicola	seedcluster_BUC.eps
Escheria coli	seedcluster_ECO.eps
Mycoplasma mobile	seedcluster_MMO.eps
Wolbachia	seedcluster_WOL.eps
Chlamydia pneumoniae	seedcluster_CPJ.eps
Homo sapiens	seedcluster_HSA.eps
Saccharomyces cerevisiae	seedcluster_SCE.eps
Rickettsia prowazekii	seedcluster_RPR.eps
Tropheryma whipplei	seedcluster_TWH.eps

Table 3: List of organisms for which metabolic resource graphs are provided.

5 Comparison to Metagrowth

The results of our analysis have been compared to the information of the Metagrowth database for the examples *Rickettsia prowazekii* (table 5) and *Tropheryma whipplei* (table 6). For that, metabolites for which a deficiency in their synthesis pathways has been noted in Metagrowth and seeds that have been found as essential in our calculations have been listed. The essential seed compounds have been taken from the organism specific resource graphs which can be found in the supplementary files "seedcluster_RPR.eps" and "seedcluster_TWH.eps". The lists show, that in 55% (*R. prowazekii*) and 78% (*T. whipplei*) of the cases our results and the Metagrowth data predict the same necessary metabolites. Certain metabolites were not predicted by our algorithm because they were either not necessary for synthesis of any compound in the target set or absent from the organism's metabolic network as dervied from KEGG.

Resource type	Member metabolites
Acetate	Acetate
Bases	Uracil, Adenine, Xanthine, Hypoxanthine, Thymidine, Guanine, Inosine
Biotin	Biotin, 7,8-Diaminononanoate, Dethiobiotin, 8-Amino-7-oxononanoate, 6-Carboxyhexanoate
CO2	HCO3-
D-Fructose	D-Fructose
D-Glucose	D-Glucose, beta-D-Glucose, alpha-D-Glucose
D-Mannose	D-Mannose
D-Ribose	D-Ribose
Folate	Tetrahydrofolate, 10-Formyltetrahydrofolate, Folate, 5,10-Methylenetetrahydrofolate, 5,10-Methenyltetrahydrofolate, 5-Formyltetrahydrofolate, 5-Methyltetrahydrofolate, Dihydrofolate, 4-Aminobenzoate, 2-Amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine, 2-Amino-4-hydroxy-6-(D-erythro-1,2,3-trihydroxypropyl)-7,8-dihydropteridine
Glycerol	Glycerol, D-Glyceraldehyde, D-Glycerate
Glycine	Glycine
Isopentenyl	Isopentenyl diphosphate, Dimethylallyl diphosphate, 1-Hydroxy-2-methyl-2-butenyl 4-diphosphate, 2-C-Methyl-D-erythritol 2,4-cyclodiphosphate
L-Alanine	L-Alanine, D-Alanine, Selenocysteine, Selenide, L-Selenocysteine
L-Arginine	L-Arginine
L-Asparagine	L-Asparagine
L-Aspartate	L-Aspartate
L-Cysteine	L-Cysteine
L-Glutamate	L-Glutamate, L-Glutamine
L-Histidine	L-Histidine, Urocanate, L-Histidinal, L-Histidinol
L-Isoleucine	L-Isoleucine, (S)-3-Methyl-2-oxopentanoic acid
L-Leucine	L-Leucine, 4-Methyl-2-oxopentanoate
L-Lysine	L-Lysine
L-Methionine	L-Methionine, Methanethiol
L-Ornithine	L-Ornithine
L-Phenylalanine	L-Phenylalanine, Phenylpyruvate, L-Arogenate, D-Phenylalanine, Prephenate
L-Proline	L-Proline
L-Serine	L-Serine
L-Threonine	L-Threonine
L-Tryptophan	L-Tryptophan, Indole
L-Tyrosine	L-Tyrosine, 3-(4-Hydroxyphenyl)pyruvate
L-Valine	L-Valine
Maltose	Maltose
Nicotinate	NAD+, NADH, NADPH, NADP+, Nicotinamide, Nicotinate, Nicotinamide D-ribonucleotide, Deamino-NAD+, Nicotinate D-ribonucleotide, Pyridine-2,3-dicarboxylate, Nicotinate D-ribonucleoside
Nucleoside	Adenosine, Deoxyguanosine, Deoxyadenosine, Deoxyinosine
OrganicAcids	2-Oxobutanoate, D-erythro-3-Methylmalate, 2-Hydroxybutanoic acid, 3-Methyl-2-oxobutanoic acid, 3-Hydroxy-3-methyl-2-oxobutanoic acid, (S)-2-Acetolactate, 2,3-Dihydroxy-3-methylbutanoate, (R)-2,3-Dihydroxy-3-methylbutanoate, 2-Acetolactate
Orthophosphate	Orthophosphate, sn-Glycerol 3-phosphate
Pantetheine	Pantetheine, Pantothenate, Dephospho-CoA, Dihydropteroate, Pantetheine 4'-phosphate, D-4'-Phosphopantothenate, (R)-4'-Phosphopantothenoyl-L-cysteine
Propanoate	Propanoyl-CoA, Propanoate, L-3-Amino-isobutanoate, 3-Hydroxypropanoate, (S)-Methylmalonate semialdehyde, (S)-3-Hydroxyisobutyrate, 3-Hydroxy-2-methylpropanoate, 2-Methyl-3-oxopropanoate, Propanoyl phosphate
Riboflavin	FAD, FADH2, FMN, Riboflavin, 6,7-Dimethyl-8-(1-D-ribityl)lumazine, 4-(1-D-Ribitylamino)-5-amino-2,6-dihydroxypyrimidine
Shikimate	3,4-Dihydroxybenzoate, 3-Dehydroquinate, Shikimate, 3-Dehydroshikimate, 5-Dehydroshikimate
Succinate	Succinate, Succinate semialdehyde, Fumarate
Sucrose	Sucrose
Sulfate	Sulfate, Hydrogen sulfide, Thiosulfate, Sulfite, Taurine
Thiamin	Thiamin diphosphate, alpha,beta-Dihydroxyethyl-TPP, 2-(alpha-Hydroxyethyl)thiamine diphosphate, Thiamin, Thiamin monophosphate, 5-(2-Hydroxyethyl)-4-methylthiazole, 4-Methyl-5-(2-phosphoethyl)-thiazole, 4-Amino-5-hydroxymethyl-2-methylpyrimidine, 4-Amino-2-methyl-5-phosphomethylpyrimidine, 2-Methyl-4-amino-5-hydroxymethylpyrimidine diphosphate
beta-Alanine	beta-Alanine, beta-Aminopropion aldehyde, 3-Oxopropanoate

Table 4: List of resource types and their equivalent compounds.

Compounds	Metagrowth	seed calculation	remarks
Amino acid	+	+	our method showed several aminoacids as essential
AMP/dAMP	+	+	
Biotin	+	+	
CoA	+	+	
FAD/Riboflavin-5-phosphate	+	+	
Glutathione	+	-	not in target set
GMP/dGMP	+	+	
HCO3	-	+	
Heme/Protoporphyrin	+	-	not in target set
Isopentenyl diphosphate	-	+	
Malonyl-CoA	+	-	not in target set
Mannose 6 phosphate	-	+	
N-Acetyl-D-mannosamine	-	+	
NAD+	+	+	
Phosphatidate/sn-Glycerol 3-phosphate/Fatty acid	+	o	sn-Glycerol 3-phosphate is the initial compound for the production of Phosphatidate and Fatty acid. Both can neither be produced nor are in the target set.
Propanoate	-	+	
Pyridoxine/Pyridoxal phosphate	+	-	not in target set
Pyruvate	+	+	
Ribose 5 phosphate	-	+	
S-Adenosyl-L-methionine	+	+	
Tetrahydrofolate/Folic acid/Dihydropteroate	+	+	
Thiamin diphosphate/Thiamin (vitamin B1)	+	+	
Thymidine/CMP/UMP	+	+	
Ubiquinone (Coenzyme Q)/Chorismate	+	-	not in target set

Table 5: Comparison of Metagrowth data to the calculations performed in this paper for *Rickettsia prowazekii* (RPR): "+" indicates a metabolite deficiency in Metagrowth or an essential seed compound in the calculations, "-" indicates a compound not identified as an essential nutrient and "o" indicates a potential, non essential nutrient.

Compounds	Metagrowth	seed calculation	remarks
1-Acyl-sn-glycerol 3-phosphatePhosphatidate	+	+	glycerol and sn-glycerol 3 phosphate predicted by seed calculation
beta-Alanine	+	+	
Biotin	+	+	
D-Glucose	o	+	not defined as deficiency in Metagrowth, but marked as primary energy source
D-Glutamate	+	o	generic compound "Glutamate" (i.e. D- or L-Glutamtate) found in seed calculation
Heme/Protoporphyrin	+	-	not in target set
L-Arginine	+	+	
L-Asparagine	+	+	
L-Aspartate	-	+	
L-Cysteine	+	o	
L-Glutamate	+	o	
L-Glutamine	+	o	
L-Histidine	+	+	
L-Leucine	+	+	
L-Lysine	+	+	
L-Methionine	+	+	
L-Phenylalanine	+	+	
L-Proline	+	+	
L-Tryptophan	+	+	
Malonyl-CoA	+	-	not in target set
Nicotinamide/Nicotinate	+	+	found in 2 different clusters in seed calculation
Propanoyl-Coa	-	+	
Pyridoxal phosphate/Pyridoxine	+	-	not in target set
(R)-Pantoate	+	+	Pantetheine found in seed calculation
Succinate	-	+	
Sugar phosphates	+	+	
Tetrahydrofolate	+	+	found in 2 different clusters in seed calculation
Thiamin diphosphate/Thiamin	+	+	
UDP-N-acetyl-D-glucosamine/N-Acetyl-D-glucosamine 1-phosphate	+	+	

Table 6: Comparison of Metagrowth data to the calculations performed in this paper for *Tropheryma whipplei* (TWH): "+" indicates a metabolite deficiency in Metagrowth or an essential seed compound in the calculations, "-" indicates a compound not identified as an essential nutrient and "o" indicates a potential, non essential nutrient.