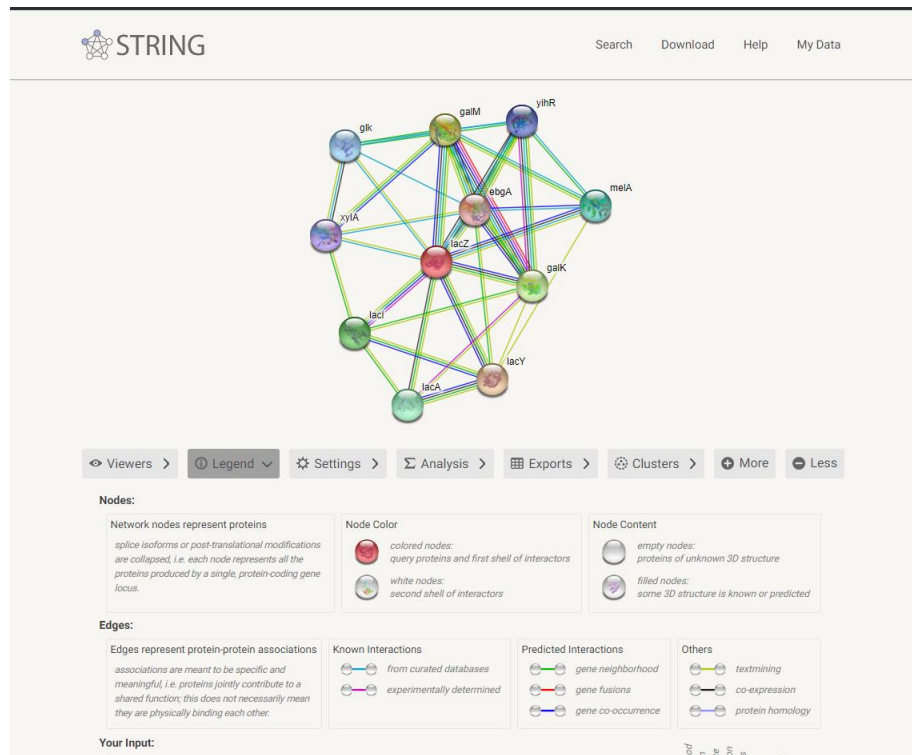


BCB 570 Assignment: Functional Annotation

-Urminder Singh

Sol1.



Sol2.

The gene lacZ encodes for β -galactosidase which is an enzyme which cleaves disaccharide lactose into glucose and galactose. lacZ is part of the lac operon which consists of other two genes as well. These genes are

lacY encodes for Beta-galactoside permease (LacY), a transmembrane symporter that pumps β -galactosides including lactose into the cell.

lacA encodes β -galactoside transacetylase (LacA), an enzyme that transfers an acetyl group from acetyl-CoA to β -galactosides.

As lacY and lacA are in the same operon as lacZ they are highly associated with each other.

According to STRING, lacZ seems to be associated with the following ten genes:

lacY: Lactose permease; Responsible for transport of beta-galactosides into the cell,

galM Aldose 1-epimerase; Mutarotase converts alpha-aldose to the beta-anomer.

galK Galactokinase; Catalyzes the transfer of the gamma-phosphate of ATP to D-galactose to form alpha-D-galactose-1-phosphate

lacI DNA-binding transcriptional repressor; Repressor of the lactose operon.

lacA Thiogalactoside acetyltransferase; May assist cellular detoxification by acetylating non-metabolizable pyranosides,

melA Alpha-galactosidase, NAD(P)-binding (451 aa)

glk Glucokinase; Not highly important in E. coli as glucose is transported into the cell by the PTS system

yihR Predicted aldose-1-epimerase (308 aa)


xylA D-xylose isomerase (440 aa)


It looks like STRING associates more than two genes to lacZ. This is because other genes are associated with lacZ via other pathways. E.g. galM and galK are associated with lacZ as these are, broadly, involved in galactose metabolism. Other criterion for association by STRING is by using text mining, co-expression, protein homology, gene fusions, experimentally determined etc. Therefore, using all these different criterion STRING is able to associate more genes to lacZ.

Sol3.

According to KEGG lacZ appears in following pathways:

eco00052	Galactose metabolism
eco00511	Other glycan degradation
eco00600	Sphingolipid metabolism
eco01100	Metabolic pathways


Escherichia coli K-12 MG1655: b0344
Help

Entry	b0344	CDS	T00007
Gene name	lacZ		
Definition	(RefSeq) beta-D-galactosidase KO K01190 beta-galactosidase [EC:3.2.1.23]		
Organism	eco Escherichia coli K-12 MG1655		
Pathway	eco00052 Galactose metabolism eco00511 Other glycan degradation eco00600 Sphingolipid metabolism eco01100 Metabolic pathways		
Brite	KEGG Orthology (KO) [BR:eco00001] Metabolism Carbohydrate metabolism 00052 Galactose metabolism b0344 (lacZ) Lipid metabolism 00600 Sphingolipid metabolism b0344 (lacZ) Glycan biosynthesis and metabolism 00511 Other glycan degradation b0344 (lacZ) Enzymes [BR:eco01000] 3. Hydrolases 3.2. Glycosylases 3.2.1. Glycosidases, i.e. enzymes that hydrolyse O- and S-glycosyl co 3.2.1.23 beta-galactosidase b0344 (lacZ)		
<div> <div>←</div> <div>→</div> </div> <div>BRITE hierarchy</div>			
SSDB	<div> <div>Ortholog</div> <div>Paralog</div> <div>Gene cluster</div> <div>GFIT</div> </div>		
Motif	Pfam: Glyco_hydro_2_C Glyco_hydro_2_N Bgal_small_N DUF4981 Glyco_hydro_2 <div>Motif</div>		
Other DBs	NCBI-GeneID: 945006 NCBI-ProteinID: NP_414878 Pasteur: lacZ RegulonDB: ECK120000520 ECOCYC: EG10527 ASAP: ABE-0001183 UniProt: P00722 G0ZKW2		
Structure	<div> <div> PDB: 1JZ7 1JZ8 1JYN 4TTG 1PX3 1PX4 1DP0 3T09 3DYP 1JYV 1JYX 1JZ3 3DYO 1JYN 1JZ5 3T20 3T0A 3MUZ 3T0D 3VD4 3T08 3IAP 3VD9 3VDB 3SEP 3DYM 3CZJ 4DUV 3I3E 1JZ2 1JZ4 1JZ6 5A1A 4DUW 3MV0 3MV1 3I3B 3I3D 4DUX 3T2Q 3T0B 3VDA 3MUJ 3VDC 3T2P 3VD5 3IAQ 3VD3 1F4A 1F4H 3VD7 3E1F 1HN1 3J7H 4CKD </div> <div>  <div>Jmol</div> </div> </div> <div>Thumbnail</div>		
Position	complement(363231..366305)		

All links

- Ontology (2)
- KEGG BRITE (2)
- Pathway (4)
- KEGG PATHWAY (4)
- Chemical substance (11)
- KEGG COMPOUND (11)
- Chemical reaction (6)
- KEGG ENZYME (1)
- KEGG REACTION (5)
- Genome (1)
- KEGG GENOME (1)
- Gene (28)
- KEGG ORTHOLOGY (1)
- RefGene (17)
- NCBI-PROTEINID (1)
- NCBI-Gene (1)
- ASAP (1)
- ECOCYC (1)
- OC (1)
- PASTEUR (1)
- PROTHERM (3)
- REGULONDB (1)
- Protein sequence (4)
- UniProt (2)
- SWISS-PROT (1)
- RefSeq(pep) (1)
- Protein domain (5)
- Pfam (5)
- All databases (61)

Download RDF

After searching the associated genes in these pathways, I found that there are many other genes present than listed by STRING in the lacZ associated pathways except lacA and lacY were not present in the pathways. This may be because lacA and lacY may have an indirect role in the lacZ associated pathways.

Sol 4

I repeated the above experiment for lacY and lacA.

I found that lacA and lacY are not associated with any pathway according to kegg. We can try to answer why, based on the function of these genes. lacY is involved in coding for a protein galactoside permease which controls flow of lactose molecules into the cell. Thus, it is not directly involved in breaking lactose into glucose and hence is not associated with the lacZ pathway.

lacA codes for a protein β -galactoside transacetylase which is an enzyme that transfers an acetyl group from acetyl-CoA to galactosides but not to the beta-galactosidase which is coded by lacZ. Further, the precise role of lacA's product is not clearly understood in the lac operon. This shows that lacA is not directly involved in any pathway and thus is absent from lacZ's pathways.

Sol 5.

ECOCYC
A member of the KEGG database collection

LOGIN | Iowa State University subscriber | Help | Create New Account

Enter a gene, protein, metabolite or pathway: Quick Search | Gene Search
Searching *Escherichia coli* K-12 substr. MG1655 (ECOCYC) change organism database

Sites Search Genome Metabolism Analysis SmartTables Help

gene lacZ <i>Escherichia coli</i> K-12 substr. MG1655	enzyme β-galactosidase	Add to SmartTable
Synonyms: β -D-galactoside galactohydrolase, β -D-galactosidase, lactase		
Accession IDs EG10527 (EcoCyc) b0144 ECK0341 P00722 (UniProt)	Length: 3075 bp / 1024 aa Map Position: [363,231 <- 366,305] (7.83 centisomes, 28°) View in Genome Browser Location: cytosol	
Reactions: $\text{lactose} + \text{H}_2\text{O} \rightarrow \beta\text{-D-galactose} + \text{D-glucopyranose}$ $\alpha\text{-lactose} \leftrightarrow \text{allolactose}$ $\text{lactulose} + \text{H}_2\text{O} \rightarrow \beta\text{-D-galactose} + \text{D-fructofuranose}$ $3\text{-O-}\beta\text{-D-galactopyranosyl-}\beta\text{-arabinose} + \text{H}_2\text{O} \rightarrow \beta\text{-D-galactose} + \text{D-arabinose}$		
Pathway: lactose degradation III		
Evidence: Inferred by functional complementation [Hall78] Inferred from experiment [Jobe72] Assay of partially-purified protein [Burstine65, Lederberg50] Assay of protein purified to homogeneity [Steers67, Craven65]		

OPERATIONS

Sequences

- Protein Sequence
- Nucleotide Sequence
- Save Nucleotide Sequence to file
- Save Protein Sequence to file

Comparison Operations

- Show this gene in another database
- Change organisms/databases for comparison operations
- Search for this gene in other databases
- Show orthologs (with operon diagrams) in multiple databases
- Align in Multi-Genome Browser
- Align gene nucleotide sequence with orthologs
- Align gene product amino acid

Sol6.

I repeated 3 to 5 with GalK

According to kegg, GalK is involved in following metabolic pathways

eco00052	Galactose metabolism
eco00520	Amino sugar and nucleotide sugar metabolism
eco01100	Metabolic pathways

Genes involved in Galactose metabolism are: GalM, lacZ, galT, galU, melA, yihR etc

Genes involved in Amino sugar and nucleotide sugar metabolism are: galT, glf, galT, arnC, arnD etc.

It is interesting that galM, galK and galT all are in the gal operon. Again, this has other genes associate which are not present in STRING. And genes which are present only in STRING. This is because STRING has different criterion to associate genes one of which is looking ate the biological pathways but it only reports strong associations.

Then, I found galK in ecocyc too

Next, From STRING I found that there are 7 publications in PubMed which co-mentions lacZ and xylA. Thus, using text mining STRING associates lacZ and xylA.