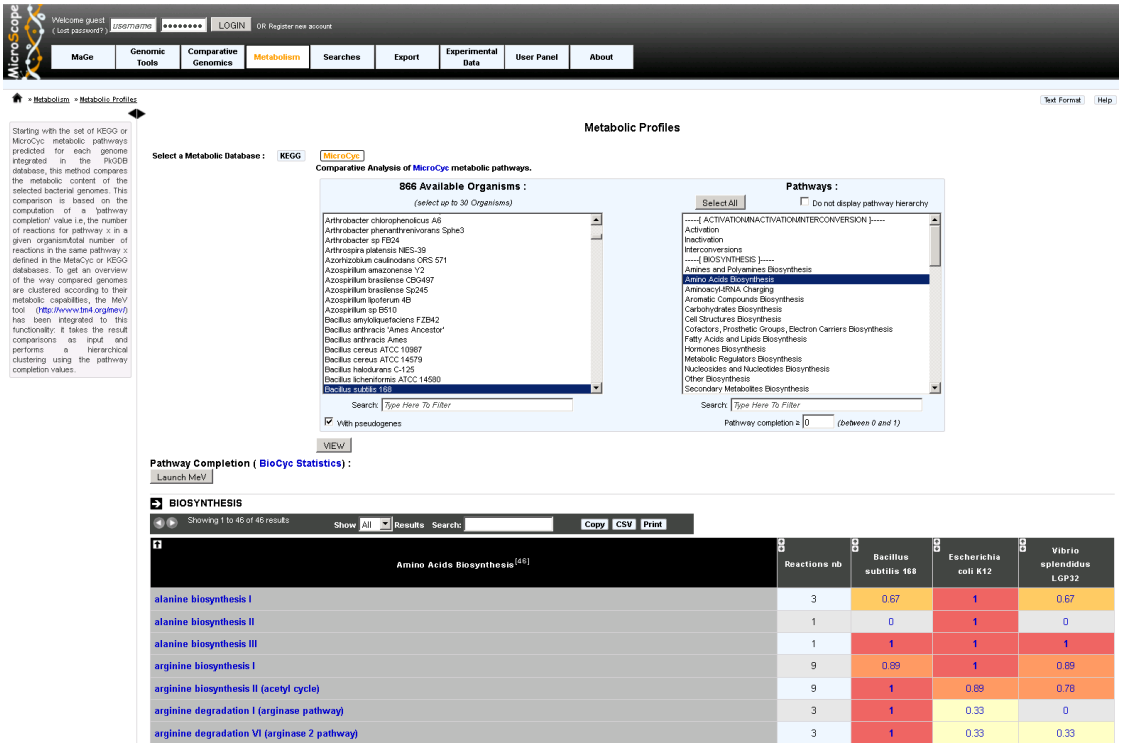
Curation of metabolic data using the MicroScope platform: Corrected exercises

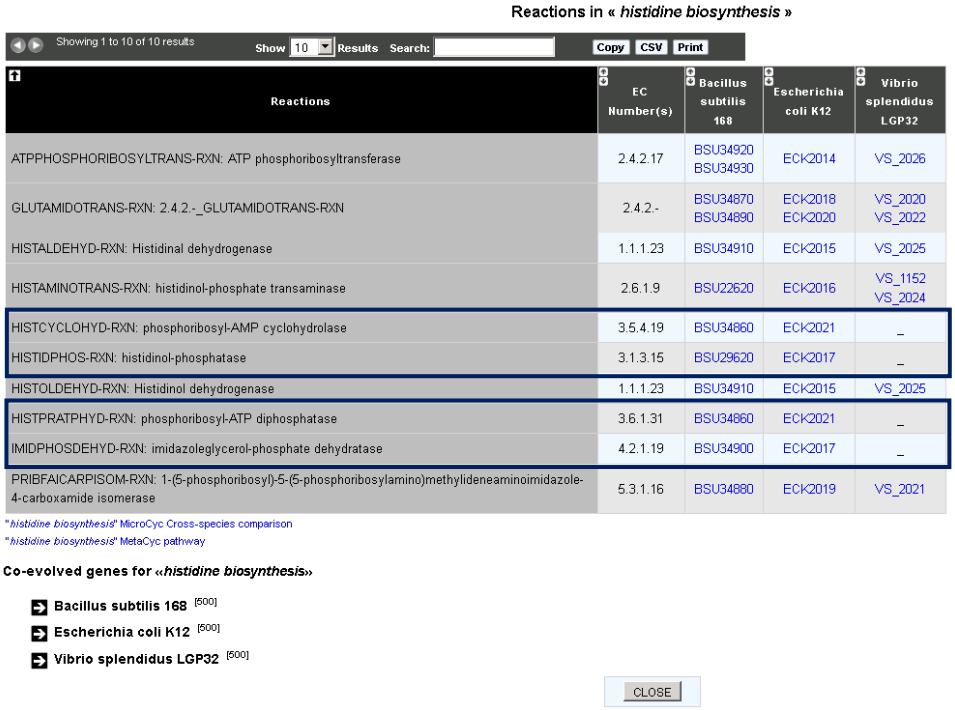
<u>Ex. T-1</u>: Metabolic profiles and co-evolved genes

Use the MicroCyc Metabolic profile tool for *E. coli* K12, *Bacillus subtilis* 168 and *Vibrio splendidus* LGP32. How many pathway holes are present in histidine biosynthesis pathway in LGP32? In the Metabolism section of Microscope main bar, select Metabolic profiles. In the Metabolic profiles page, select the genomes of interest and the pathways you want to compare. In the context of this exercise, we select Amino Acids biosynthesis, which includes histidine biosynthesis pathway. By clicking on the VIEW button, the results are displayed in the form of the degree of completion of selected pathways in the selected organisms:



We can see that the pathway is incomplete in LGP32 (completion score < 1.00). By clicking on the histidine biosynthesis hyperlink, we can display the pathway table with the reactions and

predicted genes in the selected genomes. We can see that **four enzymatic activities appear absents in LGT32:**



Reactions in « *histidine biosynthesis* »

Showing 1 to 10 of 10 results | Show 10 Results | Search: | Copy | CSV | Print

Reactions	EC Number(s)	Bacillus subtilis 168	Escherichia coli K12	Vibrio splendidus LGP32
ATPPHOSPHORIBOSYLTRANS-RXN: ATP phosphoribosyltransferase	2.4.2.17	BSU34920 BSU34930	ECK2014	VS_2026
GLUTAMIDOTRANS-RXN: 2.4.2.-_GLUTAMIDOTRANS-RXN	2.4.2.-	BSU34870 BSU34890	ECK2018 ECK2020	VS_2020 VS_2022
HISTALDEHYD-RXN: Histidinal dehydrogenase	1.1.1.23	BSU34910	ECK2015	VS_2025
HISTAMINOTRANS-RXN: histidinol-phosphate transaminase	2.6.1.9	BSU22620	ECK2016	VS_1152 VS_2024
HISTCYCLOHYD-RXN: phosphoribosyl-AMP cyclohydrolase	3.5.4.19	BSU34860	ECK2021	_
HISTIDPHOS-RXN: histidinol-phosphatase	3.1.3.15	BSU29620	ECK2017	_
HISTOLDEHYD-RXN: Histidinol dehydrogenase	1.1.1.23	BSU34910	ECK2015	VS_2025
HISTPRATPHYD-RXN: phosphoribosyl-ATP diphosphatase	3.6.1.31	BSU34860	ECK2021	_
IMIDPHOSDEHYD-RXN: imidazoleglycerol-phosphate dehydratase	4.2.1.19	BSU34900	ECK2017	_
PRIBFAICARPISOM-RXN: 1-(5-phosphoribosyl)-5-(5-phosphoribosylamino)methylideneaminoimidazole-4-carboxamide isomerase	5.3.1.16	BSU34880	ECK2019	VS_2021

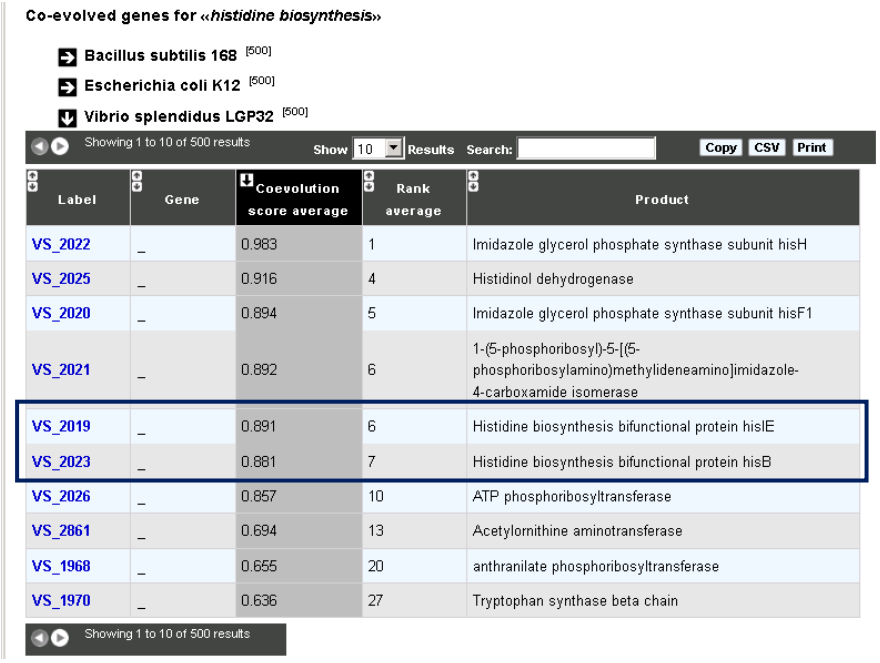
"*histidine biosynthesis*" MicroCyc Cross-species comparison
"*histidine biosynthesis*" MetaCyc pathway

Co-evolved genes for «*histidine biosynthesis*»

- **Bacillus subtilis 168** [500]
- **Escherichia coli K12** [500]
- **Vibrio splendidus LGP32** [500]

CLOSE

Using the Co-evolved genes in LGP32, find the candidate genes for missing enzymatic activities? Check blast results and synteny conservation to confirm your predictions: In the pathway table, we can display list of Co-evolved genes for histidine biosynthesis pathway in *V. splendidus* LGP32 in order to see if there are genes not initially associated to histidine biosynthesis but that shows significant co-evolution scores with other predicted genes of the pathway, indicative of a common evolutionary patterns of presence-absence across Microscope genomes that supports a common functional role:



Co-evolved genes for «*histidine biosynthesis*»

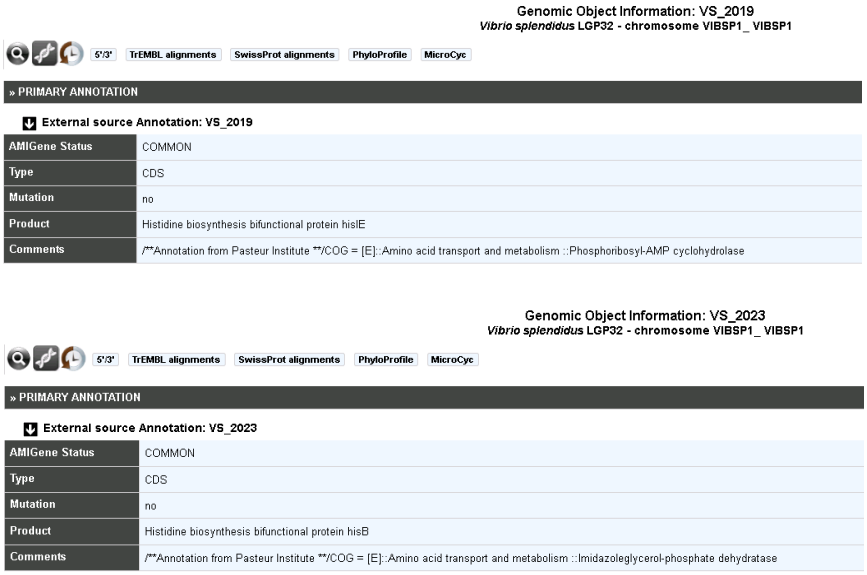
- **Bacillus subtilis 168** [500]
- **Escherichia coli K12** [500]
- **Vibrio splendidus LGP32** [500]

Showing 1 to 10 of 500 results | Show 10 Results | Search: | Copy | CSV | Print

Label	Gene	Coevolution score average	Rank average	Product
VS_2022	_	0.983	1	Imidazole glycerol phosphate synthase subunit hisH
VS_2025	_	0.916	4	Histidinol dehydrogenase
VS_2020	_	0.894	5	Imidazole glycerol phosphate synthase subunit hisF1
VS_2021	_	0.892	6	1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino]imidazole-4-carboxamide isomerase
VS_2019	_	0.891	6	Histidine biosynthesis bifunctional protein hisIE
VS_2023	_	0.881	7	Histidine biosynthesis bifunctional protein hisB
VS_2026	_	0.857	10	ATP phosphoribosyltransferase
VS_2861	_	0.694	13	Acetylornithine aminotransferase
VS_1968	_	0.655	20	anthranilate phosphoribosyltransferase
VS_1970	_	0.636	27	Tryptophan synthase beta chain

Showing 1 to 10 of 500 results

There are two CDS (VS_2019 and VS_2023) that are not initially associated to the projected pathway but that shows the fifth and sixth best co-evolution scores among *V. splendidus* LGP32 CDSs with the other genes of the pathway (Notice that 1st to 5th best co-evolution scores corresponds with CDS already associated to histidine biosynthesis pathway). These CDS are annotated as Histidine biosynthesis bifunctional proteins not linked to any EC number in Microscope annotations, and as consequence are not initially associated to any step of Histidine biosynthesis pathway:



Genomic Object Information: VS_2019
Vibrio splendidus LGP32 - chromosome VIBSP1_ VIBSP1

5'/3' | TrEMBL alignments | SwissProt alignments | PhyloProfile | MicroCyc

» PRIMARY ANNOTATION

External source Annotation: VS_2019

AMIGene Status	COMMON
Type	CDS
Mutation	no
Product	Histidine biosynthesis bifunctional protein hisIE
Comments	/**Annotation from Pasteur Institute **/COG = [E]::Amino acid transport and metabolism ::Phosphoribosyl-AMP cyclohydrolase

Genomic Object Information: VS_2023
Vibrio splendidus LGP32 - chromosome VIBSP1_ VIBSP1

5'/3' | TrEMBL alignments | SwissProt alignments | PhyloProfile | MicroCyc

» PRIMARY ANNOTATION

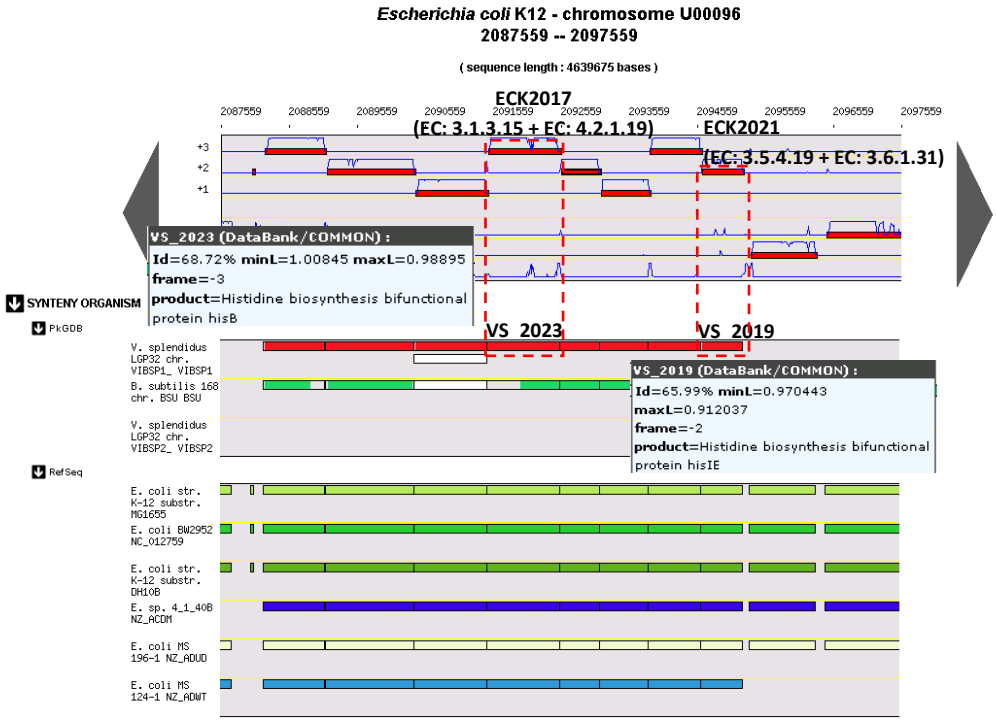
External source Annotation: VS_2023

AMIGene Status	COMMON
Type	CDS
Mutation	no
Product	Histidine biosynthesis bifunctional protein hisB
Comments	/**Annotation from Pasteur Institute **/COG = [E]::Amino acid transport and metabolism ::Imidazoleglycerol-phosphate dehydratase

By comparing with the projections in *E. coli* K12, where the pathway is completely projected, we can see that missing enzymatic reactions in *V. splendidus* LGP32 corresponds to bifunctional enzymes in *E. coli* K12:

- **ECK2021 (HisIE):** Fused phosphoribosyl-AMP cyclohydrolase ; phosphoribosyl-ATP pyrophosphatase (EC:3.5.4.19 + EC:3.6.1.31)
- **ECK2017 (HisB):** Fused histidinol-phosphatase ; imidazoleglycerol-phosphate dehydratase (EC: 3.1.3.15 + EC: 4.2.1.19)

The analysis of genomic context conservation between *E. coli* K12 and *V. splendidus* LGP32 by selecting *E. coli* K12 as pivot genome in Microscope genome browser and *V. splendidus* LGP32 and *B. subtilis* 168 in the synteny maps confirms the predictions of co-evolution scores. **The histidine biosynthesis operon of *E. coli* K12 is completely conserved in *V. splendidus* LGP32, with VS_2019 being the counterpart of ECK2021 (Bifunctional EC: 3.5.4.19 + EC: 3.6.1.31) and VS_2023 being the counterpart of ECK2017 (Bifunctional EC: 3.1.3.15 + EC: 4.2.1.19)**.



Ex. T-2: Check if CanOE predicts a metabolon covering histidine biosynthesis pathway in *V. splendidus* LGP32: Following the same procedure as in Ex. T-1 to compare the metabolic profile of Histidine biosynthesis pathway in V. splendidus LGP32 and E. coli K12 and display the pathway table to compare GPR assignments in both genomes:

Reactions in « *histidine biosynthesis* »

Showing 1 to 10 of 10 results Show 10 Results Search: Copy CSV Print

Reactions	EC Number(s)	Bacillus subtilis 168	Escherichia coli K12	Vibrio splendidus LGP32
ATPPHOSPHORIBOSYLTRANS-RXN: ATP phosphoribosyltransferase	2.4.2.17	BSU34920 BSU34930	ECK2014	VS_2026
GLUTAMIDOTRANS-RXN: 2.4.2.-_GLUTAMIDOTRANS-RXN	2.4.2.-	BSU34870 BSU34890	ECK2018 ECK2020	VS_2020 VS_2022
HISTALDEHYD-RXN: Histidinal dehydrogenase	1.1.1.23	BSU34910	ECK2015	VS_2025
HISTAMINOTRANS-RXN: histidinol-phosphate transaminase	2.6.1.9	BSU22620	ECK2016	VS_1152 VS_2024
HISTCYCLOHYD-RXN: phosphoribosyl-AMP cyclohydrolase	3.5.4.19	BSU34860	ECK2021	_
HISTIDPHOS-RXN: histidinol-phosphatase	3.1.3.15	BSU29620	ECK2017	_
HISTOLDEHYD-RXN: Histidinol dehydrogenase	1.1.1.23	BSU34910	ECK2015	VS_2025
HISTPRATPHYD-RXN: phosphoribosyl-ATP diphosphatase	3.6.1.31	BSU34860	ECK2021	_
IMIDPHOSDEHYD-RXN: imidazoleglycerol-phosphate dehydratase	4.2.1.19	BSU34900	ECK2017	_
PRIBFAICARPISOM-RXN: 1-(5-phosphoribosyl)-5-(5-phosphoribosylamino)methylideneaminoimidazole-4-carboxamide isomerase	5.3.1.16	BSU34880	ECK2019	VS_2021

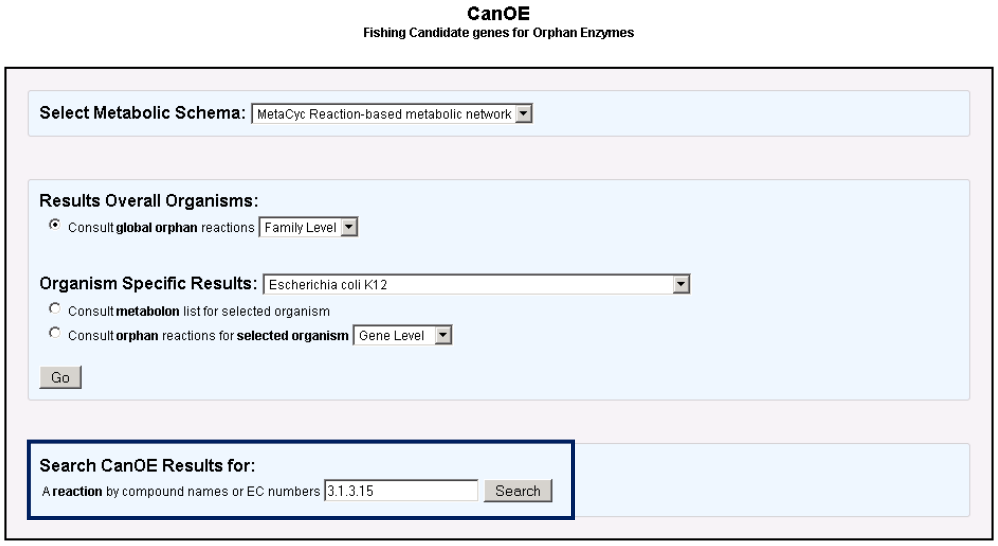
"*histidine biosynthesis*" MicroCyc Cross-species comparison
"*histidine biosynthesis*" MetaCyc pathway

Co-evolved genes for «*histidine biosynthesis*»

- **Bacillus subtilis 168** [500]
- **Escherichia coli K12** [500]
- **Vibrio splendidus LGP32** [500]

CLOSE

On order to find if CanOE predicts any metabolon in V. splendidus LGP32 covering histidine biosynthesis pathway, go to Metabolism → CanOE section of MicroScope main bar and search CanOE results by filtering on EC: 3.1.3.15:



Result of “search”:

CanOE Search

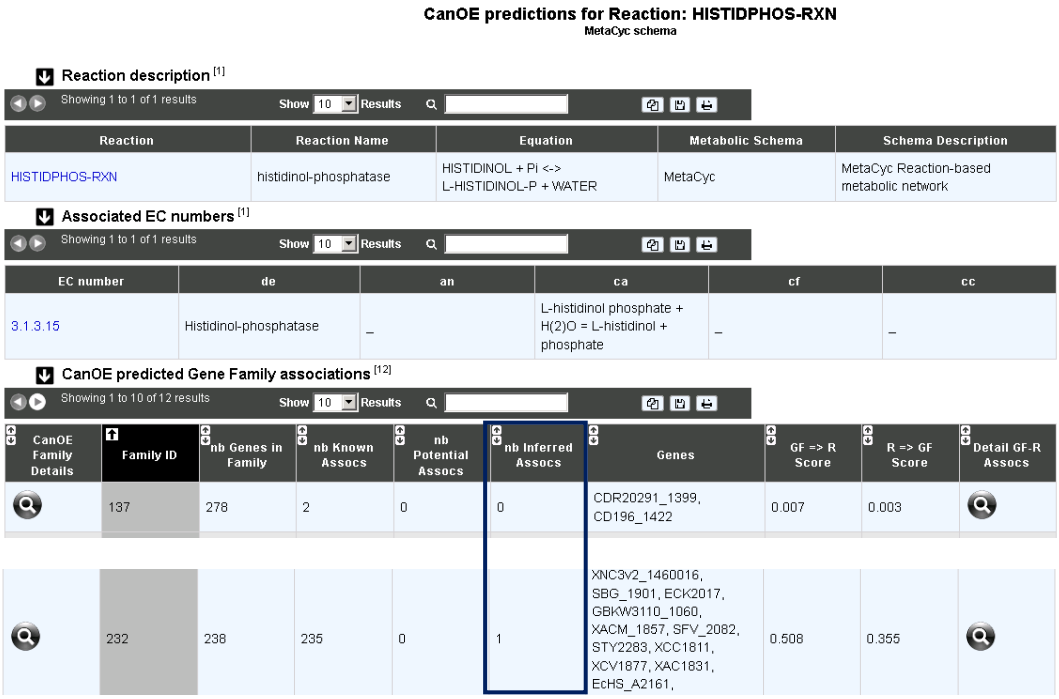
Reactions in CanOE metabolons [1]

Showing 1 to 1 of 1 results | Show 10 Results | Search: | Copy | CSV | Print

CanOE Reaction Details	Reaction	Metabolic Schema	EC number	Equation
	HISTIDPHOS-RXN	MetaCyc	3.1.3.15	HISTIDINOL + Pi <-> L-HISTIDINOL-P + WATER

Click on the icon glass ‘Canoe Reaction Details” and open “CanOE predicted gene associations”:

We find this reaction associated to 12 different families defined by the CanOE method. Within these 12 families, there is only one of them (family ID 232; 238 genes, 235 “known” associations) for which CanOE strategy predicts a single inferred association, meaning that this association is included within a predicted metabolon (conserved genomic and metabolic contexts)**:**



CanOE predictions for Reaction: HISTIDPHOS-RXN
MetaCyc schema

Reaction description [1]

Showing 1 to 1 of 1 results — Show 10 Results

Reaction	Reaction Name	Equation	Metabolic Schema	Schema Description
HISTIDPHOS-RXN	histidinol-phosphatase	HISTIDINOL + PI <-> L-HISTIDINOL-P + WATER	MetaCyc	MetaCyc Reaction-based metabolic network

Associated EC numbers [1]

Showing 1 to 1 of 1 results — Show 10 Results

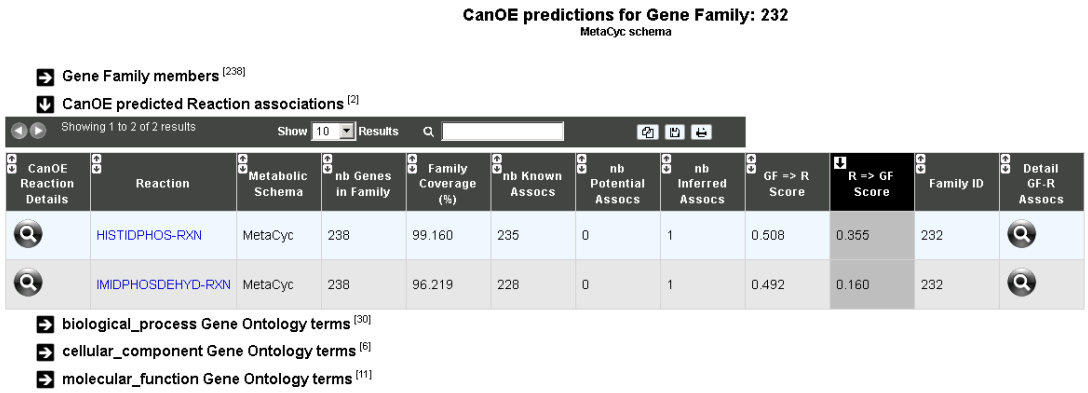
EC number	de	an	ca	cf	cc
3.1.3.15	Histidinol-phosphatase	–	L-histidinol phosphate + H(2)O = L-histidinol + phosphate	–	–

CanOE predicted Gene Family associations [12]

Showing 1 to 10 of 12 results — Show 10 Results

CanOE Family Details	Family ID	nb Genes in Family	nb Known Assocs	nb Potential Assocs	nb Inferred Assocs	Genes	GF => R Score	R => GF Score	Detail GF-R Assocs
	137	278	2	0	0	CDR20291_1399, CD196_1422	0.007	0.003	
	232	238	235	0	1	XNC3v2_1460016, SBG_1901, ECK2017, GBkW3110_1060, XACM_1857, SFV_2082, STY2283, XCC1811, XCV1877, XAC1831, EcHS_A2161,	0.508	0.355	

By clicking on the hyperlink of "CanOE family details" we can access the different reactions predicted for genes of this family included in metabolons. In this case, genes of this family are mainly associated to two different reactions that correspond to EC: 3.1.3.15 (HISTIDPHOS-RXN) and EC: 4.2.1.19 (IMIDPHOSDEHYD-RXN)



CanOE predictions for Gene Family: 232
MetaCyc schema

Gene Family members [238]

CanOE predicted Reaction associations [2]

Showing 1 to 2 of 2 results — Show 10 Results

CanOE Reaction Details	Reaction	Metabolic Schema	nb Genes in Family	Family Coverage (%)	nb Known Assocs	nb Potential Assocs	nb Inferred Assocs	GF => R Score	R => GF Score	Family ID	Detail GF-R Assocs
	HISTIDPHOS-RXN	MetaCyc	238	99.160	235	0	1	0.508	0.355	232	
	IMIDPHOSDEHYD-RXN	MetaCyc	238	96.219	228	0	1	0.492	0.160	232	

biological_process Gene Ontology terms [30]

cellular_component Gene Ontology terms [6]

molecular_function Gene Ontology terms [11]

By clicking on the hyperlink "*Detail GF-R assocs*" for both reactions, we see that both reactions has been inferred by CanOE strategy associated to the same CDSs, which corresponds to VS_2023 from **V. splendidus LGP32 (single Inferred association by CanOE strategy associated to Gene Family 232):**

CanOE associations between: IMIDPHOSDEHYD-RXN / Family 232
MetaCyc schema

CanOE predicted associations for reaction: IMIDPHOSDEHYD-RXN

CanOE predicted associations for family: 232

Inferred Gene-Reaction associations [1]

Showing 1 to 1 of 1 results Show 10 Results

Organism	Label	Product	R => GF Rank	GF => R Rank	View Metabolon
Vibrio splendidus LGP32	VS_2023	Histidine biosynthesis bifunctional protein hisB	–	–	

Known Gene-Reaction associations [228]

CanOE associations between: HISTIDPHOS-RXN / Family 232
MetaCyc schema

CanOE predicted associations for reaction: HISTIDPHOS-RXN

CanOE predicted associations for family: 232

Inferred Gene-Reaction associations [1]

Showing 1 to 1 of 1 results Show 10 Results

Organism	Label	Product	R => GF Rank	GF => R Rank	View Metabolon
Vibrio splendidus LGP32	VS_2023	Histidine biosynthesis bifunctional protein hisB	–	–	

Known Gene-Reaction associations [235]

⇨ In both cases, by clicking on **"View metabolon" icon, we access to the predicted metabolon in V. splendidus LGP32, that covers histidine biosynthesis operon including the single inferred association of VS_2023 with both HISTIDPHOS-RXN and IMIDPHOSDEHYD-RXN**

Metabolon Viewer: Metabolon 100885
MetaCyc schema

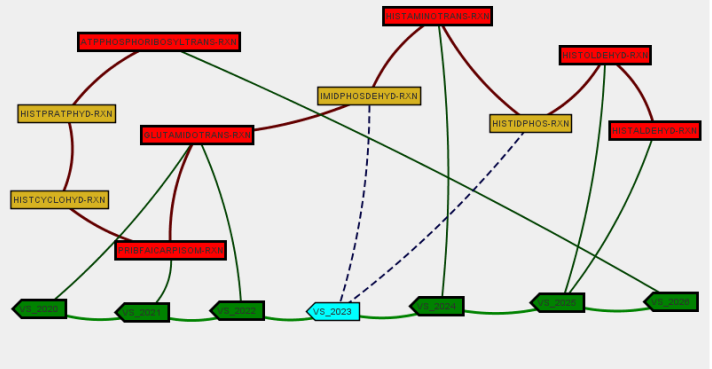
Consult Metabolon list for Vibrio splendidus LGP32

Note: Java Runtime Environment version >1.6.0 is required to view the metabolon

LEGEND

Gene colour	
Green	has associated reaction(s)
Cyan	metabolic gene with no associated reaction
Teal	gene of undetermined type
Gray	non-metabolic gene
Reaction colour	
Red	non orphan
Orange	local orphan
Yellow	global orphan
Gene-Reaction Edge colour	
Green	known association
Purple	potential association
Blue	inferred association

Mouse Mode



CanOE predicted Gene-Reaction associations [9]

Showing 1 to 9 of 9 results Show 10 Results

Association Type	Reaction	CanOE Reaction Details	Family ID	CanOE Family Details	Label	R => GF Score	R => GF Rank	GF => R Score	GF => R Rank
Inferred	HISTIDPHOS-RXN		232		VS_2023	0.355	_	0.508	_
Inferred	IMIDPHOSDEHYD-RXN		232		VS_2023	0.160	_	0.492	_
Known	ATPPHOSPHORIBOSYLTRANS-RXN		2092		VS_2026	0.041	_	1.000	_
Known	GLUTAMIDOTRANS-RXN		519		VS_2020	0.069	_	0.973	_

So using the CanOE result we are able to find a gene candidate for the histidinol-phosphatase (EC 3.1.3.15) and the imidazoleglycerol-phosphate dehydratase (4.2.1.19) => VS_2023 (Bifunctional enzyme as in E. coli).

In comparison with the results of Ex. T-1, VS_2019 is not included in the metabolon despite by analyzing genomic and metabolic context we found that corresponds to bifunctional phosphoribosyl-AMP cyclohydrolase ; phosphoribosyl-ATP pyrophosphatase (EC:3.5.4.19 + EC:3.6.1.31). This can be explained by the fact that this gene is located in the boundaries of the metabolon, which covers from VS_2020 to VS_2026. In this context, CanOE requires that the metabolon should be flanked by known gene-reaction associations.

In comparison, in E. coli K12, where all gene-reaction associations are known and the genes are co-localized in an operonic structure, the predicted metabolon covers the whole histidine biosynthesis pathway:

Reactions in « *histidine biosynthesis* »

Showing 1 to 10 of 10 results — Show 10 Results — Search: — Copy CSV Print

Reactions	EC Number(s)	Escherichia coli K12
ATPPHOSPHORIBOSYLTRANS-RXN: ATP phosphoribosyltransferase	2.4.2.17	ECK2014
GLUTAMIDOTRANS-RXN: 2.4.2.-_GLUTAMIDOTRANS-RXN	2.4.2.-	ECK2020 ECK2018
HISTALDEHYD-RXN: Histidinal dehydrogenase	1.1.1.23	ECK2015
HISTAMINOTRANS-RXN: histidinol-phosphate transaminase	2.6.1.9	ECK2016
HISTCYCLOHYD-RXN: phosphoribosyl-AMP cyclohydrolase	3.5.4.19	ECK2021
HISTIDPHOS-RXN: histidinol-phosphatase	3.1.3.15	ECK2017
HISTOLDEHYD-RXN: Histidinol dehydrogenase	1.1.1.23	ECK2015
HISTPRATPHYD-RXN: phosphoribosyl-ATP diphosphatase	3.6.1.31	ECK2021
IMIDPHOSDEHYD-RXN: imidazoleglycerol-phosphate dehydratase	4.2.1.19	ECK2017
PRIBFAICARPISOM-RXN: 1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneaminoimidazole-4-carboxamide isomerase	5.3.1.16	ECK2019

"histidine biosynthesis" MicroCyc pathway
"histidine biosynthesis" MetaCyc pathway

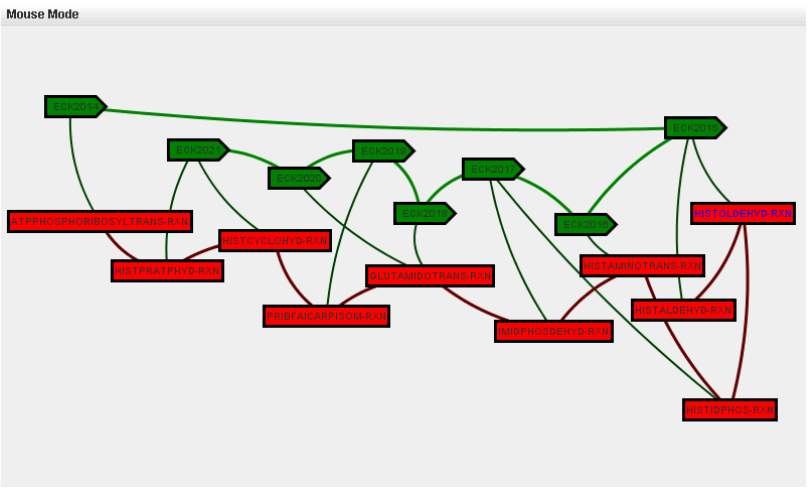
Metabolon Viewer: Metabolon 7477
MetaCyc schema

Consult Metabolon list for Escherichia coli K12

Note: Java Runtime Environment version >=1.6.0 is required to view the metabolon

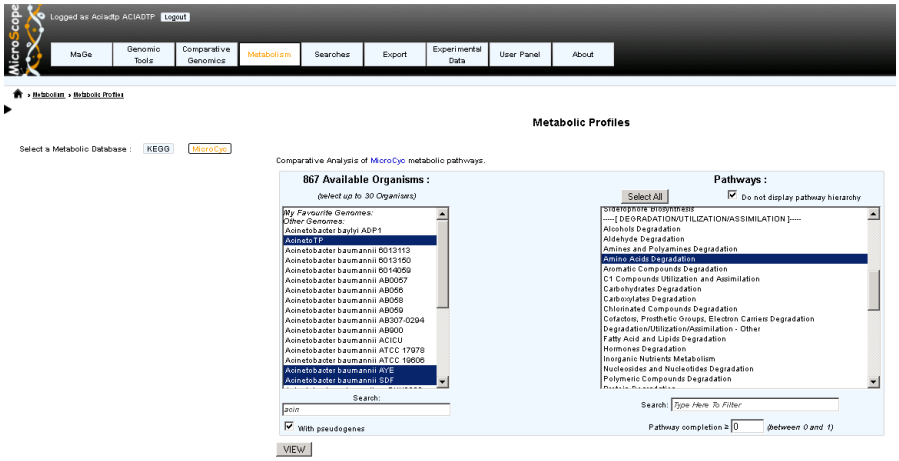
LEGEND

Gene colour	
Green	has associated reaction(s)
Cyan	metabolic gene with no associated reaction
Teal	gene of undetermined type
Gray	non-metabolic gene
Reaction colour	
Red	non orphan
Orange	local orphan
Yellow	global orphan
Gene-Reaction Edge colour	
Green	known association
Purple	potential association
Blue	inferred association



<u>Ex. A-1</u>: Expert annotation of histidine degradation I pathway: Compare the metabolic profile of this pathway in AcinetoTP with *A. baumannii* AYE and *A. baumannii* SDF

In the Metabolism section of Microscope main bar, select Metabolic profiles. In the Metabolic profiles page, select the genomes of interest and the pathways you want to compare. In the context of this exercise, we select Amino Acids Degradation, which includes histidine degradation pathway



By clicking on the VIEW button, is displayed a table with rows representing pathways, columns representing genomes, and cells representing the completion index of the pathway in a given genome calculated as the ratio between number of projected reactions and total number of reactions in the pathway:

Showing 1 to 34 of 34 results | Show All Results | Search: | Copy | CSV | Print

Pathways [34]	Reactions nb	Acinetobacter baumannii AYE	Acinetobacter baumannii SDF	AcinetoTP
2-ketoglutarate dehydrogenase complex	3	1	1	1
alanine degradation I	2	1	1	1
alanine degradation III	1	0	0	1
arginine degradation II (AST pathway)	5	1	1	1
arginine degradation III (arginine decarboxylase/agmatinase pathway)	2	0	0	0.50
asparagine degradation I	1	1	1	1
aspartate degradation II	2	1	1	1
beta-alanine degradation I	2	1	0.50	0.50
beta-alanine degradation II	2	1	0	0.50
citrulline degradation	2	0.50	0.50	0.50
D-serine degradation	1	1	0	1
glutamate degradation II	2	1	0.50	1
glutamate degradation X	1	1	1	0
glutamine degradation I	1	1	1	1
glutamine degradation II	1	1	1	1
glycine cleavage complex	3	1	1	1
histidine degradation I	4	1	1	0.25
isoleucine degradation I	6	0.50	0.50	0.50
L-cysteine degradation II	1	1	1	1

Histidine degradation pathway; incomplete in AcinetoTP (1/4 reactions projected)

By clicking in the histidine degradation I hyperlink, a new window is displayed with the details of the pathway reactions and the corresponding CDS in the analysed genomes. We can see that only ACIAD0574 has been predicted in AcinetoTP associated to EC: 4.3.1.3 reaction (histidine ammonia-lyase). In contrast, the pathway is complete in *A. baumannii* AYE and *A. baumannii* SDF, with predicted genes co-linear in the same chromosome region:

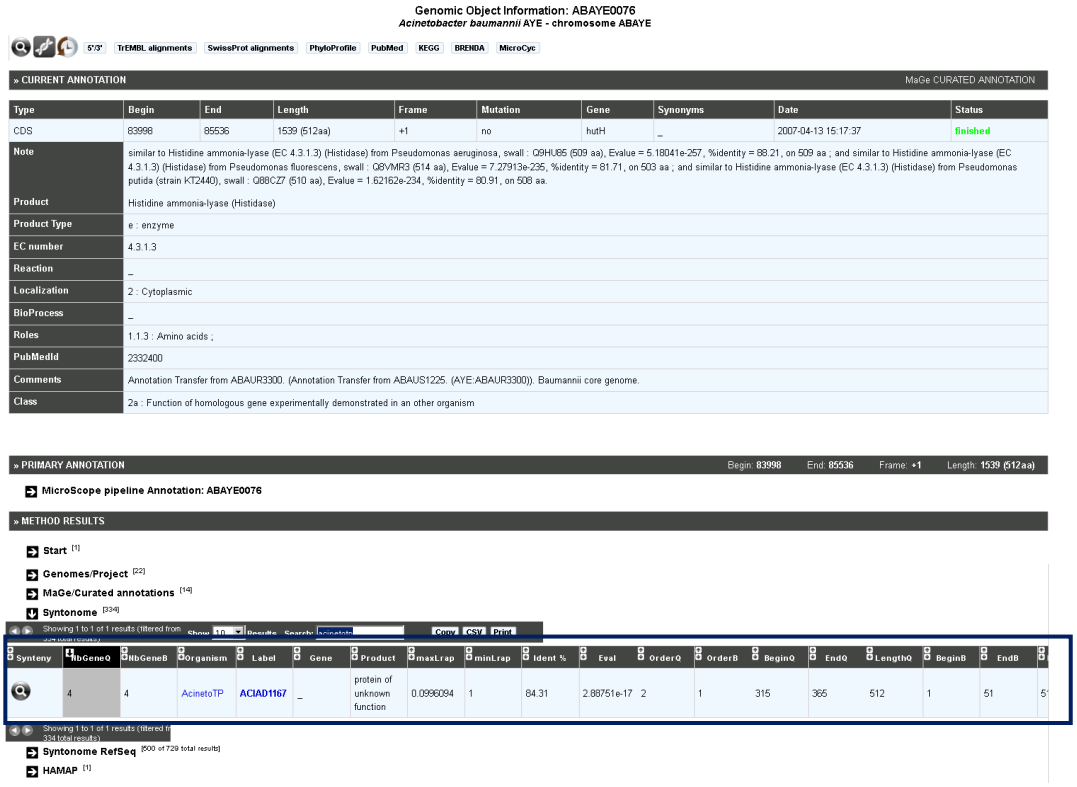
Reactions in « *histidine degradation I* »

Showing 1 to 4 of 4 results Show 10 Results Search: Copy CSV Print

Reactions	EC Number(s)	Acinetobacter baumannii AYE	Acinetobacter baumannii SDF	AcinetoTP
FORMIMINOGLUTAMASE-RXN: formimidoylglutamase	3.5.3.8	ABAYE0079	ABSDF3580	_
HISTIDINE-AMMONIA-LYASE-RXN: histidine ammonia-lyase	4.3.1.3	ABAYE0076	ABSDF3583	ACIAD0574
IMIDAZOLONEPROPIONASE-RXN: imidazolonepropionase	3.5.2.7	ABAYE0078	ABSDF3581	_
UROCANATE-HYDRATASE-RXN: urocanate hydratase	4.2.1.49	ABAYE0075	ABSDF3584	_

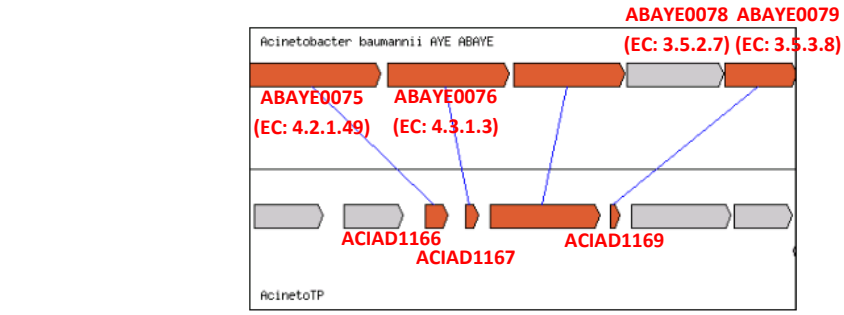
"*histidine degradation I*" MicroCyc Cross-species comparison
"*histidine degradation I*" MetaCyc pathway

Look at the conserved synteny between AcinetoTP and AYE. Is the histidine degradation operon conserved in AcinetoTP? In the pathway table, click on ABAYE0076 hyperlink to access to Microscope gene annotation interface. Go to the section *Syntonome* to explore the genome context conservation between A. baumannii AYE and AcinetoTP. Filter the Syntonome table by searching for AcinetoTP:



There is a synton of 4 orthologous genes between both genomes with a conserved genomic organization. By clicking on the glass icon, a graphical representation of the synton is displayed, where we can see that the histidine degradation operon in AcinetoTP is conserved but apparently highly degraded, with AcinetoTP genes that are shorter than *A. baumannii* AYE counterparts, consequence of a process of pseudogenization (gene inactivation):

Synton #231_1321_6850_6851



Acinetobacter baumannii AYE ABAYE

GO_label	GO_gene_name	GO_type	GO_product
ABAYE0075	hutU	CDS	urocanate hydratase (Urocanase) (Imidazolonepropionate hydrolase)
ABAYE0076	hutH	CDS	Histidine ammonia-lyase (Histidase)
ABAYE0077		CDS	putative histidine transport protein (APC family)
ABAYE0078	hutI	CDS	Imidazolonepropionase (Imidazolone-5-propionate hydrolase)
ABAYE0079	hutG	CDS	formimidoylglutamase (Formiminoglutamase) (Formiminoglutamate hydrolase)

Correspondences

ident	matchlength	minlrap	length_1	length_2	order_1	order_2
59.55	89	0.946809	558	94	1	1
84.31	51	1	512	51	2	1
91.8	451	0.980435	470	460	1	1
62.16	37	0.925	343	40	1	1

AcinetoTP

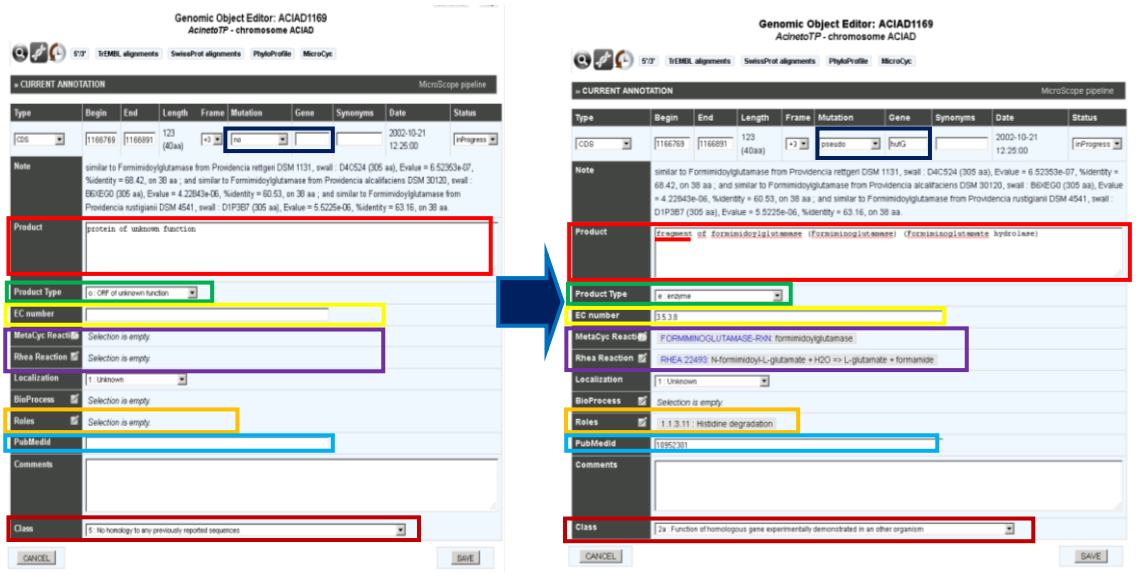
GO_label	GO_gene_name	GO_type	GO_product
ACIAD1164	hchA	CDS	Hsp31 molecular chaperone
ACIAD1165		CDS	putative transcriptional regulator
ACIAD1166		CDS	protein of unknown function
ACIAD1167		CDS	protein of unknown function
ACIAD1168	proY	CDS	proline transporter
ACIAD1169		CDS	protein of unknown function
ACIAD1170		CDS	protein of unknown function
ACIAD1171		CDS	protein of unknown function
ACIAD1172		CDS	DSBA oxidoreductase

BLASTP analysis of A. baumannii AYE against AcinetoTP proteome confirms the pseudogenization of AcinetoTP genes involved in histidine degradation pathway. For example, BLASTP of ABAYE0075 (EC: 4.2.1.49; 558 amino acids length) against AcinetoTP proteome:

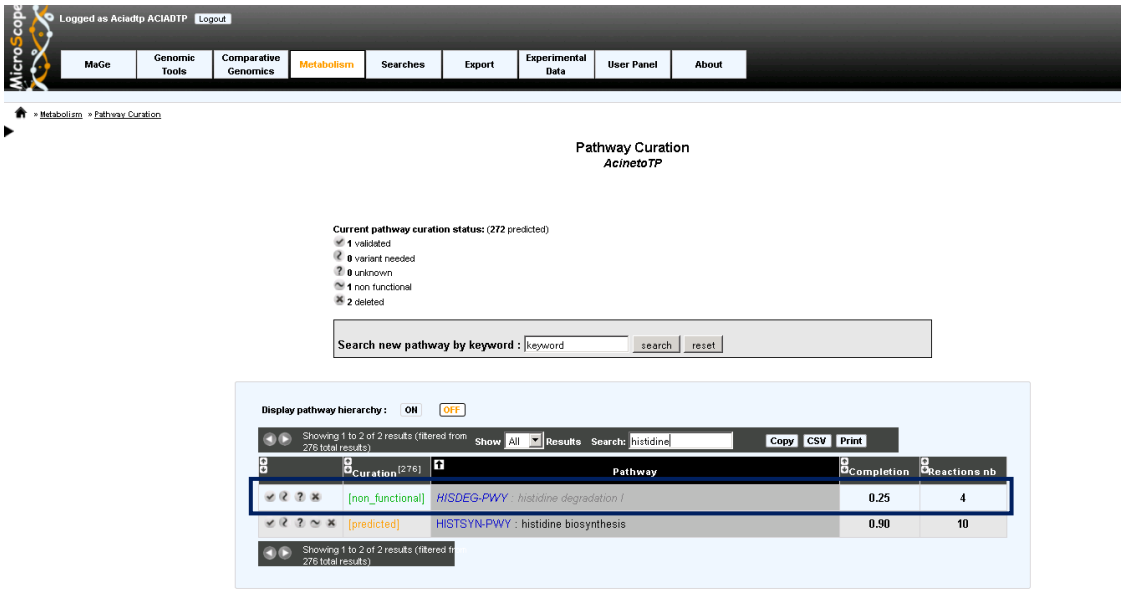
```
Number of Results: 11.

 - 1 - ACIAD1166 [AcinetoTP chromosome ACIAD]
       protein of unknown function | automatic/finished

      1      89     ABAYE0075      429     517     S= 263 E= 4.61014e-24 Bits= 105.916 LD= 94      LQ= 558 L= 89 Id= 59.55 % Pos= 71.91 %

Q:      429  MVRSGELSAPIVIGRDHLDSGSVASPNRETEAMQDGSDAVSDWPLLNALLNTAGGATWVS  488
             |||+ ||| ||+|||||||| | | | ||| |+||  +|  |+  | | ||+  |||  +|
S:        1  MVRNRELSDPILIGRDHLDFGLVTSLNRREIESMQADADTASNLTLFNTLLSVTGGAACIS  60

Q:      489  LHHGGGVGMGFSQHSGVVIVCDGTDEAAA  517
             +|||||||| |||   || +||||+|| +
S:       61  VHHGGGVGMKFSQCFEVVTICDGTEEAVS  89

 - 2 - ACIAD2725 [AcinetoTP chromosome ACIAD]
       nemA | N-ethylmaleimide reductase, FMN-linked | 1.-.-.- | automatic/finished

     24      103     ABAYE0075      397     493     S= 60 E= 1.59555 Bits= 27.7202 LD= 357  LQ= 558 L= 97 Id= 22.68 % Pos= 38.14 %

Q:      397  MARERISFQGLPARICWVGLGLRAKLGLAFNEMVRSGELSAPIVIGRDHLDSGSVASPNR  456
             | | |  | | |+ +       || +||   |  +  |              | +|+|
S:       24  MTRSRAQFDGTPSELAAEYYTQRASVGLIIAEGAQPSEDG----------QGYLATPGI  72

Q:      457  ETEAMQDGSDAVSDWPLLNALLNTAGGATWVSLHHGG  493
             |++       ++ |  +    ++   ||  ++ | | |
S:       73  YTDSH------IAGWKKITNAIHAKGGRIFIQLMHVG  103

 - 3 - ACIAD3479 [AcinetoTP chromosome ACIAD]
       waaA | 3-deoxy-D-manno-2-octulosonate transferase | automatic/finished

    320      384     ABAYE0075      413     468     S= 60 E= 1.59555 Bits= 27.7202 LD= 427  LQ= 558 L= 65 Id= 30.77 % Pos= 47.69 %

Q:      413  WVGLGLRAKLGLAFNE------MVRSGELSAPIVIGRDHLDSGSVASPNRETEAM---QD  463
             |  |   | +| + ||      ++    |  |+|||  + +  ++    |||||+   ||
S:      320  WYALAQTAFVGGSLNEPGGGHNILEPMALKIPVVIGPRYFNFQTIVDEFRETEAIDIVQD  379

Q:      464  GSDAV  468
              + ||
S:      380  ATQAV  384
```

Homology of ACIAD1166 (whole protein length) with C-terminal part of ABAYE0075

Based on these evidences, we can update annotation of ACIAD1166, ACIAD1167 and ACIAD1169 as pseudogenes of the corresponding orthologs in A. baumannii AYE, transferring the corresponding annotations from A. baumannii AYE to AcinetoTP CDSs and validating the corresponding MetaCyc and RHEA reactions, which can be retrieved by EC number synchronization. For example, for ACIAD1169, which corresponds to the pseudogenized form of ABAYE0079 (formimidoylglutamase; EC: 3.5.3.8):



Based on the results obtained, is this pathway functional in AcinetoTP? Validate the pathway accordingly by using the Pathway validation interface of Microscope: According to the previous results, we can conclude that histidine degradation pathway is non-functional in AcinetoTP consequence of a process of pseudogenization. As consequence, a "non-functional" status can be assigned in the Pathway curation interface of Microscope:



<u>Ex. A-2</u>: Expert annotation of glucarate/galactarate degradation pathway: Compare the metabolic profile of glucarate/galactarate degradation pathways in AcinetoTP with *E. coli* K12

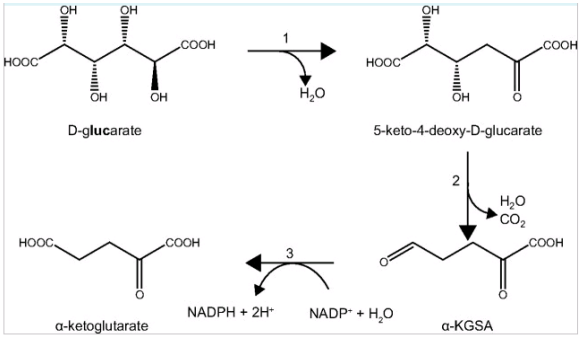
In the Metabolism section of Microscope main bar, select Metabolic profiles. In the Metabolic profiles page, select the genomes of interest and the pathways you want to compare. In the context of this exercise, we select Secondary Metabolites Degradation, which includes the

pathways for degradation of naturally occurring dicarboxylates D-glucarate and D-galactarate. We can see that there are two possible pathway variants for the degradation of each compound:

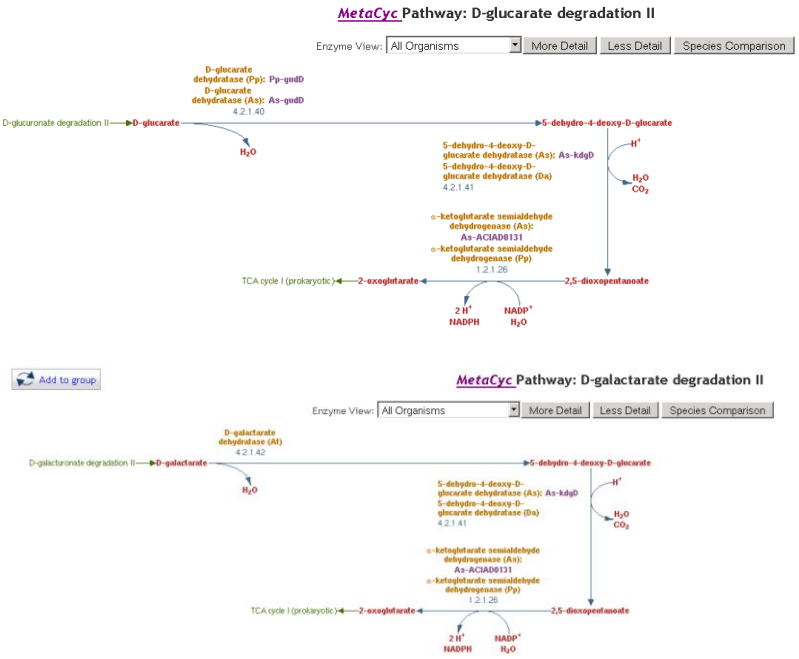
Secondary Metabolites Degradation[21]	Reactions nb	AcinetoTP	Escherichia coli K12
1,6-anhydro-*N*-acetylmuramic acid recycling	2	0.50	1
5-dehydro-4-deoxy-D-glucuronate degradation	4	0.25	0.75
curcumin degradation	2	0	1
D-galactarate degradation I	4	0.50	1
D-galactarate degradation II	3	0.67	0.33
D-galactonate degradation	3	0	1
D-galacturonate degradation I	4	0	1
D-glucarate degradation I	4	0.50	1
D-glucarate degradation II	3	0.67	0.33
fructoselysine and psicoselysine degradation	3	0	1
galactitol degradation	3	0	1
glucose degradation (oxidative)	5	0.40	0.20
ketogluconate metabolism	8	0.63	0.75
L-galactonate degradation	1	0	1
L-idonate degradation	3	0.67	1
mannitol degradation I	1	0	1

Refine the annotation of the CDSs ACIAD0126 to 0131 by using MicroScope tools and the publication PMID: 18364348: In this reference, the pathway of D-glucarate biosynthesis is experimentally verified in *A. baylyi* ADP1, demonstrating that it proceeds in a different manner as in *E. coli* K12. Whereas *E. coli* K12 degrades D-glucarate to 3-phosphoglycerate in 4 enzymatic steps (corresponds to D-glucarate degradation I pathway variant), *A. baylyi* ADP1 degrades D-glucarate to 2-oxoglucarate in 3 enzymatic steps, which corresponds to D-glucarate degradation II pathway variant:

Figure 1
PMID: 18364348



In the same reference, knockout experiments in A. baylyi ADP1 identifies a cluster of genes comprising from ACIAD0126 to ACIAD0131 that are responsible of the transport and degradation of D-glucarate and D-galactarate, that is carried out by specific dehydratases for D-glucarate (EC:4.2.1.40) and D-galactarate (EC:4.2.1.42) that produces the common intermediate 5-dehydro-4-deoxy-D-glucarate, which is subsequently converted in 2-oxoglucarate in two reaction steps catalyzed by EC:4.2.1.41 and EC:1.2.1.26. This metabolic profile is represented in MetaCyc in two different pathway variants that correspond to D-glucarate degradation II and D-galactarate degradation II:



By comparing the metabolic profile of AcinetoTP for both pathway variants, in both cases there is 1 pathway hole that corresponds to EC: 1.2.1.26 activity:

Reactions in « *D-galactarate degradation II* »

Showing 1 to 3 of 3 results | Show 10 Results | Search: | Copy | CSV | Print

Reactions	EC Number(s)	AcinetoTP	Escherichia coli K12
4.2.1.41-RXN: 5-dehydro-4-deoxyglucarate dehydratase	4.2.1.41	ACIAD0130	_
25-DIOXOVALERATE-DEHYDROGENASE-RXN: 2,5-dioxovalerate dehydrogenase	1.2.1.26	_	_
GALACTARDEHYDRA-RXN: galactarate dehydratase	4.2.1.42	ACIAD0126	ECK3116

D-galactarate degradation II MicroCyc Cross-species comparison
D-galactarate degradation II MetaCyc pathway

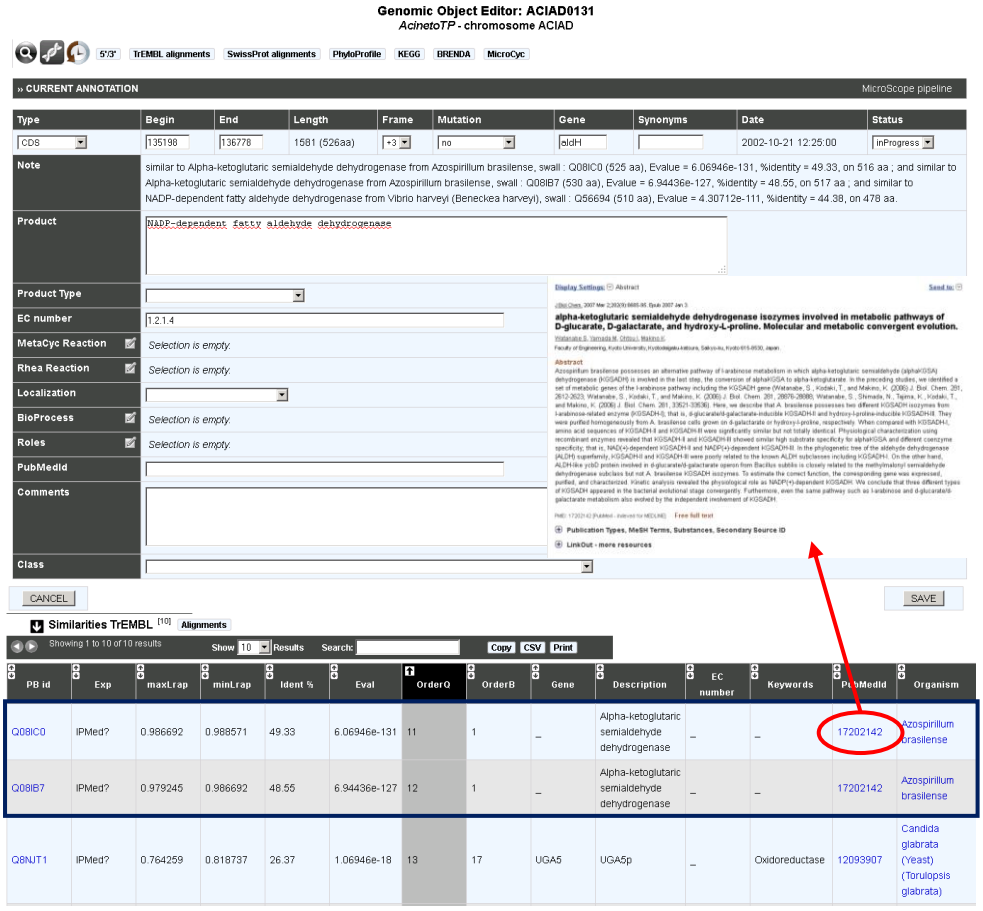
Reactions in « *D-glucarate degradation II* »

Showing 1 to 3 of 3 results | Show 10 Results | Search: | Copy | CSV | Print

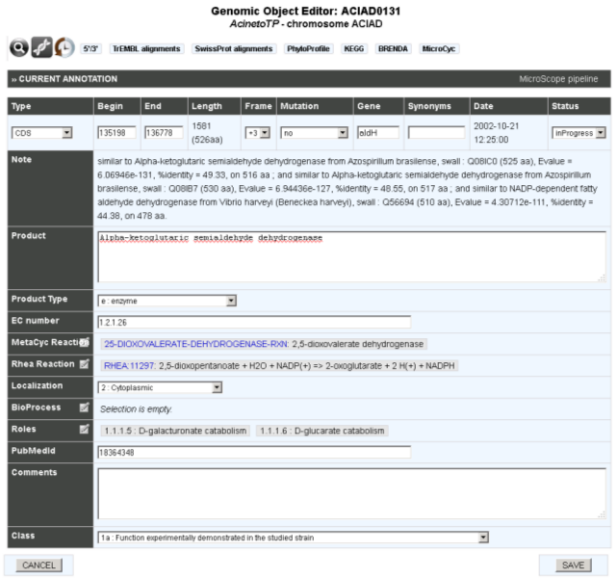
Reactions	EC Number(s)	AcinetoTP	Escherichia coli K12
4.2.1.41-RXN: 5-dehydro-4-deoxyglucarate dehydratase	4.2.1.41	ACIAD0130	_
25-DIOXOVALERATE-DEHYDROGENASE-RXN: 2,5-dioxovalerate dehydrogenase	1.2.1.26	_	_
GLUCARDEHYDRA-RXN: glucarate dehydratase	4.2.1.40	ACIAD0128	ECK2782 ECK2781

D-glucarate degradation II MicroCyc Cross-species comparison
D-glucarate degradation II MetaCyc pathway

In the same reference is experimentally verified that 2-ketoglutarate semialdehyde dehydrogenase enzyme corresponding to EC: 1.2.1.26 is encoded by the gene ACIAD0131. This gene is annotated in MicroScope as NADP-dependent fatty aldehyde dehydrogenase associated to EC: 1.2.1.4. However, by analyzing the similarity profiles in BLASTP searches against SwissProt/TrEMBL entries with experimental evidence of their annotation, we find that the best hits in TrEMBL searches are against two *Azospirillum brasilense* entries annotated as Alpha-ketoglutaric semialdehyde dehydrogenase, with experimental evidence of this activity, although without EC number association.



Based on these evidences, we should update the annotation of ACIAD0131 to Alpha-ketoglutaric semialdehyde dehydrogenase associated to EC: 1.2.1.26. The corresponding reactions in MetaCyc and RHEA should be also validated and the PubMed ID of the publication where the activity in A. baylyi ADP1 has been experimentally verified (PMID: 18364348):



Based on the results obtained, validate the 4 projected pathways accordingly by using the Pathway validation interface of Microscope: Accordingly with the evidences reported, we should assign a *"Deleted"* status to *D-glucarate degradation I* and *D-galactarate degradation I*

pathway variants (*E. coli* K12 variants not functional in AcinetoTP) in MicroScope Pathway Curation interface. In a similar manner, *D-glucarate degradation II* and *D-galactarate degradation II* should be assigned a *"Validated"* status (functional pathways in AcinetoTP).