

# Functional Annotation

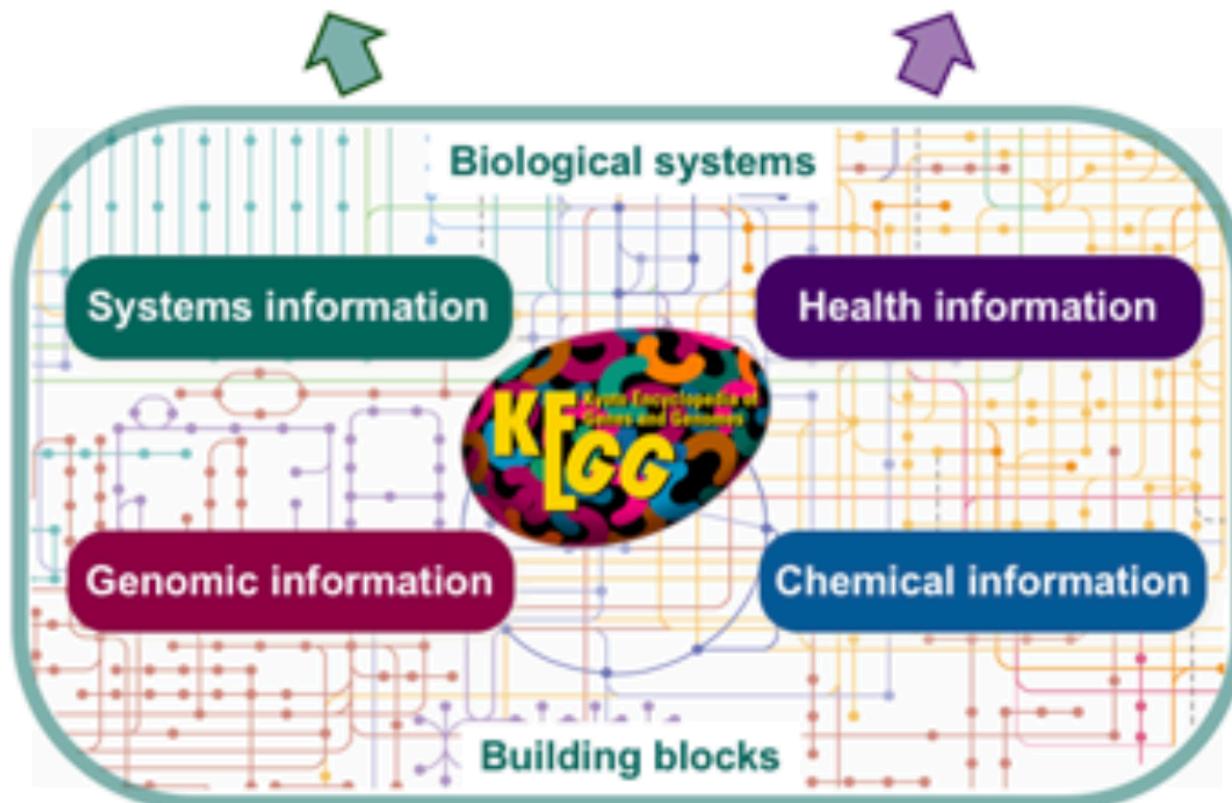
Lesson 6: Annotating with the KOALAs and KEGG database

Arkadiy Garber

Ben Tully

Basic understanding  
of biological systems

Practical applications  
for use in society



Genomes  
Metagenomes

Metabolomes

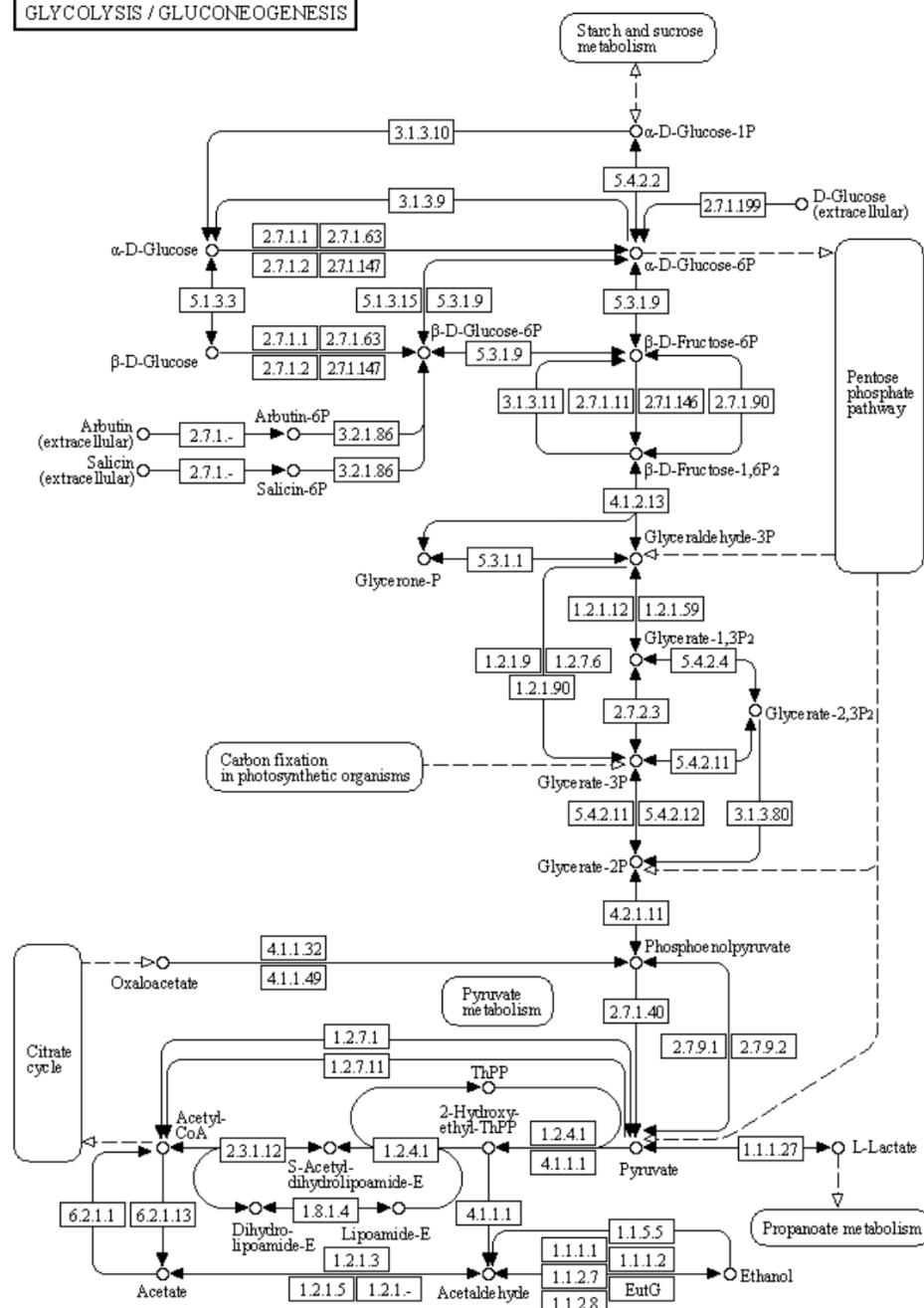
Personal genomes  
Pathogen genomes

**Table 1.** The summary of KEGG release 12.0 (October 1999)

Database	Content
PATHWAY	2706 entries for pathway diagrams constructed from 143 manually drawn diagrams
GENES	110 018 entries in 24 complete genomes and 12 partial genomes
LIGAND	5645 entries in the COMPOUND section 3705 entries in the ENZYME section 5207 reactions in the REACTION section



Category	Database	Content	Color
Systems information	KEGG PATHWAY	KEGG pathway maps	KEGG
	KEGG BRITE	BRITE hierarchies and tables	KEGG
	KEGG MODULE	KEGG modules	KEGG
Genomic information	KEGG ORTHOLOGY (KO)	Functional orthologs	KEGG
	KEGG GENOME	KEGG organisms (complete genomes)	KEGG
	KEGG GENES	Genes and proteins	KEGG
Chemical information	KEGG SSDB	GENES sequence similarity	KEGG
	KEGG COMPOUND	Small molecules	KEGG
	KEGG GLYCAN	Glycans	KEGG
	KEGG REACTION	Biochemical reactions	KEGG
	KEGG RCLASS	Reaction class	KEGG
Health information	KEGG ENZYME	Enzyme nomenclature	KEGG
	KEGG NETWORK	Disease-related network elements	KEGG
	KEGG VARIANT	Human gene variants	KEGG
	KEGG DISEASE	Human diseases	KEGG
	KEGG DRUG	Drugs	KEGG
	KEGG DGROUP	Drug groups	KEGG
	KEGG ENVIRON	Health-related substances	KEGG



# KEGG: Kyoto Encyclopedia of Genes and Genomes

Minoru Kanehisa\* and Susumu Goto

Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan

Received September 29, 1999; Accepted October 4, 1999

## Metabolism

- Global and overview maps
- Carbohydrate metabolism
- ▼ Energy metabolism
  - 00190 Oxidative phosphorylation
  - 00195 Photosynthesis
  - 00196 Photosynthesis - antenna proteins
  - 00710 Carbon fixation in photosynthetic organisms
  - 00720 Carbon fixation pathways in prokaryotes
- 00680 Methane metabolism
- 00910 Nitrogen metabolism
- 00920 Sulfur metabolism
- Lipid metabolism
- Nucleotide metabolism
- Amino acid metabolism
- Metabolism of other amino acids
- Glycan biosynthesis and metabolism
- Metabolism of cofactors and vitamins
- Metabolism of terpenoids and polyketides
- Biosynthesis of other secondary metabolites
- Xenobiotics biodegradation and metabolism
- Chemical structure transformation maps

## Genetic Information Processing

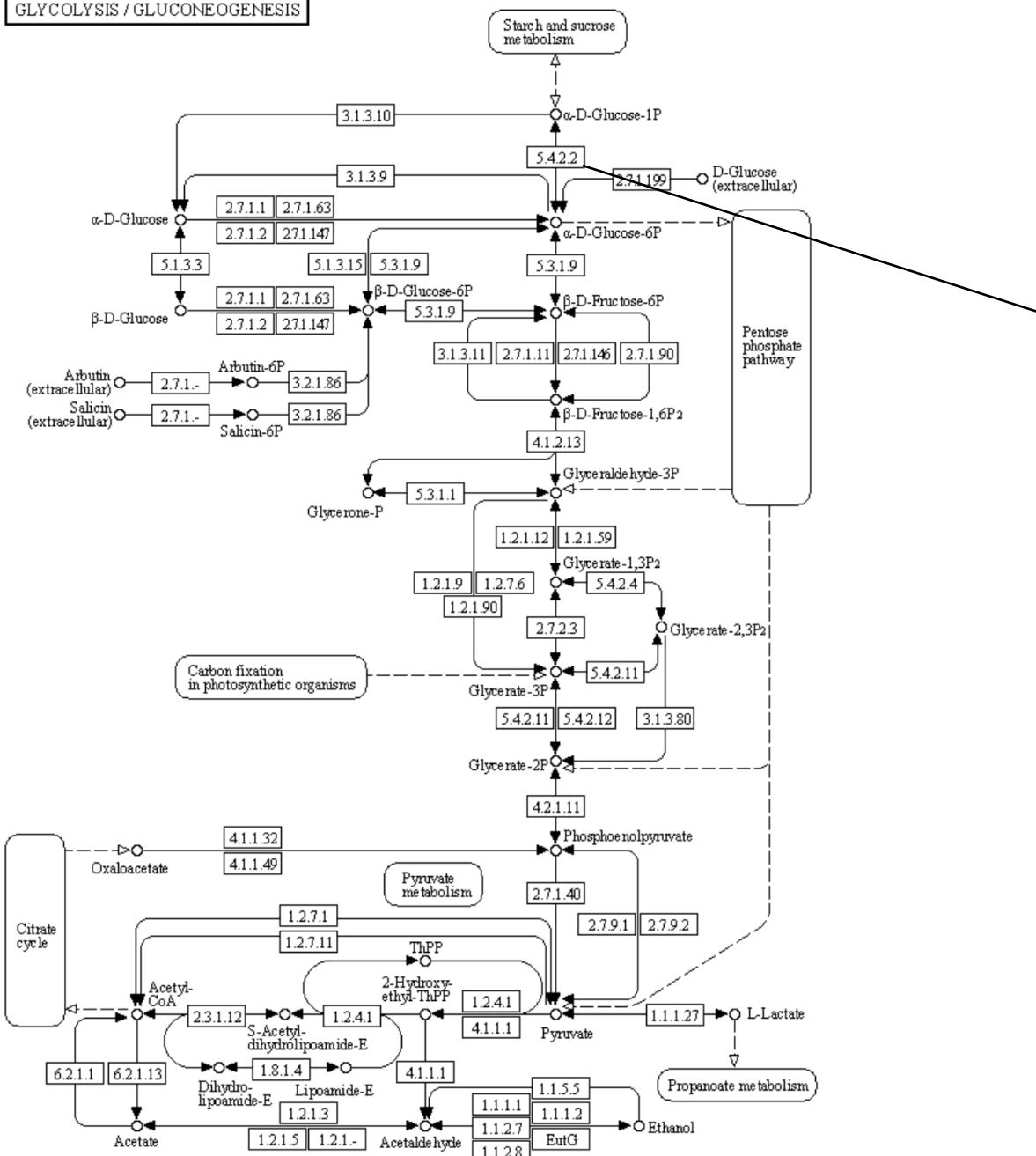
## Environmental Information Processing

## Cellular Processes

## Organismal Systems

## Human Diseases

## Drug Development



KEGG: Kyoto Encyclopedia of Genes and Genomes

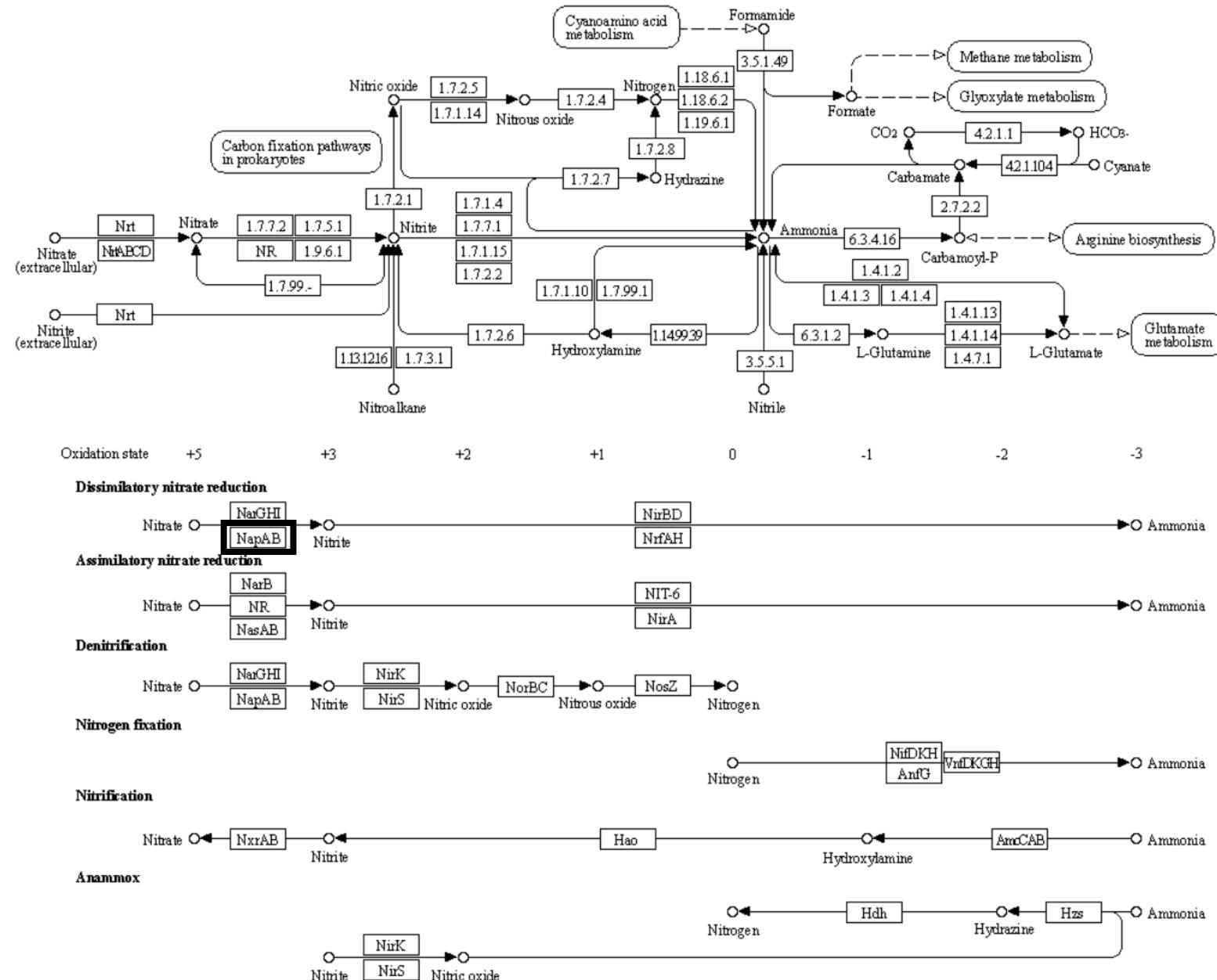
**Minoru Kanehisa\* and Susumu Goto**

Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan

Received September 29, 1999; Accepted October 4, 1999

ORTHOLOGY: K01835		
Entry	K01835	KO
Name	pgm	
Definition	phosphoglucomutase [EC:5.4.2.2]	
Pathway	<a href="#">ko00010</a> Glycolysis / Gluconeogenesis <a href="#">ko00030</a> Pentose phosphate pathway <a href="#">ko00052</a> Galactose metabolism <a href="#">ko00230</a> Purine metabolism <a href="#">ko00500</a> Starch and sucrose metabolism <a href="#">ko00520</a> Amino sugar and nucleotide sugar metabolism <a href="#">ko00521</a> Streptomycin biosynthesis <a href="#">ko01100</a> Metabolic pathways <a href="#">ko01110</a> Biosynthesis of secondary metabolites <a href="#">ko01120</a> Microbial metabolism in diverse environments	
Module	<a href="#">M00549</a> Nucleotide sugar biosynthesis, glucose => UDP-glucose <a href="#">M00855</a> Glycogen degradation, glycogen => glucose-6P	
Disease	<a href="#">H00069</a> Glycogen storage disease <a href="#">H00118</a> Congenital disorders of glycosylation type I <a href="#">H01762</a> Muscle glycogen storage disease <a href="#">H01954</a> Glycogen storage disease type XIV	
Brite	KEGG Orthology (KO) [BR: <a href="#">ko00001</a> ] <a href="#">09100</a> Metabolism <a href="#">09101</a> Carbohydrate metabolism <a href="#">00010</a> Glycolysis / Gluconeogenesis <a href="#">K01835</a> pgm; phosphoglucomutase <a href="#">00030</a> Pentose phosphate pathway <a href="#">K01835</a> pgm; phosphoglucomutase <a href="#">00052</a> Galactose metabolism <a href="#">K01835</a> pgm; phosphoglucomutase <a href="#">00500</a> Starch and sucrose metabolism <a href="#">K01835</a> pgm; phosphoglucomutase <a href="#">00520</a> Amino sugar and nucleotide sugar metabolism <a href="#">K01835</a> pgm; phosphoglucomutase <a href="#">09104</a> Nucleotide metabolism <a href="#">00230</a> Purine metabolism <a href="#">K01835</a> pgm; phosphoglucomutase <a href="#">09110</a> Biosynthesis of other secondary metabolites <a href="#">00521</a> Streptomycin biosynthesis <a href="#">K01835</a> pgm; phosphoglucomutase Enzymes [BR: <a href="#">ko01000</a> ] 5. Isomerases 5.4 Intramolecular transferases 5.4.2 Phosphotransferases (phosphomutases) 5.4.2.2 phosphoglucomutase (alpha-D-glucose-1,6-bisphosphate- <a href="#">K01835</a> pgm; phosphoglucomutase	
BRITE hierarchy		
Other DBs	RN: <a href="#">R00959</a> <a href="#">R01057</a> <a href="#">R08639</a> COG: <a href="#">COG0033</a> GO: <a href="#">0004614</a>	

## NITROGEN METABOLISM



<b>Entry</b>	K02567	KO
<b>Name</b>	napA	
<b>Definition</b>	nitrate reductase (cytochrome) [EC:1.9.6.1]	
<b>Pathway</b>	ko00910 Nitrogen metabolism ko01100 Metabolic pathways ko01120 Microbial metabolism in diverse environments	
<b>Module</b>	M00529 Denitrification, nitrate => nitrogen M00530 Dissimilatory nitrate reduction, nitrate => ammonia	
<b>Brite</b>	KEGG Orthology (KO) [BR: <a href="#">ko00001</a> ] 09100 Metabolism 09102 Energy metabolism 00910 Nitrogen metabolism K02567 napA; nitrate reductase (cytochrome) Enzymes [BR: <a href="#">ko01000</a> ] 1. Oxidoreductases 1.9 Acting on a heme group of donors 1.9.6 With a nitrogenous group as acceptor 1.9.6.1 nitrate reductase (cytochrome) K02567 napA; nitrate reductase (cytochrome) <a href="#">BRITE hierarchy</a>	
<b>Other DBs</b>	RN: <a href="#">R00798</a> <a href="#">R01106</a> <a href="#">R03071</a> COG: <a href="#">COG0243</a> GO: <a href="#">0050140</a>	
<b>Genes</b>	ECO: <a href="#">b2206</a> (napA) ECJ: <a href="#">JW2194</a> (napA) ECD: <a href="#">ECDH10B_2363</a> (napA) EBW: <a href="#">BWG_1979</a> (napA) ECOK: <a href="#">ECMDS42_1773</a> (napA) ECE: <a href="#">Z3463</a> (napA) ECS: <a href="#">Ecs3095</a> ECF: <a href="#">ECH74115_3343</a> (napA) ETW: <a href="#">ECSP_3085</a> (napA) ELX: <a href="#">CDC0157_2858</a> » show all <a href="#">Taxonomy</a> <a href="#">KOALA</a> <a href="#">UniProt</a>	
<b>Reference</b>	<b>Authors</b> PMID:11921398 Stolz JF, Basu P <b>Title</b> Evolution of nitrate reductase: molecular and structural variations on a common function. <b>Journal</b> Chembiochem 3:198-206 (2002) DOI: <a href="#">10.1002/1439-7633(20020301)3:2/3&lt;198::AID-CBIC198&gt;3.0.CO;2-C</a>	
<b>Reference</b>	<b>Authors</b> PMID:17130127 Jepson BJ, Mohan S, Clarke TA, Gates AJ, Cole JA, Butler CS, Butt JN, Hemmings AM, Richardson DJ <b>Title</b> Spectropotentiometric and structural analysis of the periplasmic nitrate reductase from Escherichia coli. <b>Journal</b> J Biol Chem 282:6425-37 (2007) DOI: <a href="#">10.1074/jbc.M607353200</a> <b>Sequence</b> [eco: <a href="#">b2206</a> ]	

<b>Entry</b>	K02568	KO
<b>Name</b>	napB	
<b>Definition</b>	nitrate reductase (cytochrome), electron transfer subunit	
<b>Pathway</b>	ko00910 Nitrogen metabolism ko01100 Metabolic pathways ko01120 Microbial metabolism in diverse environments	
<b>Module</b>	M00529 Denitrification, nitrate => nitrogen M00530 Dissimilatory nitrate reduction, nitrate => ammonia	
<b>Brite</b>	KEGG Orthology (KO) [BR: <a href="#">ko00001</a> ] 09100 Metabolism 09102 Energy metabolism 00910 Nitrogen metabolism K02568 napB; nitrate reductase (cytochrome), electron transfe <a href="#">BRITE hierarchy</a>	
<b>Other DBs</b>	RN: <a href="#">R00798</a> COG: <a href="#">COG3043</a>	
<b>Genes</b>	ECO: <a href="#">b2203</a> (napB) ECJ: <a href="#">JW5367</a> (napB) ECD: <a href="#">ECDH10B_2360</a> (napB) EBW: <a href="#">BWG_1976</a> (napB) ECOK: <a href="#">ECMDS42_1770</a> (napB) ECE: <a href="#">Z3460</a> (napB) ECS: <a href="#">Ecs3092</a> (napB) ECF: <a href="#">ECH74115_3340</a> (napB) ETW: <a href="#">ECSP_3082</a> (napB) ELX: <a href="#">CDC0157_2855</a> (napB) » show all <a href="#">Taxonomy</a> <a href="#">KOALA</a> <a href="#">UniProt</a>	
<b>Reference</b>	<b>Authors</b> PMID:11921398 Stolz JF, Basu P <b>Title</b> Evolution of nitrate reductase: molecular and structural variations on a common function. <b>Journal</b> Chembiochem 3:198-206 (2002) DOI: <a href="#">10.1002/1439-7633(20020301)3:2/3&lt;198::AID-CBIC198&gt;3.0.CO;2-C</a>	
<b>Reference</b>	<b>Authors</b> PMID:17130127 Jepson BJ, Mohan S, Clarke TA, Gates AJ, Cole JA, Butler CS, Butt JN, Hemmings AM, Richardson DJ <b>Title</b> Spectropotentiometric and structural analysis of the periplasmic nitrate reductase from Escherichia coli. <b>Journal</b> J Biol Chem 282:6425-37 (2007) DOI: <a href="#">10.1074/jbc.M607353200</a> <b>Sequence</b> [eco: <a href="#">b2203</a> ]	

## ENZYME: 1.9.6.1

Help

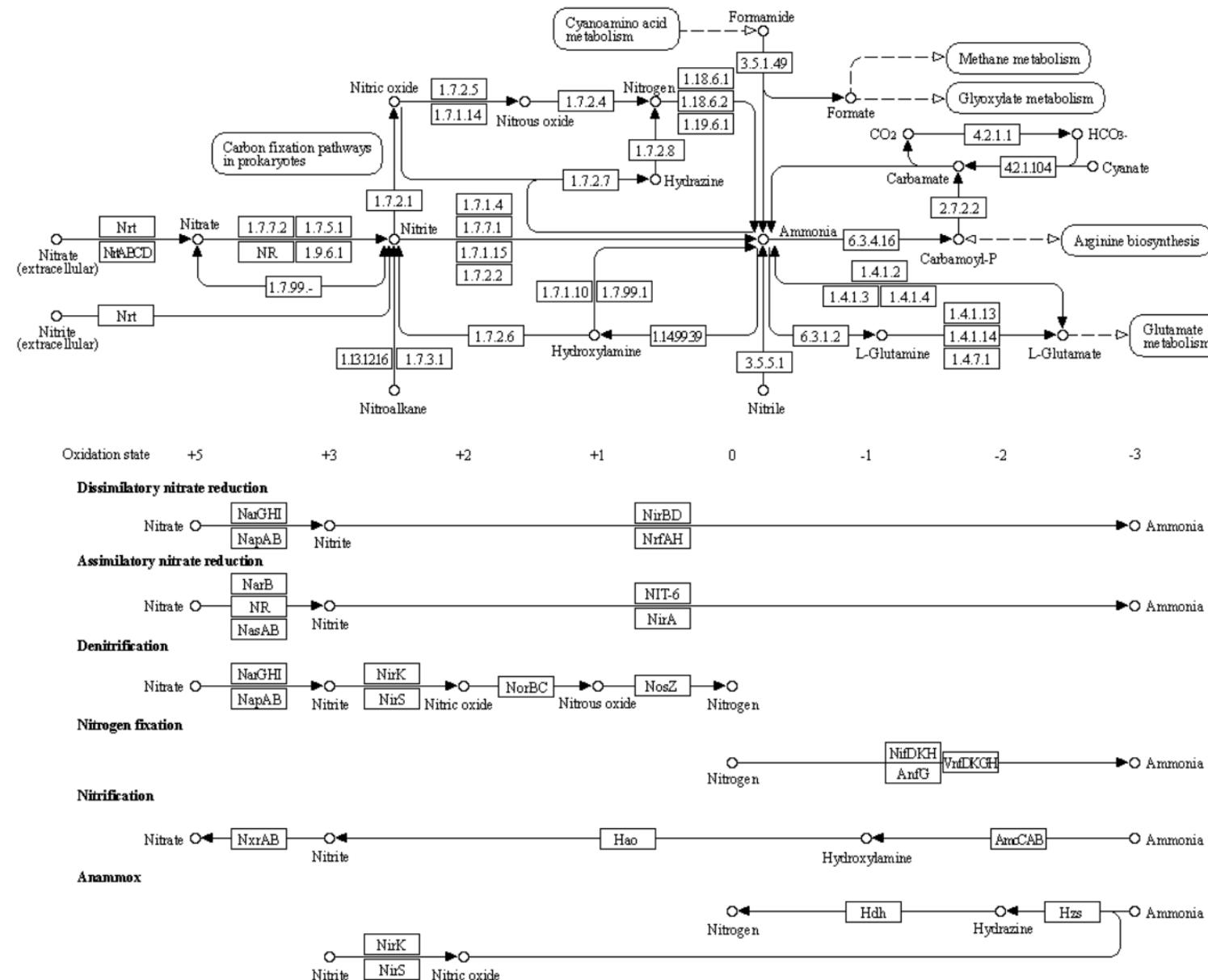
<b>Entry</b>	EC 1.9.6.1	Enzyme
<b>Name</b>	nitrate reductase (cytochrome); respiratory nitrate reductase; benzyl viologen-nitrate reductase	
<b>Class</b>	Oxidoreductases; Acting on a heme group of donors; With a nitrogenous group as acceptor <a href="#">BRITE hierarchy</a>	
<b>Sysname</b>	ferrocytochrome:nitrate oxidoreductase	
<b>Reaction (IUBMB)</b>	$2 \text{ ferrocytochrome} + 2 \text{ H}^+ + \text{nitrate} = 2 \text{ ferricytochrome} + \text{nitrite}$ [RN: <a href="#">R03071</a> ]	
<b>Reaction (KEGG)</b>	R03071 > R00792 <a href="#">Reaction</a>	
<b>Substrate</b>	ferrocytochrome [CPD: <a href="#">C00924</a> ]; $\text{H}^+$ [CPD: <a href="#">C00080</a> ]; nitrate [CPD: <a href="#">C00244</a> ]	
<b>Product</b>	ferricytochrome [CPD: <a href="#">C00923</a> ]; nitrite [CPD: <a href="#">C00088</a> ]	
<b>History</b>	EC 1.9.6.1 created 1961	
<b>Pathway</b>	<a href="#">ec00910</a> Nitrogen metabolism <a href="#">ec01100</a> Metabolic pathways	
<b>Orthology</b>	K02567 nitrate reductase (cytochrome)	
<b>Genes</b>	ECO: <a href="#">b2206</a> (napA) ECJ: <a href="#">JW2194</a> (napA) ECD: <a href="#">ECDH10B_2363</a> (napA) EBW: <a href="#">BWG_1979</a> (napA) ECOK: <a href="#">ECMDS42_1773</a> (napA) ECE: <a href="#">Z3463</a> (napA) ECS: <a href="#">ECs3095</a> ECF: <a href="#">ECH74115_3343</a> (napA) ETW: <a href="#">ECSP_3085</a> (napA) ELX: <a href="#">CDC0157_2858</a> » show all <a href="#">Taxonomy</a>	
<b>Reference</b>	1 [PMID: <a href="#">13412117</a> ] SADANA JC, MCELROY WD.	
<b>Authors</b>		
<b>Title</b>	Nitrate reductase from Achromobacter fischeri; purification and properties: function of flavines and cytochrome.	
<b>Journal</b>	Arch Biochem Biophys 67:16-34 (1957) DOI: <a href="#">10.1016/0003-9861(57)90242-4</a>	
<b>Other DBs</b>	ExplorEnz - The Enzyme Database: <a href="#">1.9.6.1</a> IUBMB Enzyme Nomenclature: <a href="#">1.9.6.1</a> ExPASy - ENZYME nomenclature database: <a href="#">1.9.6.1</a> UM-BBD (Biocatalysis/Biodegradation Database): <a href="#">1.9.6.1</a> BRENDA, the Enzyme Database: <a href="#">1.9.6.1</a> CAS: 9029-42-9	

## REACTION: R03071

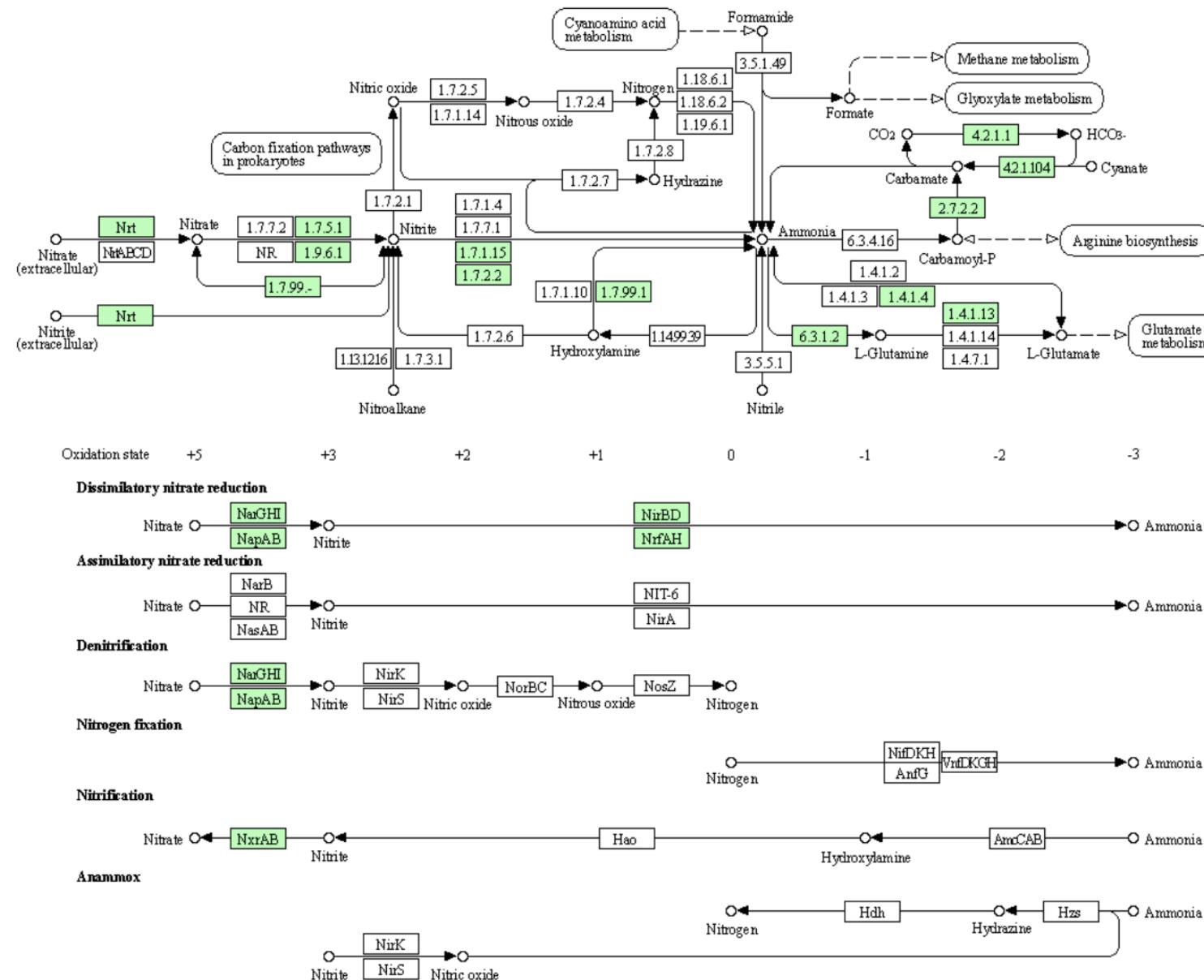
Help

<b>Entry</b>	R03071	Reaction
<b>Name</b>	ferrocytochrome:nitrate oxidoreductase	
<b>Definition</b>	$2 \text{ Ferrocytochrome} + \text{Nitrate} \rightleftharpoons 2 \text{ Ferricytochrome} + \text{Nitrite} + \text{H}_2\text{O}$	
<b>Equation</b>	$2 \text{ C00924} + \text{C00244} \rightleftharpoons 2 \text{ C00923} + \text{C00088} + \text{C00001}$	
<b>Comment</b>	nitrate reductase (cytochrome)	
<b>Reaction class</b>	<a href="#">RC02812</a> <a href="#">C00088_C00244</a>	
<b>Enzyme</b>	<a href="#">1.9.6.1</a>	
<b>Pathway</b>	<a href="#">rn00910</a> Nitrogen metabolism <a href="#">rn01100</a> Metabolic pathways	
<b>Module</b>	M00529 Denitrification, nitrate => nitrogen M00530 Dissimilatory nitrate reduction, nitrate => ammonia	
<b>Orthology</b>	K02567 nitrate reductase (cytochrome) [EC:1.9.6.1]	
<b>Other DBs</b>	RHEA: <a href="#">12912</a>	

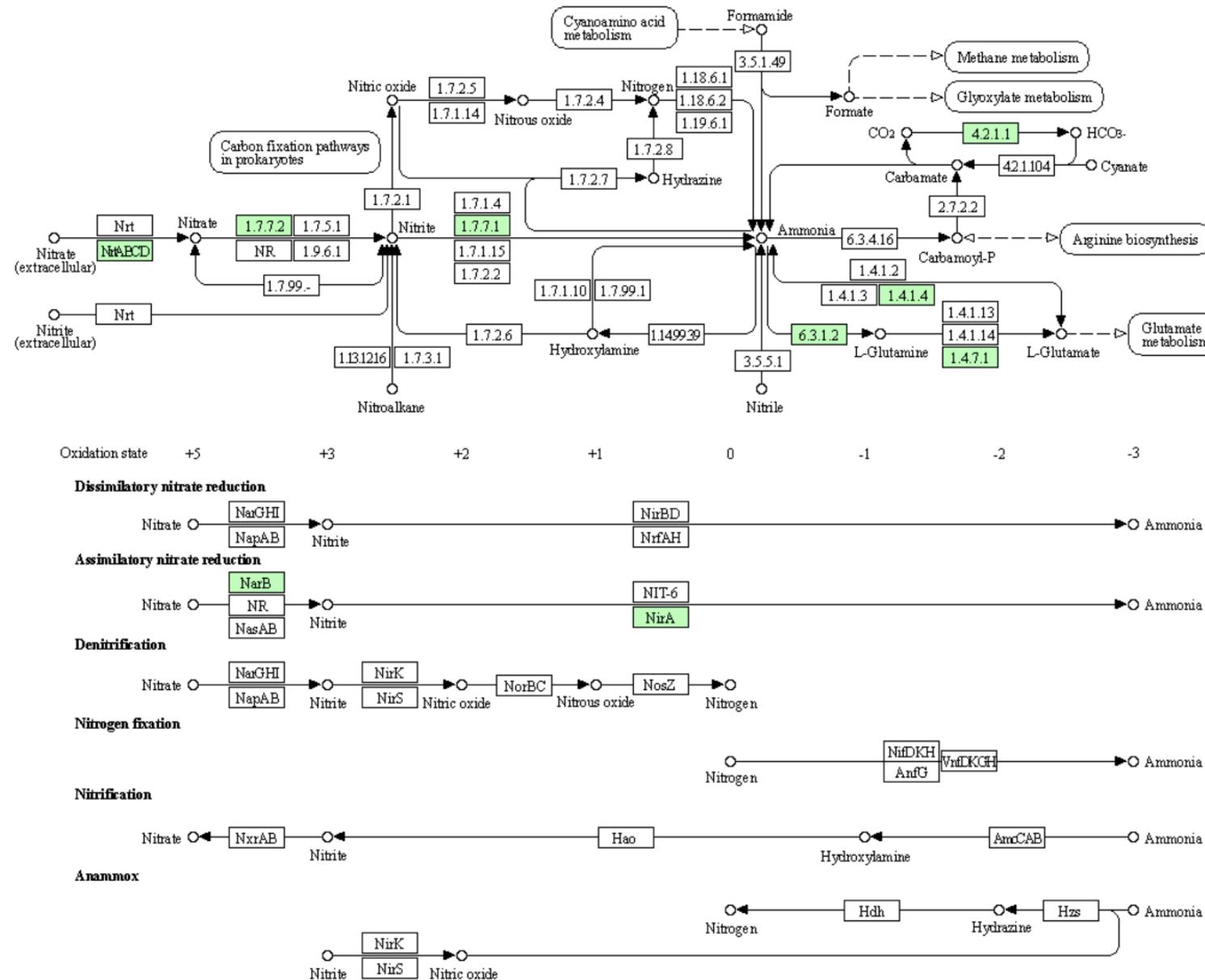
## NITROGEN METABOLISM



## NITROGEN METABOLISM



## NITROGEN METABOLISM



Category	Database	Content	Color
Systems information	KEGG PATHWAY	KEGG pathway maps	
	KEGG BRITE	BRITE hierarchies and tables	
	KEGG MODULE	KEGG modules	
Genomic information	KEGG ORTHOLOGY (KO)	Functional orthologs	
	KEGG GENOME	KEGG organisms (complete genomes)	
	KEGG GENES	Genes and proteins	
Chemical information	KEGG SSDB	GENES sequence similarity	
	KEGG COMPOUND	Small molecules	
	KEGG GLYCAN	Glycans	
	KEGG REACTION	Biochemical reactions	
	KEGG RCLASS	Reaction class	
Health information	KEGG ENZYME	Enzyme nomenclature	
	KEGG NETWORK	Disease-related network elements	
	KEGG VARIANT	Human gene variants	
	KEGG DISEASE	Human diseases	
	KEGG DRUG	Drugs	
	KEGG DGROUP	Drug groups	
	KEGG ENVIRON	Health-related substances	

# BlastKOALA, GhostKOALA, and KofamKOALA



**GhostKOALA**  
Query Data Input

**KEGG** Automatic annotation and  
KEGG mapping service

	BlastKOALA	GhostKOALA	KofamKOALA
<b>KOALA job status</b> 2020/05/16 10:10:28 (GMT+9)			
Number of jobs in the queue	Blast	Ghost	Kofam
Submission of last completed job	2020/05/16 00:02:13	2020/05/15 22:24:21	2020/05/16 07:36:40

KOALA (KEGG Orthology And Links Annotation) is KEGG's internal annotation tool for K number assignment of KEGG GENES using SSEARCH computation. BlastKOALA and GhostKOALA assign K numbers to the user's sequence data by BLAST and GHOSTX searches, respectively, against a nonredundant set of KEGG GENES. KofamKOALA is a new member of the KOALA family available at GenomeNet using the HMM profile search, rather than the sequence similarity search, for K number assignment. See [Step-by-step Instructions](#).

Reference: Kanehisa, M., Sato, Y., and Morishima, K. (2016) BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726-731. [[pubmed](#)] [[pdf](#)]

**GhostKOALA accepts a larger dataset and is suitable for annotating metagenomes**

**Upload query amino acid sequences in FASTA format**

Enter FASTA sequences

Or upload file:  no file selected

Your query data consisting of multiple amino acid sequences will be given K numbers by GhostKOALA. The file size of up to 300 MB (one million sequences with average length of 300 or three million sequences with average length of 100) may be uploaded.

**Enter KEGG GENES database file to be searched**

genus\_prokaryotes  
 genus\_prokaryotes + family\_eukaryotes  
 genus\_prokaryotes + family\_eukaryotes + viruses

The database files for GhostKOALA are somewhat different from those for BlastKOALA. For each group of KEGG organisms at the genus or family level, a nonredundant dataset is generated by taking all protein-coding genes from the representative genome and additional genes from the other genomes with two criteria. One is the same as in BlastKOALA, different K numbers, and the other is unique to GhostKOALA, different CD-HIT clusters, which are computed with 50% identity cutoff. In addition, the database file for viruses is created by CD-HIT with 90% identity cutoff from the viruses category of KEGG GENES. These additions are meant for analyzing taxonomic compositions of metagenomes.

Enter your email address



## BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences

Minoru Kanehisa<sup>1</sup>, Yoko Sato<sup>2</sup> and Kanae Morishima<sup>1</sup>

<sup>1</sup> - Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan

<sup>2</sup> - Healthcare Solutions Department, Fujitsu Kyushu Systems Ltd., Hakata-ku, Fukuoka 812-0007, Japan

**Correspondence to Minoru Kanehisa:** [kanehisa@kuicr.kyoto-u.ac.jp](mailto:kanehisa@kuicr.kyoto-u.ac.jp)

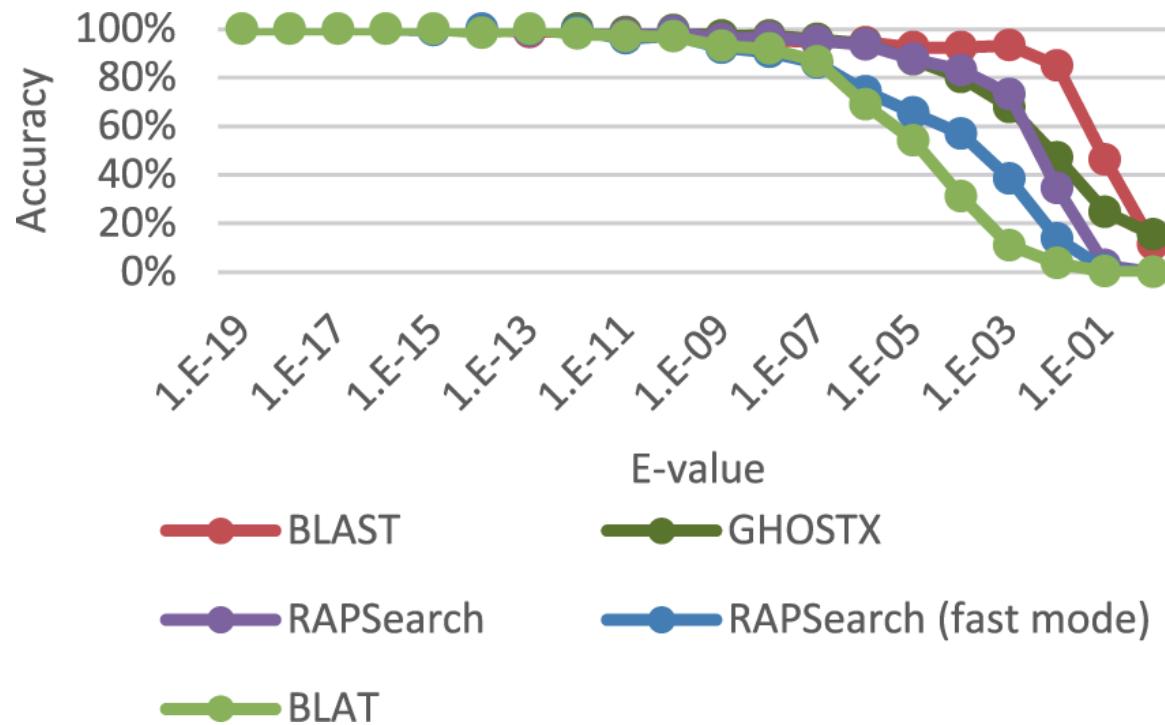
<http://dx.doi.org/10.1016/j.jmb.2015.11.006>

**Edited by M. Sternberg**

# GHOSTX: An Improved Sequence Homology Search Algorithm Using a Query Suffix Array and a Database Suffix Array

Shuji Suzuki, Masanori Kakuta, Takashi Ishida, Yutaka Akiyama\*

Graduate School of Information Science and Engineering, Tokyo Institute of Technology, Meguro-ku, Tokyo, Japan



BLASTx vs GHOSTX  
GHOSTX about 50 times faster than BLASTx

# BlastKOALA and GhostKOALA output

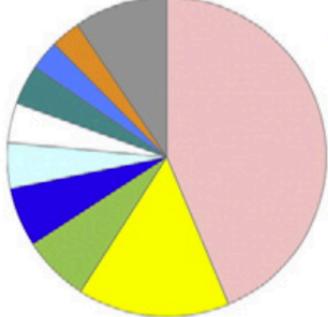
(a)

**Your BlastKOALA job**  
Query dataset: 12669 entries  
Taxonomy group: Eukaryotes, Animals  
KEGG database searched: family\_eukaryotes.pep  
Job submitted: Mon Jun 22 11:53:34 JST 2015  
Job completed: Mon Jun 22 20:21:02 JST 2015

**Annotation data** [View](#) | [Download](#)

Summary 4949 entries (39.1%) annotated

Functional category



**KEGG Mapper**

[Reconstruct Pathway](#)  
[Reconstruct Brite](#)  
[Reconstruct Module](#)

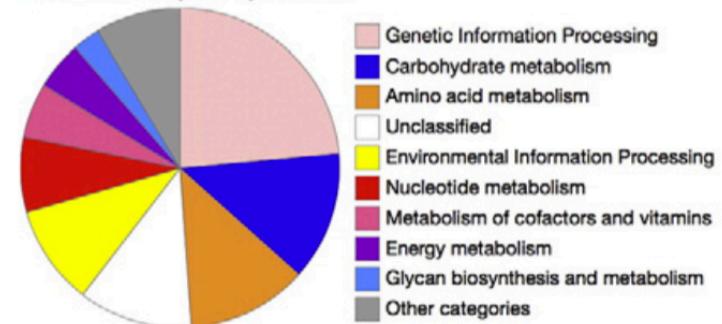
(b)

**Your GhostKOALA job**  
Query dataset: 725424 entries  
KEGG database searched: c\_family\_euk+genus\_prok+viruses  
Job submitted: Wed Jun 17 16:54:07 JST 2015  
Job completed: Wed Jun 17 18:41:26 JST 2015

**Annotation data** [Preview first 100](#) | [Download](#)

Summary 328383 entries (45.3%) annotated

Functional category



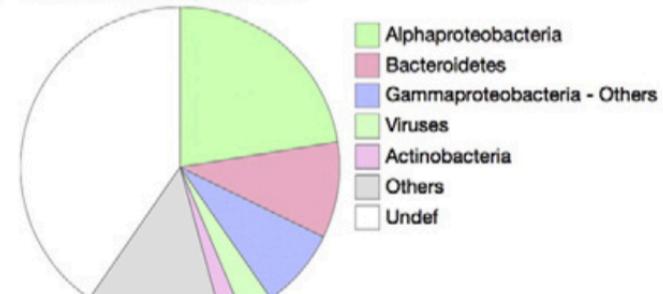
**KEGG Mapper**

[Reconstruct Pathway](#)  
[Reconstruct Brite](#)  
[Reconstruct Module](#)

**Taxonomy data**

[Preview first 100](#) | [Download](#)

Taxonomic category



[More details](#)

[View all taxonomic categories \(and perform KEGG Mapper analysis\)](#)

# GhostKOALA raw output

user.ko_definition.txt						
b_1	145					
1_2	22	K10967	1			
1_3	12	K02274	3			
1_4	K01990	ABC-2.A; ABC-2 type transport system ATP-binding protein	208			
1_5	14	K01992	5			
1_6	16	K20463	5			
1_7	55	K21430	3			
1_8	20	K06560	1			
1_9	15	K23910	4			
1_10	20	K01893	1			
1_11	K01715	crt; enoyl-CoA hydratase [EC:4.2.1.17]	199			
1_12	K00074	paaH; 3-hydroxybutyryl-CoA dehydrogenase [EC:1.1.1.157]	243			
1_13	12	K10127	6			
1_14	K00626	ACAT; acetyl-CoA C-acetyltransferase [EC:2.3.1.9]	445			
1_15	50	K19223	25			
1_16	93	K00248	66			
1_17	K00248	ACADS; butyryl-CoA dehydrogenase [EC:1.3.8.1]	306	K18244	35	
1_18	12	K03306	7			
1_19	K00995	pgsA; CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase [EC:2.7.8.5]	10			
1_20	54					
1_21	K03057	tfs; transcription factor S	85			
1_22	K00330	nuoA; NADH-quinone oxidoreductase subunit A [EC:7.1.1.2]	67			
1_23	K00331	nuoB; NADH-quinone oxidoreductase subunit B [EC:7.1.1.2]	251			
1_24	111	K0332	12			
1_25	K00333	nuoD; NADH-quinone oxidoreductase subunit D [EC:7.1.1.2]	348			
1_26	K00337	nuoH; NADH-quinone oxidoreductase subunit H [EC:7.1.1.2]	324			
1_27	K00338	nuoI; NADH-quinone oxidoreductase subunit I [EC:7.1.1.2]	159	K22164	10	
1_28	K05576	ndhE; NAD(P)H-quinone oxidoreductase subunit 4L [EC:7.1.1.2]	45	K00340	24	
1_29	K00342	nuoM; NADH-quinone oxidoreductase subunit M [EC:7.1.1.2]	175	K15863	31	
1_30	K00341	nuoL; NADH-quinone oxidoreductase subunit L [EC:7.1.1.2]	627			
1_31	K00342	nuoM; NADH-quinone oxidoreductase subunit M [EC:7.1.1.2]	324			
1_32	K03768	fppN; F420H2 dehydrogenase subunit N [EC:1.5.98.3]	399			
1_33	K22169	fpnB; F420H2 dehydrogenase subunit N [EC:1.5.98.3]	399			
1_34	K02523	ispB; octaprenyl-diphosphate synthase [EC:2.5.1.90]	190	K13787	15	
1_35	K03768	PPIB; peptidyl-prolyl cis-trans isomerase B (cyclophilin B) [EC:5.2.1.8]	111			
1_36	K03768	PPIB; peptidyl-prolyl cis-trans isomerase B (cyclophilin B) [EC:5.2.1.8]	132			
1_37	K01890	FARSb; phenylalanyl-tRNA synthetase beta chain [EC:6.1.1.20]	402			
1_38	K01889	FARSA; phenylalanyl-tRNA synthetase alpha chain [EC:6.1.1.20]	454			
1_39	K01867	WARS; tryptophanyl-tRNA synthetase [EC:6.1.1.2]	423			
1_40	306					
1_41	18	K16733	1			
1_42	K01265	map; methionyl aminopeptidase [EC:3.4.11.18]	273			
1_43	K02201	K01529	12			
1_44	K22491	litR; MerR family transcriptional regulator, light-induced transcriptional regulator	105	K00548	6	
1_45	K07071	uncharacterized protein 190				
1_46	18	K04257	3			
1_47	295					
1_48	16	K08796	5			
1_49	14	K02374	7			
1_50	14	K05546	5			
1_51	K01934	MTHFS; 5-formyltetrahydfolate cyclo-ligase. [EC:6.3.3.2]	117			
1_52	K06864	larE; pyridinium-3,2'-discarboxyllic acid mononucleotide sulfuryltransferase [EC:4.4.1.37]	243			
1_53	K06898	larB; pyridinium-3,2'-discarboxyllic acid mononucleotide synthase [EC:2.5.1.143]	225			
1_54	K09121	larC; pyridinium-3,2'-discarboxylic acid mononucleotide nickel chelatase [EC:4.99.1.12]	293			
1_55	32	K23886	8			
1_56	K00820	glmS; glutamine--fructose-6-phosphate transaminase (isomerizing) [EC:2.6.1.16]	547			
1_57	K02948	RP-S11; small subunit ribosomal protein S11	155			
1_58	16	K12567	3			
1_59	K09735	uncharacterized protein 111				
1_60	K03050	ppgE; DNA-directed RNA polymerase subunit E' [EC:2.7.7.6]	75			
1_61	K03049	ppgE1; DNA-directed RNA polymerase subunit E' [EC:2.7.7.6]	227			
1_62	13	K07305	4			
1_63	K03242	EIF2S3; translation initiation factor 2 subunit 3	456			
1_64	K02991	RP-S6e; small subunit ribosomal protein S6e	119			
1_65	K03243	EIF5B; translation initiation factor 5B	58			
1_66	K00940	ndk; nucleoside-diphosphate kinase [EC:2.7.4.6]	156			
1_67	K02896	RP-L24e; large subunit ribosomal protein L24e	72			
1_68	K02979	RP-S28e; small subunit ribosomal protein S28e	81			
1_69	K02936	RP-L7Ae; large subunit ribosomal protein L7Ae	161			
1_70	K01657	trpE; anthranilate synthase component I [EC:4.1.3.27]	415			
1_71	K01658	trpG; anthranilate synthase component II [EC:4.1.3.27]	235			
1_72	K0766	trpD; anthranilate phosphoribosyltransferase [EC:2.4.2.18]	267	K13497	31	
1_73	K01609	trpC; indole-3-glycerol phosphate synthase [EC:4.1.1.48]	184			
1_74	K01696	trpB; tryptophan synthase beta chain [EC:4.2.1.20]	524			
1_75	K01695	trpA; tryptophan synthase alpha chain [EC:4.2.1.20]	225			

user.out.top						
user:1_1	K01990	Archaea Bathyarchaeota Candidatus Bathyarchaeota archaeon BA2	barb:A0A66_1372	164.081		
user:1_2		Bacteria Actinobacteria Saccharomonospora	svi:svr_15830	51.9878		
user:1_3		Bacteria Bacteroidetes Fibrella	fae:FAES_2062	32.3426		
user:1_4		K01990 Archaea Euryarchaeota Methanomassiliicoccus	mer:MMINT_08100	219.55		
user:1_5		Bacteria Deltaproteobacteria Bacteriovorax	bsto:COV70_00670	48.521		
user:1_6		Bacteria Gammaproteobacteria - Others	Alteromonas amcIMAUe_1000555	50.447		
user:1_7		K01990 Archaea Euryarchaeota Methanobolbus	mpv:Mpsv_0275	63.1586		
user:1_8		Bacteria Alphaproteobacteria Labrenzia	labf:FJ695_22405	32.3426		
user:1_9		Archaea Thaumarchaeota Candidatus Nitrosococcus	taa:MY3_00903	84.7297		
user:1_10		Bacteria Alphaproteobacteria Nitrobacter	nha:Nham_3731	73.559		
user:1_11	K01715	Bacteria Firmicutes - Clostridia Thermanaerobacter	twi:Thewi_0615	226.868		
user:1_12	K00074	Archaea Cuniculiplasma cuniculiplasma	cdv:CPM_0579	285.804		
user:1_13	K07068	Archaea Thaumarchaeota Candidatus Nitrosopelagicus	nbv:T478_1344	91.2781		
user:1_14	K00626	Archaea Thaumarchaeota Candidatus Nitrosotalea ndv:NEVE_1984	468.003			
user:1_15		Plants Eudicots Cicer cam:cam10506405				
user:1_16		Bacteria Chlorobi Ignavibacterium iai_LALB				
user:1_17	K00248	Bacteria Firmicutes - Bacilli Lentibacillus lao:AOX59_15560				
user:1_18		Bacteria Alphaproteobacteria Candidatus Pelagibacter	pub:SAR11_0965	72.7886		
user:1_19		Bacteria Alphaproteobacteria Candidatus Pelagibacter	pub:SAR11_0965	59.3066		
user:1_20		Archaea Thaumarchaeota Candidatus Nitroscaldus	ncv:NCAV_1111	87.8113		
user:1_21	K03057	Archaea Thaumarchaeota Candidatus Nitrosotalea ndv:DEVE_2086				
user:1_22		Nitrosphaera nvn:NVIE_011670				
user:1_23	K00331	Archaea Thaumarchaeota Nitrosphaera nga:Ngar_c11520				
user:1_24		Archaea Thaumarchaeota Candidatus Nitrosococcus	taa:MY3_02144	133.265		
user:1_25	K00333	Archaea Thaumarchaeota Candidatus Nitroscaldus	ncv:NCAV_0267	460.685		
user:1_26	K00337	Archaea Thaumarchaeota Candidatus Nitrosotalea ndv:NEVE_1826				
user:1_27	K00338	Archaea Thaumarchaeota Candidatus Nitroscaldus	ncv:NCAV_0269	189.889		
user:1_28	K00339	Archaea Thaumarchaeota Candidatus Nitrosococcus	ndv:NFRAN_2933	124.02		
user:1_29		Archaea Thaumarchaeota Candidatus Nitroscaldus	ncv:NCAV_0272	86.2705		
user:1_30	K00342	Archaea Thaumarchaeota Candidatus Nitroscaldus	ncv:NCAV_0273	519.62		
user:1_31	K00341	Archaea Thaumarchaeota Nitrosphaera	nga:Ngar_c11600	665.225		
user:1_32	K00342	Bacteria Deferribacteres Calditerrivibrion	qui:Calni_0325	332.413		
user:1_33	K22169	Archaea Thaumarchaeota Candidatus Nitrospelagicus	nbv:T478_1161	426.017		
user:1_34	K02523	Archaea Thaumarchaeota Nitrosphaera nga:Ngar_c11620				
user:1_35	K03788	Bacteria Nitospirae Nitospirale	nde:NIDE1683	202.986		
user:1_36	K03768	Bacteria Nitospirae Nitospirale	nde:NIDE1683	203.756		
user:1_37	K01890	Archaea Thaumarchaeota Nitrosphaera nga:Ngar_c23540				
user:1_38	K01889	Archaea Thaumarchaeota Nitrosphaera nga:Ngar_c17500				
user:1_39	K01867	Archaea Bathyarchaeota Candidatus Nitrosotalea ndv:NEVE_1757				
user:1_40		Archaea Thaumarchaeota Candidatus Bathyarchaeota archaeon BA1	barb:A0A65_2091	451.055		
user:1_41		Candidatus Nitrototenuis	tah:SU86_003115	333.954		
user:1_42	K01265	Archaea Thaumarchaeota Candidatus Nitroscaldus	ncv:NCAV_1075	50.8322		
user:1_43	K02201	Archaea Thaumarchaeota Candidatus Nitrospelagicus	nbv:T478_0094	288.5		
user:1_44	K22491	Archaea Thaumarchaeota Candidatus Nitrospelagicus	nbv:T478_0659	207.994		
user:1_45	K07071	Bacteria Planctomycetes Gemmati gmar:GMAR_30250				
user:1_46		Archaea Thaumarchaeota Candidatus Nitrototenuis	tah:SU86_000790	107.071		
user:1_47		Archaea Thaumarchaeota Candidatus Nitrosococcus	taa:MY3_01131	404.831		
user:1_48		Animals Placozoa Trichoplax	tad:TRIADRAFT_49890	33.4982		
user:1_49		Archaea Thaumarchaeota Candidatus Caldarchaeota	csu:CSUB_C0824	57.3806		
user:1_50	K01934	Archaea Bathyarchaeota Candidatus Bathyarchaeota archaeon BA2	barb:A0A66_1018	45.4394		
user:1_51	K06864	Bacteria Planctomycetes Gemmati ges:PYT8_29015				
user:1_52	K06898	Archaea Bathyarchaeota archaeon BA2	barb:A0A66_0335	248.44		
user:1_53	K09121	Archaea Bathyarchaeota archaeon BA2	barb:A0A66_0334	322.783		
user:1_54	K09121	Candidatus Bathyarchaeota archaeon BA2				
user:1_55	K23886	Archaea Crenarchaeota Caldivirga cmv:Cmv_1075				
user:1_56	K00820	Archaea Thaumarchaeota Nitrosphaera nga:Ngar_c21900				
user:1_57	K02948	Archaea Thaumarchaeota Nitrosphaera nga:Ngar_cw4580				
user:1_58	K09735	Archaea Thaumarchaeota Candidatus Nitrosopelagicus	nbv:T478_1094	52.373		
user:1_59	K03050	Archaea Thaumarchaeota Candidatus Nitrosopelagicus	nbv:T478_0497	120.168		
user:1_60	K03049	Archaea Thaumarchaeota Candidatus Nitrosopelagicus				
user:1_61	K07158	Archaea Thaumarchaeota Candidatus Nitrosopelagicus				
user:1_62	K02842	Archaea Thaumarchaeota Candidatus Nitrosopelagicus				
user:1_63	K02921	Archaea Thaumarchaeota Candidatus Nitrosopelagicus				
user:1_64	K03243	Archaea Thaumarchaeota Candidatus Nitrosopelagicus				
user:1_65	K00940	Archaea Thaumarchaeota Candidatus Nitrosopelagicus				
user:1_66	K02896	Archaea Thaumarchaeota Candidatus Nitrosopelagicus				
user:1_67	K02979	Archaea Thaumarchaeota Candidatus Nitrosopelagicus				
user:1_68	K02979	Archaea Crenarchaeota Nitrosopilum nmr:Nmr_0227				
user:1_69	K02936	Archaea Crenarchaeota Nitrosopilum nmr:Nmr_0227				
user:1_70	K01657	Archaea Bathyarchaeota Candidatus Bathyarchaeota archaeon BA2	barb:A0A66_0467	493.426		
user:1_71	K01658	Archaea Thaumarchaeota Candidatus Bathyarchaeota archaeon BA2	barb:A0A66_0465	347.436		
user:1_72	K00766	Archaea Bathyarchaeota Candidatus Bathyarchaeota archaeon BA2	barb:A0A66_0465	347.436		
user:1_73	K01609	Archaea Bathyarchaeota Candidatus Bathyarchaeota archaeon BA2	barb:A0A66_0465	347.436		
user:1_74	K01696	Archaea Thaumarchaeota Candidatus Nitrosopilum nmr:Nmr_0914				
user:1_75	K01695	Archaea Thaumarchaeota Candidatus Nitrosopilum nmr:Nmr_0914				

[Code](#)[Issues 0](#)[Pull requests 0](#)[Actions](#)[Projects 0](#)[Wiki](#)[Security 0](#)[Insights](#)[Settings](#)

Script for converting raw KO-number output from GhostKOALA and BlastKOALA into hierarchically-structures CSV format

[Edit](#)[Manage topics](#)

6 commits

1 branch

0 packages

0 releases

1 contributor

Branch: master

[New pull request](#)[Create new file](#)[Upload files](#)[Find file](#)[Clone or download](#) Arkadiy-Garber Update README.md

Latest commit 7d39131 21 seconds ago

 [KO-convert-script.py](#)

Add files via upload

4 hours ago

 [KO\\_Orthology\\_ko00001.txt](#)

Add files via upload

3 hours ago

 [README.md](#)

Update README.md

21 seconds ago

## README.md

### Installation

```
git clone https://github.com/Arkadiy-Garber/KOALA-formatter.git
```

Put KO-convert-script.py in your \$PATH

### Basic usage

```
KO-convert-script.py -db KO_Orthology_ko00001.txt -ko user_ko_definition.txt -taxa user.out.top -out ghostokoala-summary.csv
```

KO\_Orthology\_ko00001.txt can be found in this repository. This file contains gene and pathway information for every KO number. This file is sourced from Elaina Graham's GhostKoalaParser repo: <https://github.com/edgraham/GhostKoalaParser>

user\_ko\_definition.txt is the main file that is created when you download the results from your finished GhostKOALA run.

user.out.top is the output file that is created when you download the taxonomy data. This input is optional.

ghostokoala-summary.csv is what we want to name our output file

Incorporation of  
GhostKOALA results  
into Anvio's  
metagenomic workflow

# What does the final output look like?

ORF ↑	KO	KEGG category	KEGG family	KEGG system	gene	taxa1	taxa2	taxa3
1_102	K03273	01005 Lipopolysaccharide biosynthesis proteins [BR:ko01005]	09107 Glycan biosynthesis and metabolism	09100 Metabolism	gmhB; D-glycero-D-manno-heptose 1;7-bisphosphate phosphatase	Bacteria	Elusimicrobia	Endomicrobium
1_103	K03271	01005 Lipopolysaccharide biosynthesis proteins [BR:ko01005]	09107 Glycan biosynthesis and metabolism	09100 Metabolism	gmhA; lpcA; D-sedoheptulose 7-phosphate isomerase	Archaea	Thaumarchaeota	Nitrosopumilus
1_104	K07031	00540 Lipopolysaccharide biosynthesis [PATH:ko00540]	09107 Glycan biosynthesis and metabolism	09100 Metabolism	hddA; D-glycero-alpha-D-manno-heptose-7-phosphate kinase	Archaea	Thaumarchaeota	Nitrosopumilus
1_11	K01715	00650 Butanoate metabolism [PATH:ko00650]	09101 Carbohydrate metabolism	09100 Metabolism	crt; enoyl-CoA hydratase	Bacteria	Firmicutes - Clostridia	Thermoanaerobacter
1_12	K00074	00362 Benzoate degradation [PATH:ko00362]	09111 Xenobiotics biodegradation and metabolism	09100 Metabolism	paaH; hbd; fadB; mmgB; 3-hydroxybutyryl-CoA dehydrogenase	Archaea	Euryarchaeota	Cuniculiplasma
1_14	K00626	04147 Exosome [BR:ko04147]	09141 Transport and catabolism	09140 Cellular Processes	E2.3.1.9; atoB; acetyl-CoA C-acetyltransferase	Archaea	Thaumarchaeota	Candidatus Nitrosotalea
1_17	K00248	00280 Valine; leucine and isoleucine degradation [PATH:ko00280]	09105 Amino acid metabolism	09100 Metabolism	ACADS; bcd; butyryl-CoA dehydrogenase	Bacteria	Firmicutes - Bacilli	Lentibacillus
1_19	K00995	00564 Glycerophospholipid metabolism [PATH:ko00564]	09103 Lipid metabolism	09100 Metabolism	pgsA; PGS1; CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatase	Bacteria	Alphaproteobacteria	Candidatus Pelagibacter
1_21	K03057	03021 Transcription machinery [BR:ko03021]	09121 Transcription	09120 Genetic Information Processing	tfS; transcription termination factor TFS	Archaea	Thaumarchaeota	Candidatus Nitrosotalea
1_22	K00330	00190 Oxidative phosphorylation [PATH:ko00190]	09102 Energy metabolism	09100 Metabolism	nuoA; NADH-quinone oxidoreductase subunit A	Archaea	Thaumarchaeota	Nitrososphaera
1_23	K00331	00190 Oxidative phosphorylation [PATH:ko00190]	09102 Energy metabolism	09100 Metabolism	nuoB; NADH-quinone oxidoreductase subunit B	Archaea	Thaumarchaeota	Nitrososphaera
1_25	K00333	00190 Oxidative phosphorylation [PATH:ko00190]	09102 Energy metabolism	09100 Metabolism	nuoD; NADH-quinone oxidoreductase subunit D	Archaea	Thaumarchaeota	Candidatus Nitrosocaldus
1_26	K00337	00190 Oxidative phosphorylation [PATH:ko00190]	09102 Energy metabolism	09100 Metabolism	nuoH; NADH-quinone oxidoreductase subunit H	Archaea	Thaumarchaeota	Candidatus Nitrosotalea
1_27	K00338	00190 Oxidative phosphorylation [PATH:ko00190]	09102 Energy metabolism	09100 Metabolism	nuoI; NADH-quinone oxidoreductase subunit I	Archaea	Thaumarchaeota	Candidatus Nitrosocaldus
1_29	K05576	00190 Oxidative phosphorylation [PATH:ko00190]	09102 Energy metabolism	09100 Metabolism	ndhE; NAD(P)H-quinone oxidoreductase subunit 4L	Archaea	Thaumarchaeota	Candidatus Nitrosocaldus
1_30	K00342	00190 Oxidative phosphorylation [PATH:ko00190]	09102 Energy metabolism	09100 Metabolism	nuoM; NADH-quinone oxidoreductase subunit M	Archaea	Thaumarchaeota	Candidatus Nitrosocaldus
1_31	K00341	00190 Oxidative phosphorylation [PATH:ko00190]	09102 Energy metabolism	09100 Metabolism	nuoL; NADH-quinone oxidoreductase subunit L	Archaea	Thaumarchaeota	Nitrososphaera
1_32	K00342	00190 Oxidative phosphorylation [PATH:ko00190]	09102 Energy metabolism	09100 Metabolism	nuoM; NADH-quinone oxidoreductase subunit M	Bacteria	Deferrabacteres	Calditerrivibrio
1_33	K22169	99982 Energy metabolism	09191 Metabolism	09190 Unclassified	fpoN; F420H2 dehydrogenase subunit N	Archaea	Thaumarchaeota	Candidatus Nitrosopelagicus
1_34	K02523	00900 Terpenoid backbone biosynthesis [PATH:ko00900]	09109 Metabolism of terpenoids and polyketides	09100 Metabolism	ispB; octaprenyl-diphosphate synthase	Archaea	Thaumarchaeota	Nitrososphaera
1_35	K03768	03110 Chaperones and folding catalysts [BR:ko03110]	09123 Folding; sorting and degradation	09120 Genetic Information Processing	PPIB; ppiB; peptidyl-prolyl cis-trans isomerase B (cyclophilin B)	Bacteria	Nitrospirae	Nitrospira
1_36	K03768	03110 Chaperones and folding catalysts [BR:ko03110]	09123 Folding; sorting and degradation	09120 Genetic Information Processing	PPIB; ppiB; peptidyl-prolyl cis-trans isomerase B (cyclophilin B)	Bacteria	Nitrospirae	Nitrospira
1_37	K01890	00970 Aminoacyl-tRNA biosynthesis [PATH:ko00970]	09122 Translation	09120 Genetic Information Processing	FARSB; phET; phenylalanyl-tRNA synthetase beta chain	Archaea	Thaumarchaeota	Nitrososphaera
1_38	K01889	00970 Aminoacyl-tRNA biosynthesis [PATH:ko00970]	09122 Translation	09120 Genetic Information Processing	FARSA; phES; phenylalanyl-tRNA synthetase alpha chain	Archaea	Thaumarchaeota	Candidatus Nitrosotalea
1_39	K01867	00970 Aminoacyl-tRNA biosynthesis [PATH:ko00970]	09122 Translation	09120 Genetic Information Processing	WARS; trpS; tryptophanyl-tRNA synthetase	Archaea	Bathyarchaeota	Candidatus Bathyarchaeota
1_4	K01990	02000 Transporters [BR:ko02000]	09131 Membrane transport	09130 Environmental Information Processing	ABC-2.A; ABC-2 type transport system ATP-binding protein	Archaea	Euryarchaeota	Methanomassiliicoccus
1_42	K01265	01002 Peptidases [BR:ko01002]	09112 Enzyme families	09100 Metabolism	map; methionyl aminopeptidase	Archaea	Thaumarchaeota	Candidatus Nitrosopelagicus
1_43	K02201	00770 Pantothenate and CoA biosynthesis [PATH:ko00770]	09108 Metabolism of cofactors and vitamins	09100 Metabolism	E2.7.7.3B; pantetheine-phosphate adenyltransferase	Archaea	Thaumarchaeota	Candidatus Nitrosopelagicus
1_45	K07071	99997 Function unknown	09194 Poorly characterized	09190 Unclassified	K07071; uncharacterized protein	Bacteria	Planctomycetes	Gimesia
1_51	K01934	00670 One carbon pool by folate [PATH:ko00670]	09108 Metabolism of cofactors and vitamins	09100 Metabolism	MTHFS; 5-formyltetrahydrofolate cyclo-ligase	Bacteria	Firmicutes - Clostridia	Ruminiclostridium
1_52	K06864	99987 Metabolism of cofactors and vitamins	09191 Metabolism	09190 Unclassified	larE; pyridinium-3;5-biscarboxylic acid mononucleotide sulfurtransferase	Bacteria	Planctomycetes	Gemmata
1_53	K06898	99987 Metabolism of cofactors and vitamins	09191 Metabolism	09190 Unclassified	larB; nicotinic acid adenine dinucleotide carboxylase/hydrolase	Archaea	Bathyarchaeota	Candidatus Bathyarchaeota
1_54	K09121	99987 Metabolism of cofactors and vitamins	09191 Metabolism	09190 Unclassified	larC; pyridinium-3;5-bisthiocarboxylic acid mononucleotide nucleotidyl transferase	Archaea	Bathyarchaeota	Candidatus Bathyarchaeota
1_56	K00820	04931 Insulin resistance [PATH:ko04931]	09166 Endocrine and metabolic diseases	09160 Human Diseases	glmS; GFPT; glucosamine---fructose-6-phosphate aminotransferase	Archaea	Thaumarchaeota	Nitrososphaera
1_57	K02948	03011 Ribosome [BR:ko03011]	09122 Translation	09120 Genetic Information Processing	RP-S11; MRPS11; rpsK; small subunit ribosomal protein S11	Archaea	Thaumarchaeota	Nitrososphaera
1_59	K09735	99997 Function unknown	09194 Poorly characterized	09190 Unclassified	K09735; uncharacterized protein	Archaea	Thaumarchaeota	Candidatus Nitrosopelagicus
1_60	K03050	03021 Transcription machinery [BR:ko03021]	09121 Transcription	09120 Genetic Information Processing	rpoE2; DNA-directed RNA polymerase subunit E"	Archaea	Thaumarchaeota	Nitrososphaera
1_61	K03049	03021 Transcription machinery [BR:ko03021]	09121 Transcription	09120 Genetic Information Processing	rpoE1; DNA-directed RNA polymerase subunit E'	Archaea	Thaumarchaeota	Nitrososphaera
1_63	K03242	03012 Translation factors [BR:ko03012]	09122 Translation	09120 Genetic Information Processing	EIF2S3; translation initiation factor 2 subunit 3	Archaea	Thaumarchaeota	Nitrososphaera
1_64	K02991	01521 EGFR tyrosine kinase inhibitor resistance [PATH:ko01521]	09168 Drug resistance	09160 Human Diseases	RP-S6e; RPS6; small subunit ribosomal protein S6e	Archaea	Thaumarchaeota	Candidatus Nitrosopelagicus
1_65	K03243	03012 Translation factors [BR:ko03012]	09122 Translation	09120 Genetic Information Processing	EIF5B; translation initiation factor 5B	Archaea	Thaumarchaeota	Candidatus Nitrosocaldus



## KofamKOALA - KEGG Orthology Search

K number assignment based on KO-dependent scoring criteria

BlastKOALA

GhostKOALA

KofamKOALA

KOALA job status 2020/05/17 14:53:49 (GMT+9)

	Blast	Ghost	Kofam
Number of jobs in the queue	0	0	0
Submission of last completed job	2020/05/17 09:24:40	2020/05/17 11:04:35	2020/05/17 12:31:06

KofamKOALA assigns K numbers to the user's sequence data by HMMER/HMMSEARCH against KOfam (a customized HMM database of KEGG Orthologs (KOs)). K number assignments with scores above the predefined thresholds for individual KOs are more reliable than other proposed assignments. Such high score assignments are highlighted with asterisks '\*' in the output. The K number assignments facilitate the interpretation of the annotation results by linking the user's sequence data to the KEGG pathways and EC numbers.

Enter FASTA Sequences

or upload a sequence file

Choose File GCF\_00021919..protein.faa

E-value

Hits with scores above the predefined adaptive thresholds and E-values lower than or equal to the specified threshold will be reported with '\*'.

0.01

E-mail

rkdgarber@gmail.com

Compute

Clear

Genome analysis

# KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold

Takuya Aramaki<sup>1</sup>, Romain Blanc-Mathieu<sup>1</sup>, Hisashi Endo<sup>1</sup>, Koichi Ohkubo<sup>1,2</sup>, Minoru Kanehisa<sup>1</sup>, Susumu Goto<sup>3</sup> and Hiroyuki Ogata<sup>1,\*</sup>

<sup>1</sup>Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611-0011, <sup>2</sup>Hewlett-Packard Japan Ltd., Koto-ku, Tokyo 136-8711 and <sup>3</sup>Database Center for Life Science, Research Organization of Information and Systems, Kashiwa, Chiba 277-0871, Japan

\*To whom correspondence should be addressed.

Associate Editor: Alfonso Valencia

Received on April 9, 2019; revised on October 2, 2019; editorial decision on November 14, 2019; accepted on November 16, 2019

**KOfamscan about 83 times faster than BLASTx  
About twice as fast as GHOSTX**

## Current release

- ver. 2020-04-02
  - KEGG release 94.0

## Version history

## Download

KOfam - HMM profiles for KEGG/KO with predefined score thresholds ([download](#))  
 KofamScan - Software to search KOfam ([download](#))

## Reference

- Aramaki T., Blanc-Mathieu R., Endo H., Ohkubo K., Kanehisa M., Goto S., Ogata H.  
 KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold.  
*Bioinformatics*. 2019 Nov 19. pii: btz859. doi: [10.1093/bioinformatics/btz859](https://doi.org/10.1093/bioinformatics/btz859).

Last updated: November 25, 2019



## Search result

[Above threshold](#)[All hits](#)[KEGG Mapper](#)[Download text file](#)

Name	KO (threshold)	HMM score	E-value	Definition
* WP_013975834.1	K02970 (29.03)	61.1	1.2e-18	small subunit ribosomal protein S21
* WP_013975839.1	K01940 (133.33)	280.7	2.5e-85	argininosuccinate synthase [EC:6.3.4.5]
* WP_013975841.1	K05524 (105.03)	131.5	3.4e-40	ferredoxin
* WP_013975843.1	K04043 (790.20)	874.7	6e-265	molecular chaperone DnaK
* WP_013975844.1	K01626 (128.73)	181.3	3.3e-55	3-deoxy-7-phosphoheptulonate synthase [EC:2.5.1.54]
* WP_013975846.1	K00549 (42.33)	484.0	6.6e-147	5-methyltetrahydropteroylglutamate--homocysteine methyltransferase [EC:2.1.1.14]
* WP_013975848.1	K02968 (25.10)	46.8	3e-14	small subunit ribosomal protein S20
* WP_013975849.1	K02871 (117.00)	143.3	1.1e-43	large subunit ribosomal protein L13
* WP_013975850.1	K02996 (109.93)	142.2	2.2e-43	small subunit ribosomal protein S9
* WP_013975853.1	K09924 (39.83)	73.9	2.3e-22	uncharacterized protein
* WP_013975857.1	K02959 (25.60)	83.0	3.4e-25	small subunit ribosomal protein S16
* WP_013975858.1	K02884 (21.27)	30.7	2.2e-09	large subunit ribosomal protein L19
* WP_013975859.1	K00297 (162.47)	220.2	5.1e-67	methylenetetrahydrofolate reductase (NADPH) [EC:1.5.1.20]
* WP_013975861.1	K04077 (162.77)	858.1	4.6e-260	chaperonin GroEL
* WP_013975862.1	K04078 (26.00)	130.9	5.3e-40	chaperonin GroES
* WP_013975863.1	K02337 (887.40)	1080.2	0	DNA polymerase III subunit alpha [EC:2.7.7.7]
* WP_013975866.1	K00053 (113.97)	394.7	7.1e-120	ketol-acid reductoisomerase [EC:1.1.1.86]
* WP_013975868.1	K02956 (30.83)	79.8	2.8e-24	small subunit ribosomal protein S15
* WP_013975870.1	K02899 (37.17)	135.5	2.9e-41	large subunit ribosomal protein L27

# KEGG-Decoder: Evaluation and visualization KEGG KOALA output

<https://github.com/bjtully/BioData/tree/master/KEGGDecoder>

 README.md 

---

## KEGG-Decoder

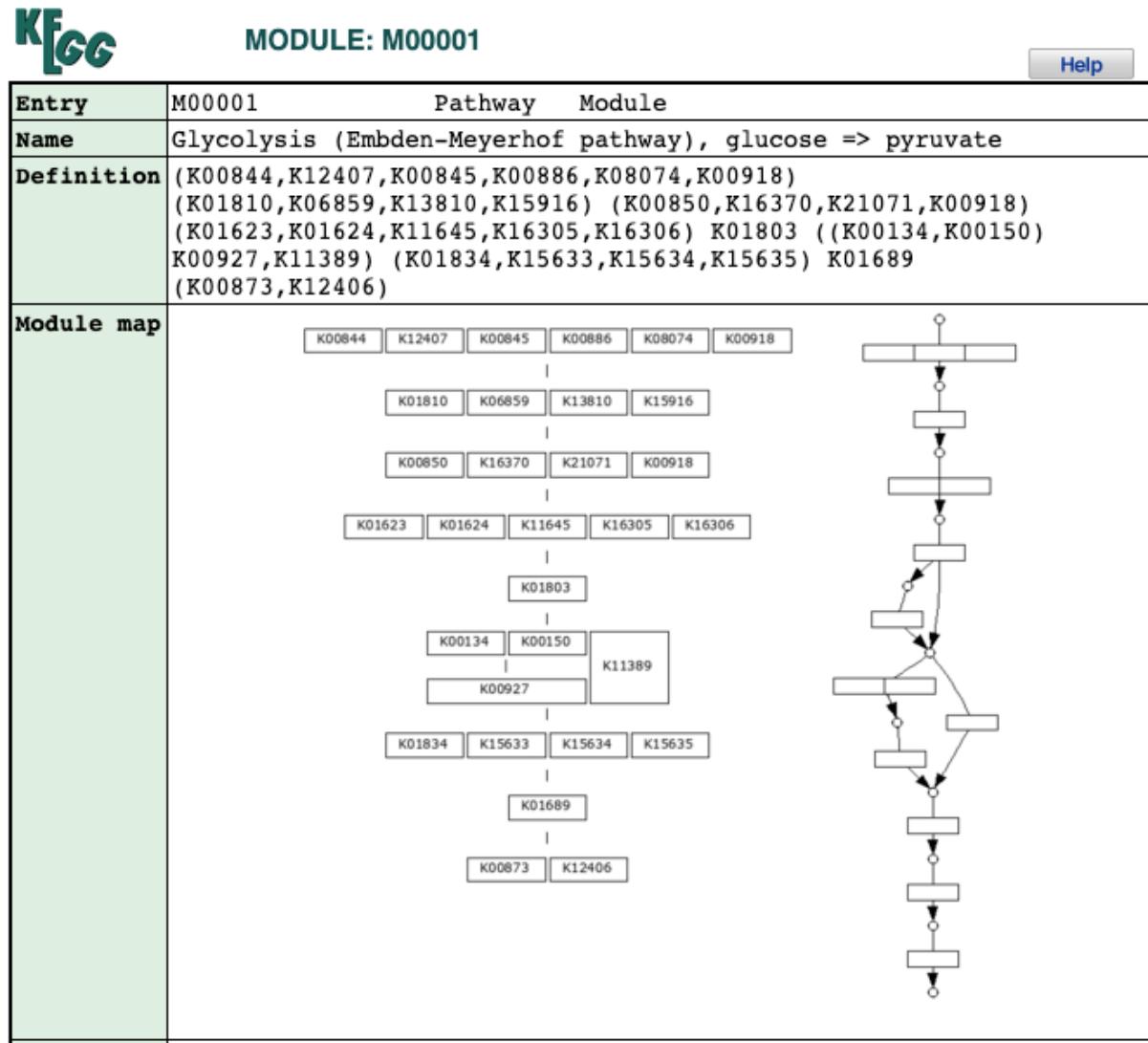
---

### Description

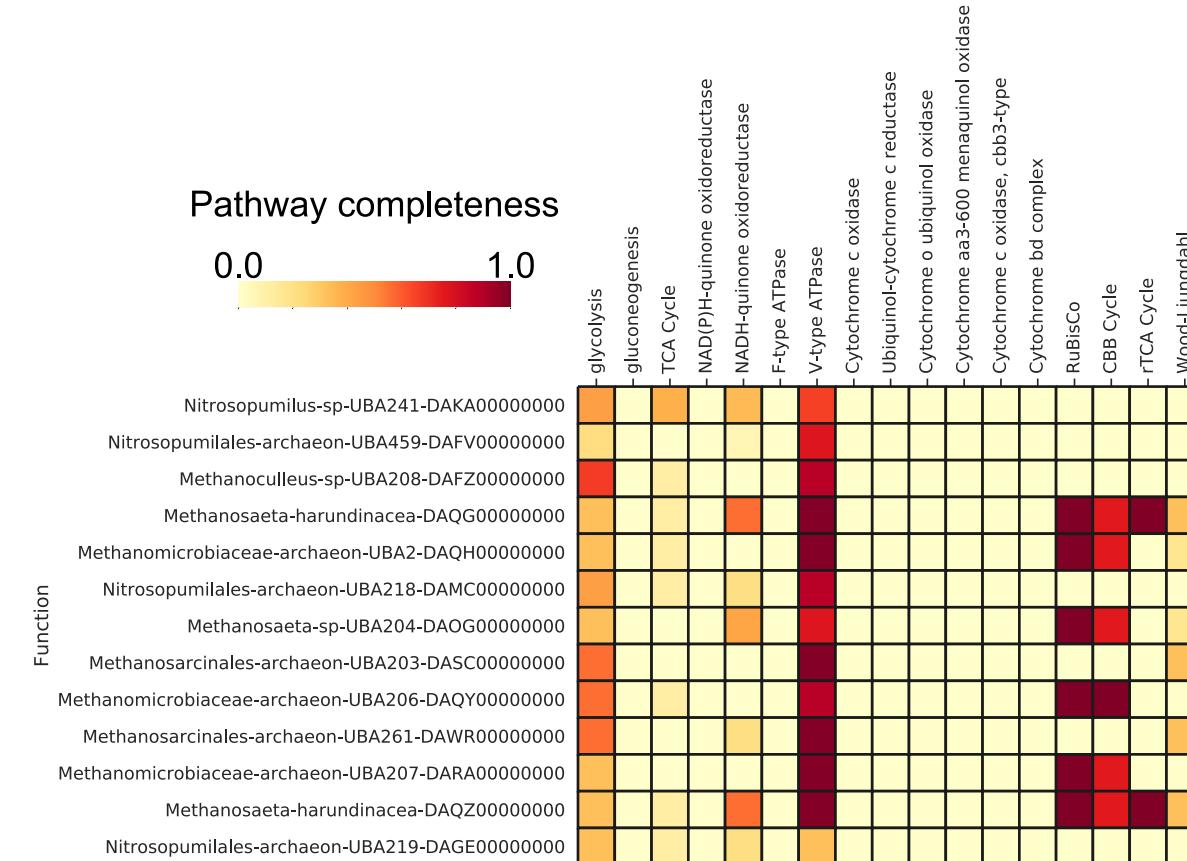
Designed to parse through a KEGG-Koala outputs (including blastKOALA, ghostKOALA, KOFAMSCAN) to determine the completeness of various metabolic pathways.

- This module was constructed using manually curated "canonical" pathways described as part of KEGG Pathway Maps.  
For information regarding which KOs are used to predict a metabolic pathway see the KOALA\_definitions.txt
- if you are interested in certain pathway and the genes are listed in KEGG it is possible to add it to file (with some Python scripting)

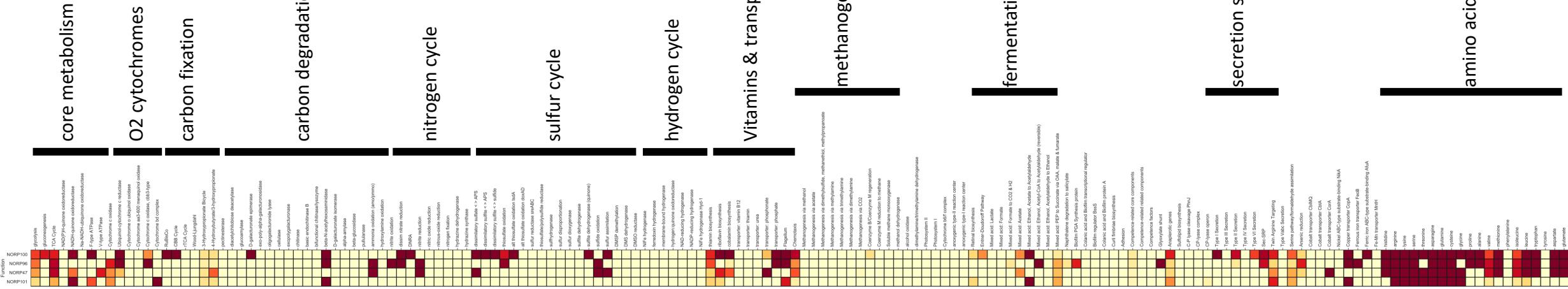
# KEGG Modules



# KEGG Decoder heatmap



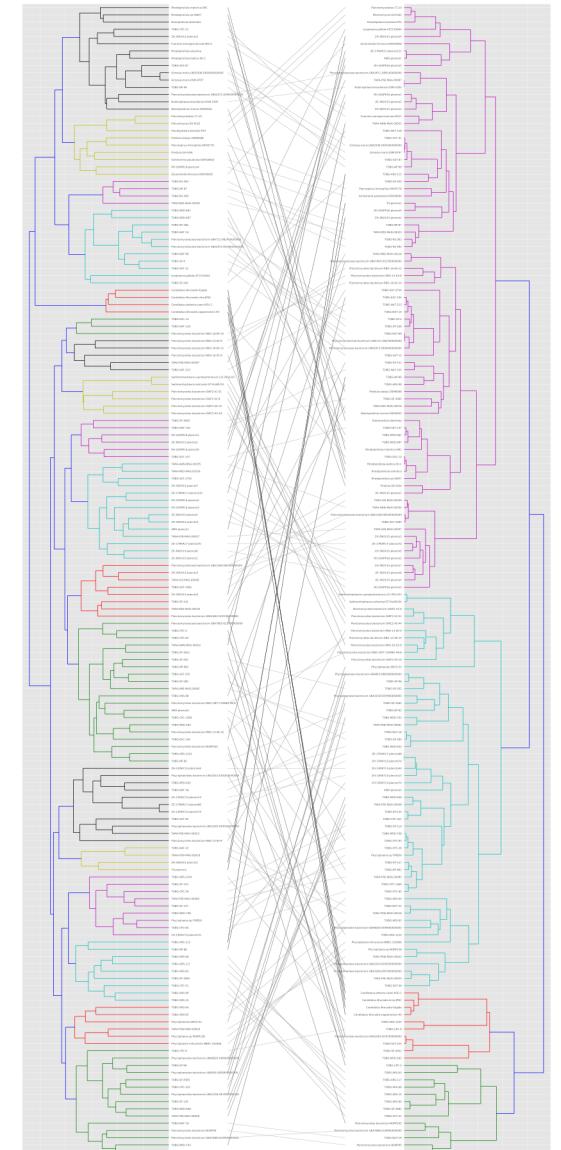
# KEGG-Decoder



Designed to provide a high-level overview of metabolic processes → additional confirmation required

# KEGG-Decoder Visualization Options

- Static
- Interactive → HTML version with zoom and hover features
  - Ordered by user input
- Tanglegram





Onto the Jupyter Binder Tutorial:

