



Computational identification of common enzymes and motifs in different metabolic pathways of *Escherichia coli* and *Arabidopsis thaliana*

Prashant Ankur Jain¹, A. K. Gupta², Krishna Misra³ and Satendra Singh^{1,*}

¹Department of Computational Biology & Bioinformatics, JSBB, SHIATS, Allahabad-211007.

²Department of Forensic Science, SHIATS, Allahabad-211007.

³Division of Bioinformatics, IITA, Allahabad-211007.

ARTICLE INFO

Article history:

Received: 19 June 2013;

Received in revised form:

15 July 2013;

Accepted: 23 July 2013;

Keywords

A. thaliana,

E. coli,

Metabolic pathway,

Motif.

ABSTRACT

Resources hub NCBI and EBI with databases like KEGG, METACYC etc. were used for computational comparative analysis of metabolic pathways in *E. coli* and *A. thaliana* to determine common enzymes and motifs according to same EC number. The extent of conservation in the metabolic pathways like glycolysis, citrate pathway, pyruvate pathway, nucleotide sugar pathway, riboflavin pathway, carboxylate pathway, galactose pathway, methane pathway, urea pathway and pentose phosphate pathway present in both *E. coli* and *A. thaliana* was analyzed. Among the ten metabolic pathways shared by both organisms, seven enzymes were identified in pentose phosphate pathway of both organisms according to their EC number. Nine motifs were present in methane pathway and nucleotide sugar pathway of *E. coli*. In *A. thaliana* maximum eight motifs in methane pathway and one in each pentose phosphate pathway, riboflavin pathway and urea pathway was identified. Comparison of *E. coli* and *A. thaliana* metabolic pathways shows that central set of pathway is largely conserved in terms of pathways, domain architect and motifs present. There was difference in the position of the motifs present in the enzymes that perform the same function so the variation in the stretch of amino acid conferring a specific structure needs to be distinguished.

© 2013 Elixir All rights reserved.

Introduction

In the last few decades, advances in molecular biology and the equipment available for research in this field have allowed the increasingly rapid sequencing of large portion of genome of several species. In fact, to date, several bacterial genomes, as well as those of simple eukaryotes (*Arabidopsis thaliana*) have been sequenced in full [1-2]. The most pressing tasks in bioinformatics involved the analysis of sequence information. Computational biology is the name given to this process. This present work studies how the enzyme sequence can be used to study metabolic pathways, the interrelationship over ranges of different organism and the comparative characteristics of specific enzyme with in the pathways. The metabolic pathway was studied in the two organisms *E. coli* and *A. thaliana*. Bioinformatics tools were used for the identification of the similarity amongst the two mentioned organism in metabolic pathways [3-4]. *E. coli* is the organism most studied the source of much of our information on molecular biology, metabolic pathways, regulation and biochemistry and continue to be a source of new insights into how cells work. Comparative analysis of metabolic pathways in *E. coli* and *A. thaliana* can give insights into the understanding of evolutionary and organizational relationships among species. This type of analysis allows one to measure the evolution of complete processes (with different functional roles) rather than the individual elements of a conventional analysis [6-8].

Materials and method

The bioinformatics tools were used to carry out of the comparative analysis of metabolic pathways in selected two

organisms. Firstly, we have selected two organisms i.e. *E. coli* and *A. thaliana*. We have selected these organisms because both having the features of model organism. We have searched different metabolic pathways (glycolysis pathway, citric acid pathway, urea pathway, riboflavin pathway, methane pathway, pentose phosphate pathway, galactose pathway, nucleotide sugar pathway, carboxylate pathway, pyruvate pathway) of *E. coli* and *A. thaliana* in KEGG database and searched about *E. coli* in Metacyc database which is a database of metabolic pathways and enzymes located at <http://MetaCyc.org/>. Its goal is to serve as a metabolic encyclopedia, containing a collection of non-redundant pathways central to small molecule metabolism, which have been reported in the experimental literature. KEGG was also employed for the study. From the KEGG web site, protein network is selected under the metabolic pathway.

The metabolic reference pathway indicates the metabolites and enzymes involved in the pathway, with their EC classification and associated metabolic pathway [9-10]. Selection of ortholog table on the reference pathway page results in a list of organism with the enzymes of the pathway. From the list, the study was centralized on two organisms by selecting them. The two organisms selected were *E. coli* and *A. thaliana* because of their evolutionary significance.

The EC number or the enzyme classification system describes the functions of an enzyme in terms of four numbers organized in a hierarchical manner and there can be several enzymes that have the same EC number but present in different pathways. We have searched different enzymes by click on individual EC number and collected their flat files to

Tele:

E-mail addresses: prash1721@gmail.com

© 2013 Elixir All rights reserved

identify their number of motifs.

Results and Discussion

The Comparative analysis of metabolic pathway in *E. coli* and *A. thaliana* proceeded as follows:-

KEGG was employed to investigate different metabolic pathways in two organisms "*E. coli* and *A. thaliana*". The KEGG reference page includes the amino acid and gene sequence for enzymes in the metabolic pathway. The reference cycle also indicates the EC number of the enzyme and the detail of the enzymes is searched by using the EC number. Different enzymes that were shared by both the organism in the metabolic pathway were selected for comparative analysis of the metabolic pathway according to EC number using computational tools.

Common enzymes according to their EC number were identified in ten metabolic pathways. Different metabolic pathways contain different number of common enzymes according to EC number. Citrate, glycolysis, pyruvate, galactose, methane, nucleotide sugar, urea, pentose phosphate, riboflavin and carboxylate pathway respectively contain five, six, six, three, four, three, five, seven, five and three enzymes accordingly in both organisms (Table 1).

In study of the motifs, we have found that in citrate pathway the maximum number of motifs in *E. coli* are six and minimum one and in *A. thaliana* maximum number of motifs are six and minimum is two.

➤ In glycolysis pathway the maximum number of motifs in *E. coli* are three and minimum one and in *A. thaliana* maximum number of motifs are two.

➤ In carboxylate pathway the maximum number of motifs in *E. coli* are five and minimum one and in *A. thaliana* maximum number of motifs are six and minimum is two.

➤ In galactose pathway the maximum number of motifs in *E. coli* are eight and minimum one and in *A. thaliana* maximum number of motifs are four and minimum is two.

➤ In methane pathway the maximum number of motifs in *E. coli* are nine and minimum one and in *A. thaliana* maximum number of motifs are eight and minimum is two.

➤ In nucleotide sugar pathway the maximum number of motifs in *E. coli* are nine and minimum one and in *A. thaliana* maximum number of motifs are six and minimum is two.

➤ In pentose phosphate pathway the maximum number of motifs in *E. coli* are seven and minimum one and in *A. thaliana* maximum number of motifs are seven and minimum is one.

➤ In pyruvate pathway the maximum number of motifs in *E. coli* are seven and minimum one and in *A. thaliana* maximum number of motifs are seven and minimum is two.

➤ In riboflavin pathway the maximum number of motifs in *E. coli* are three and minimum one and in *A. thaliana* maximum number of motifs are three and minimum is one.

➤ In urea pathway the maximum number of motifs in *E. coli* are five and minimum one and in *A. thaliana* maximum number of motifs are four and minimum is one.

So we found that the maximum number of motifs are present in methane and nucleotide sugar pathway of *E. coli* is nine and minimum number of motifs i.e. one present in all pathways.

In *A. thaliana* maximum numbers of motifs are present in methane pathway and minimum number of motifs i.e. one, present in pentose phosphate pathway, riboflavin pathway and urea pathway. Among the ten metabolic pathways shared by both organism, maximum number of enzymes i.e. seven were identified in pentose phosphate pathway of both organism

according to their EC number and maximum number of motifs present in methane and nucleotide sugar pathway of *E. coli* is nine and minimum number of motifs i.e. one present in all pathways. In *A. thaliana* maximum number of motifs present in methane pathway is eight and minimum number of motifs i.e. one, present in methane pathway, pentose phosphate pathway, riboflavin pathway and urea pathways were identified.

By studying the motif structure of enzymes present in metabolic pathways in both organisms, the difference in the stretch of amino acid conferring a specific structure can be distinguished [11-15].

This allows study of evolutionary relationship between the two organisms so as to determine the extent of conservation in the metabolic pathways like glycolysis, citrate pathway, pyruvate pathway, nucleotide sugar pathway, riboflavin pathway, carboxylate pathway, galactose pathway, methane pathway, urea pathway, pentose phosphate pathway present in both *E. coli* and *A. thaliana*.

Conclusion

Our comparison of *E. coli* and *Arabidopsis thaliana*, metabolic pathways and enzymes shows that this central set of pathway is largely conserved in terms of pathways present and in terms of the domain architect and motifs.

The main advantage of using multiple motif sets to identify protein family's lies in the fact that homology is concentrated only in conserved region between related sequences. *E. coli* and *Arabidopsis Thaliana* are closely related to each other. The difference in the position of the motifs present in the enzymes that perform the same function so that same enzyme have difference in structure. Thus, with the help of computational tools the metabolic pathway can be interpreted in terms of the protein involved.

References

1. Bjorkland, M. (1994). The independent contrast method in comparative biology. *Cladistics* 10:425-433
2. Bjorkland, M. (1997). Are "comparative methods" always necessary? *Oikos* 80:607-612
3. Beeckmans, Kanarek. L, (1987). "Enzyme -enzyme interactions as modulators of the metabolic flux through the citric acid cycle. "Biochem soc symp; 54; 163=172. PMID: 3332992
4. Brown,C.John ,(1995),. What the heck is an *Ecoli*?
5. Mansi , E.M,Nimmo HG ,Holms WH (1986). "pyruvate metabolism and the phosphorylation state of isocitrate dehydrogenase in *Escherichia coli* ."J Gen Microbiol ;132 (Pt 3);797 -806 . PMID:3525743
6. Harold ,j.Morowitz , Jennifer D. Kostelnik Jeremy Yang , and George D.Cody (2000): The origin of intermediary metabolism , *PNas* , vol . 97/no. 14 / 7704-7708
7. Holms, W.H, (1987)."Control of flux through the citric acid cycle and the glyoxylate bypass in *Escherichia coli*." *Biochem Soc Symp* ;54;17-31.PMID :3332993
8. Laporte , DC , Walsh. K,Koshland. D.E, (1984). "The branch point effect.
9. Ultrasensitivity and subsensitivity to metabolic control" *J Biol Chem*;259 (22);14068-75.PMID:6389540
10. Michael, W.king, Ph.d/IU school of medicine /university of brescia. "TCA cycle and its regulation."
11. Milburn, D,Laskowski R.A, Thornton J.M,(1998). Sequences annotated by structure: a tool to facilitate the use of structural information in sequence analysis. *Prot. Eng.*,11,855-859

Table 1. Similarities between enzymes according to EC. No. of *E.coli* & *Arabidopsis Thaliana*

Pathways	<i>E. coli</i>	<i>A. thaliana</i>
Citrate Pathway	Dihydrolipoyl transsuccinase [EC:2.3.1.61]	2-oxoacid dehydrogenase family protein [EC:2.3.1.61]
	predicted dehydrogenase [EC:1.1.1.37]	malate dehydrogenase, cytosolic, putative [EC:1.1.1.37]
	e14 prophage; isocitrate dehydrogenase, specific for NADP+ [EC:1.1.1.42]	ICDH (ICDH); isocitrate dehydrogenase (NADP+) [EC:1.1.1.42]
	2-oxoglutarate decarboxylase, thiamin-requiring [EC:1.2.4.2]	2-oxoglutarate dehydrogenase E1 component, putative / oxoglutarate decarboxylase, putative / alpha-ketoglutaric dehydrogenase, putative [EC:1.2.4.2]
	bifunctional aconitate hydratase 2/2-methylisocitrate dehydratase [EC:4.2.1.3]	aconitate hydratase, cytoplasmic, putative / citrate hydro-lyase/aconitase, putative [EC:4.2.1.3]
Glycolysis Pathway	phosphoglyceromutase 1 [EC:5.4.2.1]	2,3-biphosphoglycerate-independent phosphoglycerate mutase, putative / phosphoglyceromutase, putative [EC:5.4.2.1]
	triosephosphate isomerase [EC:5.3.1.1]	TIM (TRIOSEPHOSPHATE ISOMERASE) [EC:5.3.1.1]
	galactose-1-epimerase (mutarotase) [EC:5.1.3.3]	aldose 1-epimerase family protein [EC:5.1.3.3]
	glucosephosphate isomerase [EC:5.3.1.9]	PGII (CHLOROPLASTIC PHOSPHOGLUCOSE ISOMERASE) [EC:5.3.1.9]
	acetyl-CoA synthetase [EC:6.2.1.1]	AMP binding / acetate-CoA ligase/ catalytic [EC:6.2.1.1]
pyruvate dehydrogenase, decarboxylase component E1, thiamin-binding [EC:1.2.4.1]	pyruvate dehydrogenase E1 component beta subunit, mitochondrial / PDHE1-B (PDH2) [EC:1.2.4.1]	
Pyruvate Pathway	predicted dehydrogenase [EC:1.1.1.37]	malate dehydrogenase, cytosolic, putative [EC:1.1.1.37]
	2-isopropylmalate synthase [EC:2.3.3.13]	IPMS1/MAML-4 (METHYLTHIOALKYLMALATE SYNTHASE-LIKE 4); 2-isopropylmalate synthase [EC:2.3.3.13]
	fused malic enzyme predicted oxidoreductase/predicted phosphotransacetylase [EC:1.1.1.40]	ATNADP-ME4 (NADP-MALIC ENZYME 4); malate dehydrogenase (oxaloacetate-decarboxylating) (NADP+)/ malic enzyme/ oxidoreductase, acting on NADH or NADPH, NAD or NADP as acceptor [EC:1.1.1.40]
	acetyl-CoA synthetase [EC:6.2.1.1]	AMP binding / acetate-CoA ligase/ catalytic [EC:6.2.1.1]
	acetyl-CoA acetyltransferase [EC:2.3.1.9]	ACAT2/EMB1276 (ACETOACETYL-COA THIOLASE 2, EMBRYO DEFECTIVE 1276); acetyl-CoA C-acetyltransferase [EC:2.3.1.9]
	pyruvate dehydrogenase, decarboxylase component E1, thiamin-binding [EC:1.2.4.1]	pyruvate dehydrogenase E1 component beta subunit, mitochondrial / PDHE1-B (PDH2) [EC:1.2.4.1]
Galactose Pathway	maltodextrin glucosidase [EC:3.2.1.20]	alpha-xylosidase, putative [EC:3.2.1.20]
	beta-D-galactosidase [EC:3.2.1.23]	beta-galactosidase, putative / lactase, putative [EC:3.2.1.23]
	alpha-galactosidase, NAD(P)-binding [EC:3.2.1.22]	alpha-galactosidase, putative / melibiase, putative / alpha-D-galactoside galactohydrolase, putative [EC:3.2.1.22]
Methane Pathway	serine hydroxymethyltransferase [EC:2.1.2.1]	SHM6 (serine hydroxymethyltransferase 6); glycine hydroxymethyltransferase [EC:2.1.2.1]
	hydroperoxidase HP(III) (catalase) [EC:1.11.1.6]	CAT3 (CATALASE 3); catalase [EC:1.11.1.6]
	5,10-methylenetetrahydrofolate reductase [EC:1.5.1.20]	MTHFR1 (METHYLENETETRAHYDROFOLATE REDUCTASE 1); methylenetetrahydrofolate reductase (NADPH) [EC:1.5.1.20]

	formate dehydrogenase-N, alpha subunit, nitrate-inducible [EC:1.2.1.2]	FDH (FORMATE DEHYDROGENASE); NAD binding / cofactor binding / oxidoreductase, acting on the CH-OH group of donors, NAD or NADP as acceptor [EC:1.2.1.2]
Nucleotide Sugar Pathway	UDP-galactose-4-epimerase [EC:5.1.3.2]	UGE1 (UDP-D-glucose/UDP-D-galactose 4-epimerase 1); UDP-glucose 4-epimerase/ protein dimerization [EC:5.1.3.2]
	dTDP-glucose 4,6 dehydratase, NAD(P)-binding [EC:4.2.1.46]	UXS5 (UDP-Xyl synthase 5); catalytic [EC:4.2.1.46]
	glucose-1-phosphate uridylyltransferase [EC:2.7.7.9]	UTP--glucose-1-phosphate uridylyltransferase, putative / UDP-glucose pyrophosphorylase, putative / UGPase, putative [EC:2.7.7.9]
Urea Pathway	ornithine carbamoyltransferase [EC:2.1.3.3]	OTC (ORNITHINE CARBAMOYLTRANSFERASE); amino acid binding / carboxyl- and carbamoyltransferase [EC:2.1.3.3]
	N-acetylglutamate synthase [EC:2.3.1.1]	arginine biosynthesis protein ArgJ family [EC:2.3.1.1 2.3.1.35]
	gamma-glutamyl kinase [EC:2.7.2.11]	P5CS2 (DELTA 1-PYRROLINE-5-CARBOXYLATE SYNTHASE 2); catalytic/ glutamate 5-kinase/ glutamate-5-semialdehyde dehydrogenase [EC:2.7.2.11 1.2.1.41]
	acetylglutamate kinase [EC:2.7.2.8]	aspartate/glutamate/uridylylase kinase family protein [EC:2.7.2.8]
	acetylornithine deacetylase [EC:3.5.1.16]	peptidase M20/M25/M40 family protein [EC:3.5.1.16]
Pentose phosphate pathway	glucose-6-phosphate dehydrogenase [EC:1.1.1.49]	G6PD4 (GLUCOSE-6-PHOSPHATE DEHYDROGENASE 4); glucose-6-phosphate 1-dehydrogenase [EC:1.1.1.49]
	ribokinase [EC:2.7.1.15]	pfkb-type carbohydrate kinase family protein [EC:2.7.1.15]
	phosphoribosylpyrophosphate synthase [EC:2.7.6.1]	ribose-phosphate pyrophosphokinase 2 / phosphoribosyl diphosphate synthetase 2 (PRS2) [EC:2.7.6.1]
	D-ribulose-5-phosphate 3-epimerase [EC:5.1.3.1]	ribulose-phosphate 3-epimerase, cytosolic, putative / pentose-5-phosphate 3-epimerase, putative [EC:5.1.3.1]
	gluconate-6-phosphate dehydrogenase, decarboxylating [EC:1.1.1.44]	6-phosphogluconate dehydrogenase family protein [EC:1.1.1.44]
	gluconate kinase 2 [EC:2.7.1.12]	shikimate kinase family protein [EC:2.7.1.12]
	glucosephosphate isomerase [EC:5.3.1.9]	PGI1 (CHLOROPLASTIC PHOSPHOGLUCOSE ISOMERASE) [EC:5.3.1.9]
Riboflavin Pathway	phosphoanhydride phosphorylase [EC:3.1.3.2 3.1.3.26]	ATPAP3/PAP3 (purple acid phosphatase 3); acid phosphatase/ protein serine/threonine phosphatase [EC:3.1.3.2]
	riboflavin synthase, alpha subunit [EC:2.5.1.9]	lumazine-binding family protein [EC:2.5.1.9]
	GTP cyclohydrolase II [EC:3.5.4.25]	riboflavin biosynthesis protein, putative [EC:3.5.4.25]
	fused diamino-hydroxyphosphoribosylaminopyrimidine deaminase and 5-amino-6-(5-phosphoribosylamino) uracil reductase [EC:3.5.4.26 1.1.1.193]	cytidine/deoxycytidylate deaminase family protein [EC:3.5.4.26 1.1.1.193]
	bifunctional riboflavin kinase/FAD synthetase [EC:2.7.1.26 2.7.7.2]	ATFMN/FHY (RIBOFLAVIN KINASE/FMN HYDROLASE); FMN adenylyltransferase/ riboflavin kinase [EC:2.7.1.26]
Carboxylate Pathway	bifunctional aconitate hydratase 2/2-methylisocitrate dehydratase [EC:4.2.1.3]	aconitate hydratase, cytoplasmic, putative / citrate hydro-lyase/aconitase, putative [EC:4.2.1.3]
	predicted dehydrogenase [EC:1.1.1.37]	malate dehydrogenase, cytosolic, putative [EC:1.1.1.37]
	acetyl-CoA synthetase [EC:6.2.1.1]	AMP binding / acetate-CoA ligase/ catalytic [EC:6.2.1.1]

Table 2. Retrieval of number of motif present in enzymes according to their EC number in *E. coli* and *Arabidopsis thaliana*

<i>E. coli</i>			<i>Arabidopsis thaliana</i>	
Pathway	EC No..	No. of motifs i.d.	EC. No.	No.of motifs i.d.
Citrate pathway	4.2.1.3	1.)Pf:Aconitase_2_N 2.)pf:Aconitase 3.)pf:ACONITASE_1 4.)pf:ACONITASE_2	4.2.1.3	1.) ps:SER_RICH 2.) pf:Aconitase 3.) ps:ACONITASE_1 4.) ps:ACONITASE_2 5.) pf:Aconitase_C 6.) pf:SpoVT_AbrB
	1.2.4.2	1.)pf:E1_dh 2.) pf:Transket_pyr	1.2.4.2	1.) pf:E1_dh 2.) pf:Transket_pyr
	2.3.1.61	1.) pf:Biotin_lipoyl 2.) ps:BIOTINYL_LIPOYL 3.) pf:TolA 4.) ps:LIPOYL 5.) pf:E3_binding 6.) pf:2-oxoacid_dh	2.3.1.61	1.) pf:Biotin_lipoyl 2.)ps:BIOTINYL_LIPOYL 3.) pf:TolA 4.) ps:LIPOYL 5.) pf:E3_binding 6.) pf:2-oxoacid_dh
	1.1.1.37	1.) pf:Ldh_2	1.1.1.37	1.) pf:Ldh_1_N 2.) pf:Epimerase 3.) ps:MDH 4.) pf:Ldh_1_C
	1.1.1.42	1.) pf:Iso_dh 2.) ps:IDH_IMDH	1.1.1.42	1.) pf:Iso_dh 2.) ps:IDH_IMDH
	2.3.3.1	1.) pf:Citrate_synt 2.) ps:CITRATE_SYNTHASE	2.3.3.1	1.) pf:Citrate_synt 2.)ps:CITRATE_SYNTHASE
	4.2.1.2	1.) pf:Lyase_1 2.) ps:FUMARATE_LYASES	4.2.1.2	1.) pf:Lyase_1 2.)ps:FUMARATE_LYASES
Glycolysis pathway	1.2.4.1	1.) pf:Transketolase_N	1.2.4.1	1.) pf:Transket_pyr 2.)pf:Transketolase_C
	5.4.2.1	1.) pf:PGAM 2.) ps:PG_MUTASE	5.4.2.1	1.)pf:iPGM_N 2.) pf:Metalloenzyme
	5.3.1.1	1.) pf:TIM 2.) pf:tRNA_synt_1c_R1 3.) ps:TIM	5.3.1.1	1.) pf:TIM 2.) pf:TIM
	5.1.3.3	1.) pf:Aldose_epim 2.) ps:ALDOSE 1 EPIMERASE	5.1.3.3	1.)pf:Aldose_epim 2.) pf:Yqai
	6.2.1.1	1.) ps:EF_HAND_1 2.) pf:AMP-binding 3.) ps:AMP_BINDING	6.2.1.1	1.) pf:AMP-binding 2.)ps:AMP_BINDING
Carboxilate pathway	4.2.1.3	1.) pf:Aconitase_2_N 2.) pf:Aconitase 3.) ps:ACONITASE_1 4.) ps:ACONITASE_2	4.2.1.3	1.) ps:SER_RICH 2.) pf:Aconitase 3.) ps:ACONITASE_1 4.) ps:ACONITASE_2 5.) pf:Aconitase_C 6.) pf:SpoVT_AbrB
	4.2.1.2	1.) pf:Lyase_1 2.) ps:FUMARATE_LYASES	4.2.1.2	1.) pf:Lyase_1 2.)ps:FUMARATE_LYASES
	1.1.1.37	1.) pf:Ldh_2	1.1.1.37	1.) pf:Ldh_1_N 2.) pf:Epimerase 3.) ps:MDH 4.) pf:Ldh_1_C
	6.2.1.1	1.) ps:EF_HAND_1 2.) pf:AMP-binding 3.) ps:AMP_BINDING	6.2.1.1	1.) pf:Ldh_1_N 2.) pf:Epimerase 3.) ps:MDH 4.) pf:Ldh_1_C
	1.1.1.42	1.) pf:Iso_dh 2.) ps:IDH_IMDH	1.1.1.42	1.) pf:Iso_dh 2.) ps:IDH_IMDH
	4.1.1.31	1.) pf:PEPcase 2.) pf:MIF4G_like 3.) pf:SCAN 4.) ps:PEPCASE_1 5.) ps:PEPCASE_2	4.1.1.31	1.) pf:PEPcase 2.) pf:DUF885 3.) ps:PEPCASE_1 4.) ps:PEPCASE_2
Galactose pathway	3.2.1.23	1.) pf:Glyco_hydro_2_N 2.)pf:Bac_rhamnosid_N 3.) pf:Glyco_hydro_2 4.) pf:ROF 5.) pf:Glyco_hydro_2_C 6.)ps:GLYCOSYL_HYDROL_F2_1 7.)ps:GLYCOSYL_HYDROL_F2_2 8.) pf:Bgal_small_N	3.2.1.23	1.)pf:Glyco_hydro_35 2.)pf:Glyco_hydro_42 3.)ps:GLYCOSYL_HYDROL_F35 4.)pf:Glyco_hydro_2_N
	3.2.1.20	1.) pf:Alpha-amylase	3.2.1.20	1.)pf:Glyco_hydro_31

				2.)ps:GLYCOSYL_HYDROL_F31_1
	3.2.1.22	1.)pf:Glyco_hydro_4 2.)ps:GLYCOSYL_HYDROL_F4	3.2.1.22	1.)pf:Melibiase 2.)ps:ALPHA_GALACTOSIDASE
Methane pathway	1.2.1.2	1.) ps:TAT 2.) pf:Molybdop_Fe4S4 3.)ps:MOLYBDOPTERIN_PROK_1 4.) pf:Molybdopterin 5.) pf:Molydop_binding 6.)ps:MOLYBDOPTERIN_PROK_3	1.2.1.2	1.) pf:2-Hacid_dh 2.) pf:2-Hacid_dh_C 3.) pf:Shikimate_DH 4.)pf:NAD_binding_2 5.) pf:DapB_N 6.)ps:D_2_HYDROXYACID_DH_1 7.)ps:D_2_HYDROXYACID_DH_2 8.)ps:D_2_HYDROXYACID_DH_3
	2.1.2.1	1.) pf:SHMT 2.) pf:Aminotran_1_2 3.) pf:Aminotran_5 4.) pf:Cys_Met_Meta_PP 5.) pf:OKR_DC_1 6.) pf:OKR_DC_1 7.) pf:DegT_DnrJ_EryC1 8.) pf:Beta_elim_lyase 9.) ps:SHMT	2.1.2.1	1.) pf:SHMT 2.)pf:Aminotran_1_2 3.)pf:Beta_elim_lyase 4.) ps:SHMT 5.)ps:WD_REPEATS_1
	1.11.1.6	1.)pf:Catalase 2.) ps:CATALASE_2 3.) ps:CATALASE_1 4.) pf:Catalase-rel	1.11.1.6	1.) pf:Catalase 2.)ps:CATALASE_1
	1.5.1.20	1.) pf:MTHFR	1.5.1.20	1.) pf:MTHFR
	1.1.1.1 1.1.1.284	1.) pf:ADH_N 2.) ps:ADH_ZINC 3.) pf:ADH_zinc_N 4.) pf:ApbA 5.) pf:TrkA_N	1.1.1.1 1.1.1.284	1.) pf:ADH_N 2.) ps:ADH_ZINC 3.)pf:ADH_zinc_N 4.) pf:ThiF
Nucleotide sugar pathway	5.1.3.2	1.) pf:adh_short 2.) pf:RmlD_sub_bind 3.) pf:Epimerase 4.) pf:KR 5.) pf:NmrA 6.) pf:Polysacc_synt_2 7.) pf:3Beta_HSD 8.) pf:NAD_binding_4	5.1.3.2	1.) pf:adh_short 2.)pf:RmlD_sub_bind 3.) pf:Epimerase 4.)pf:Polysacc_synt_2 5.) pf:3Beta_HSD 6.)pf:NAD_binding_4
	2.7.7.9	1.) pf:NTP_transferase	2.7.7.9	1.)pf:tRNA_lig_trans 2.) pf:UDPGP
	1.1.1.22	1.) pf:Ldh_1_N 2.) pf:NAD_Gly3P_dh_N 3.) pf:UDPG_MGDP_dh_N 4.) pf:3HCDH_N 5.) pf:F420_oxidored 6.) pf:Pyr_redox 7.) pf:ApbA 8.) pf:UDPG_MGDP_dh 9.) pf:UDPG_MGDP_dh_C	1.1.1.22	1.)pf:UDPG_MGDP_dh_N 2.)pf:UDPG_MGDP_dh 3.)pf:UDPG_MGDP_dh_C
	4.2.1.46	1.) pf:adh_short 2.) pf:RmlD_sub_bind 3.) pf:Epimerase 4.) pf:Polysacc_synt_2 5.) pf:3Beta_HSD 6.) pf:NAD_binding_4	4.2.1.46	1.)pf:UDPG_MGDP_dh_N 2.)pf:UDPG_MGDP_dh 3.)pf:UDPG_MGDP_dh_C
Pentose phosphate pathway	2.7.6.1	1.) ps:PRPP_SYNTHETASE 2.) pf:Pribosyltran 3.)ps:PUR_PYR_PR_TRANSFER	2.7.6.1	1.) ps:SER_RICH 2.)ps:PRPP_SYNTHETASE 3.) pf:Pribosyltran 4.)ps:PUR_PYR_PR_TRANSFER
	1.1.1.44	1.) pf:NAD_binding_2 2.) pf:NAD_Gly3P_dh_N 3.) pf:F420_oxidored 4.) pf:ApbA 5.) pf:PDH 6.) pf:6PGD 7.) ps:6PGD	1.1.1.44	1.)pf:NAD_binding_2 2.)pf:Hexokinase_2 3.) pf:6PGD
	5.1.3.1	1.) pf:TMP-TENI 2.) pf:Ribul_P_3_epim 3.) pf:OMPdecase 4.) ps:RIBUL_P_3_EPIMER_1 5.) ps:RIBUL_P_3_EPIMER_2	5.1.3.1	1.)pf:Ribul_P_3_epim 2.)ps:RIBUL_P_3_EPIMER_1 3.) pf:Aldolase 4.)ps:RIBUL_P_3_EPIMER_2
	2.7.1.12	1.) pf:APS_kinase 2.) pf:Thymidylate_kin	2.7.1.12	1.)pf:APS_kinase 2.) pf:ABC_tran

		3.) pf:SKI		3.)pf:Thymidylate_kin 4.) pf:SKI
	2.7.1.15	1.)pf:APS_kinase 2.) pf:Phos_pyr_kin 3.) pf:HK 4.) ps:RHODANESE_3 5.) ps:PFKB_KINASES_2	2.7.1.15	1.) pf:APS_kinase 2.) pf:ABC_tran 3.)pf:Thymidylate_kin 4.) pf:SKI
	5.3.1.9	1.) pf:PGI 2.)ps:P_GLUCOSE_ISOMERASE_1 3.) pf:MORN 4.)ps:P_GLUCOSE_ISOMERASE_2	5.3.1.9	1.) pf:PGI 2.)ps:P_GLUCOSE_ISOMERASE_2
	3.1.1.31	1.) pf:SGL	3.1.1.31	1.)pf:Glucosamine_iso
	2.2.1.1	1.) pf:Transketolase_N 2.) pf:E1_dh 3.) ps:TRANSKETOLASE_1 4.) pf:TPP_enzyme_C 5.) pf:Transket_pyr 6.) ps:TRANSKETOLASE_2 7.) pf:Transketolase_C	2.2.1.1	1.)pf:Transketolase_N 2.) pf:E1_dh 3.) ps:TRANSKETOLASE_1 4.)pf:TPP_enzyme_C 5.)pf:Transket_pyr 6.)ps:TRANSKETOLASE_2 7.) pf:Transketolase_C
Pyruvate pathway	6.2.1.1	1.)ps:EF_HAND_1 2.) pf:AMP-binding 3.) ps:AMP_BINDING	6.2.1.1	1.) pf:AMP-binding 2.)ps:AMP_BINDING
	2.3.3.13	1.) ps:PYPYR_CT 2.)ps:AIPM_HOMOCIT_SYNTH_1 3.) pf:HMGL-like 4.) pf:Trp_syntA 5.)ps:AIPM_HOMOCIT_SYNTH_2 6.) pf:LeuA_dimer	2.3.3.13	1.)ps:PYPYR_CT 2.)ps:AIPM_HOMOCIT_SYNTH_1 3.)pf:HMGL-like 4.)pf:Glyco_hydro_2 5.) pf:LeuA_dimer
	1.1.1.37	1.) pf:Ldh_2	1.1.1.37	1.)pf:Ldh_1_N 2.) pf:Epimerase 3.) ps:MDH 4.) pf:Ldh_1_C
	2.3.1.9	1.) pf:Thiolase_N 2.) ps:THIOLASE_1 3.) pf:Thiolase_C 4.) pf:Ketoacyl-synt_C 5.) pf:ACP_syn_III_C 6.) ps:THIOLASE_2 7.) ps:THIOLASE_3	2.3.1.9	1.)pf:Thiolase_N 2.)ps:THIOLASE_1 3.) pf:Thiolase_C 4.) pf:Ketoacyl-synt_C 5.)pf:ACP_syn_III_C 6.)ps:THIOLASE_2 7.)ps:THIOLASE_3
	1.2.4.1	1.) pf:Transketolase_N	1.2.4.1	1.)pf:Transket_pyr 2.)pf:Transketolase_C
Riboflavin pathway	2.7.1.26 2.7.7.2	1.) pf:FAD_syn 2.) pf:Flavokinase	2.7.1.26 2.7.7.2	1.) pf:Hydrolase 2.) pf:Flavokinase
	3.5.4.26	1.)pf:dCMP_cyt_deam_1 2.)ps:CYT_DCMP_DEAMINASES 3.) pf:RibD_C	3.5.4.26	1.)pf:dCMP_cyt_deam_1 2.) pf:RibD_C 3.) pf:DUF1768
	3.1.3.2	1.) pf:Acid_phosphat_A 2.)ps:HIS_ACID_PHOSPHAT_1 3.)ps:HIS_ACID_PHOSPHAT_2	3.1.3.2	1.)pf:Metallophos
	3.5.4.25	1.) pf:GTP_cyclohydro2	3.5.4.25	1.)pf:dCMP_cyt_deam_1 2.) pf:RibD_C 3.) pf:DUF1768
Urea payhway	2.7.2.11	1.)pf:AA_kinase 2.) pf:AvrPphF-ORF-2 3.) ps:GLUTAMATE_5_KINASE 4.) ps:PUA 5.) pf:PUA	2.7.2.11	1.)pf:AA_kinase 2.)pf:BtpA 3.)ps:GLUTAMATE_5_KINASE 4.) ps:PROA
	2.3.1.1	1.) pf:UPF0228 2.)pf:AA_kinase 3.) ps:GNAT 4.) pf:Acetyltransf_1	2.3.1.1	1.) pf:ArgJ
	3.5.1.16	1.) pf:Peptidase_M28 2.) ps:ARGE_DAPE_CPG2_1 3.) pf:Peptidase_M20 4.) ps:ARGE_DAPE_CPG2_2 5.) pf:M20_dimer	3.5.1.16	1.)pf:Peptidase_M20 2.) pf:M20_dimer
	2.7.2.8	1.) pf:AA_kinase	2.7.2.8	1.) pf:AA_kinase

12. Neidhardt, F.C, Curtiss III, R, Ingraham J,L, Lin ECC, Low Jr KB, Magasanik B, Reznikoff WS , Riley M , Schaechter M, Umberger HE (1996) "Escherichia coli and cellular and molecular biology, second edition. "American society of microbiology , Washington , D.C.
13. Kenneth, Todar© 2005 Kenneth Todar University of Wisconsin – Madison , Department of Bacteriology –"Text book of bacteriology."
14. Rastogi, P. and Mendiratta N (2004). Bioinformatics methods and applications, Genomics.
15. Agrawal, Abhishek and Dunham J. Valgene. Journal of the South Carolina Academy of Science 2(1):1-17. Bioinformatics: An undergraduate research/teaching tool, department of biology , Coastal Carolina University , Conway .sc29528