2017

Hub Proteins, Paralogs, and Unknown Proteins in Bacterial Interaction Networks

Neha Sakhawalkar
Virginia Commonwealth University

Hub Proteins, Paralogs, and Unknown Proteins in Bacterial Interaction Networks

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Integrative Life Sciences at Virginia Commonwealth University.

by
Neha Sakhawalkar
Master of Bioinformatics, Virginia Commonwealth University, 2012
Master of Science (Biotechnology), University of Pune, 2010
Bachelor of Science (Biotechnology), University of Pune, 2008

Advisor: Dr. Peter Uetz
Associate Professor
Center for Study of Biological Complexity

Virginia Commonwealth University
Richmond, Virginia
April, 2017

This dissertation is dedicated to my grandparents

Vasant and Vandana Sakhardande

who have motivated me always to push my limits

ACKNOWLEDGEMENT

I am grateful to my advisor, my committee members, friends, and my family. It is their guidance and support that has given me the strength and confidence to finish my dissertation.

I would like to express my sincerest gratitude to my advisor, Dr. Peter Uetz, for providing me with this wonderful opportunity and for providing me with an excellent atmosphere for doing research. This dissertation would never have been accomplished without his knowledge, wisdom and insight. I am very thankful to him for his guidance and support throughout my doctoral journey. I would also like to thank my committee member for their valuable inputs and advice. They have encouraged me to expand my knowledge and guided me to stay focused.

I am grateful to all my lab members for their help and guidance with my research. I would especially like to thank Dr. Jitender Mehla and Dr. Harry Caufield for their invaluable suggestions, advice and help with experimental work.

I would like to thank my mother Mrs. Mridula Sakhawalkar for her support, blessings and encouragement. She has been a source of inspiration for me throughout my life and especially during my doctoral studies. I am truly grateful for all the sacrifices she has made for me. I would also like to thank my grandparents for their blessings and unconditional love. They have always

motivated me to pursue a doctoral degree. I am very thankful to my brother Nitish Sakhawalkar for being there to encourage me through the ups and downs of this journey.

Finally, I would like to thank by husband Parag Kulkarni for standing by me through this long journey and being patient. He has been always there for me and pushed me to do my best.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

PPI	Protein – protein Interaction
Eco	*Escherichia coli K12 W3110*
Vch	*Vibrio cholera serotype O1 El tor*
Ype	*Yersinia pestis KIM10+*
Sau	*Staphylococcus aureus COL*
Rpr	*Rickettsia prowazekii strain Madrid E*
MS	Mass Spectrometry
Y2H	Yeast two hybrid assay/system
B2H	Bacterial two hybrid assay/system
3-AT	3-amino 1,2,4-triazole
COG	Cluster Orthologous Group
OG	Orthologous Group

ABSTRACT

HUB PROTEINS, PARALOGS, AND UNKNOWN PROTEINS IN BACTERIAL INTERACTION NETWORKS

By
Neha Sakhawalkar, MS

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2017

Major Director: Dr. Peter Uetz
Associate Professor
Center for Study of Biological Complexity

Proteins are the functional units of cells. However, a major portion of the proteome does not have a known functional annotation. This dissertation explores protein -protein interactions, involving these uncharacterized or unknown function proteins. Initially, protein – protein interactions were tested and analyzed for paralogous proteins in *Escherichia coli.* To expand this concept further and to get an overview, protein – protein interactions were analyzed using 'comparative interactomics' for four pathogenic bacterial species including *Escherichia coli*, *Yersinia pestis*, *Vibrio cholerae* and *Staphylococcus aureus*. This approach was used to study unknown function protein pairs as well as to focus on uncharacterized hub proteins. The

dissertation aims at using protein – protein interactions along with other research data about proteins as a possible approach to narrow down on functions of proteins.

CHAPTER I

INTRODUCTION

Proteins. Proteins are one of the most important macromolecules in a cell and contribute to more than 50 percent of the dry weight of cells. Proteins perform diverse functions in a cell including replication of DNA, cytoskeleton formation, facilitating transport of molecules into and out of the cell, etc. Several proteins, DNA, and RNA molecules function together in an orderly manner to perform these functions. By carrying out these diverse functions, proteins maintain the structural and functional integrity of a cell (David Whitford, 2005).

Protein Structure. Proteins are made up of amino acids in a linear polypeptide chain. These amino acids form the secondary structure of the proteins such as the α-helix or β-strands. The α-helix or β-strands maximize formation of hydrogen bonds leading to stabilization of the proteins as well as minimize steric repulsion. A β-strand generally does not exist alone; they pair up with another β-strand to form a β-sheet. In the β-sheets, the β-strands may be parallel, antiparallel or mixed. Some proteins, generally fibrous proteins have the entire backbone made up of either one of these secondary structures, e.g. keratin is made up of α-helices only whereas fibroin is made up of β-strands alone. Globular proteins on the other hand are made up of a combination of α-helices and β-strands (Gary Walsh, 2002). The interactions between the side chains in a protein comprise of the tertiary structure of the protein. Quaternary structure deals with peptides that consist of more than one polypeptide chain. The quaternary structure may consist of two similar

or dissimilar polypeptides (dimers) or can be more complex and consist of several similar or dissimilar or a combination of polypeptides, e.g. human hemoglobin has two α subunits and two β subunits (David Whitford, 2005).

Interactions. Cellular functions are performed by a cascade of signaling pathways. These signaling pathways are a result of interactions between molecules of a cell. These interactions could be DNA – protein interactions, RNA – protein interactions, small molecule – protein interactions or protein – protein interactions.

DNA – protein Interactions. DNA – protein binding is required for carrying out various housekeeping function of a cell. Transcriptional regulation, replication and DNA repair, chromosome maintenance are some of the function that require interaction between DNA and protein molecules. To study these cellular functions at the molecular level to get further insight into their mechanism various techniques have been developed. These techniques can be categorized into *in vitro* techniques, *in vivo* techniques and *in silico* techniques. The commonly used *in vitro* techniques include foot printing assay, Electrophoretic mobility shift assay (EMSA), Southwestern blotting, and Yeast one hybrid assay. Chromatin immunoprecipitation (ChIP), ChIP-ChIP, ChIP sequencing and numerous variations of ChIP are the *in vivo* techniques used. Most computational methods available are aimed at predicting transcription factor based gene regulation like TRANSFAC, Multi-genome analysis of positions and patterns of elements of regulation (MAPPER), ProNIT, DNAProt, etc (Dey et al., 2012).

RNA – protein Interactions. Study of RNA – protein interactions aid in getting a better understanding about mRNA processing like alternative splicing, editing, methylation, etc. Studying these interactions led to identification of coding RNAs. The techniques used to study the interactions are categorized into protein centric and RNA centric. Protein centric methods

rely on protein purification, followed by its interaction of the purified protein with the RNA which is then sequenced to map RNA binding proteins. On the other hand, RNA centric methods rely on mass spectrometry for identifying proteins bound to a particular RNA or a class of RNAs. (McHugh *et al.*, 2014).

Small molecule – protein Interactions. Small molecule – protein interactions find applications in drug discovery and in basic biology for studying complex pathways in cells (Wang, 2005). The methods used to study small molecule – protein interactions can be categorized into small molecule to protein and protein to small molecule strategies. Most of these techniques are based on a combination of affinity chromatography and mass spectrometry (McFedries, 2013).

Protein – protein Interactions. PPIs find applications in basic as well as applied sciences. Numerous methods have been developed to study them and these techniques can be categorized into *in vitro*, *in vivo* and *in silico*. Each of the methods used has advantages and disadvantages, however based on the research question a combination of these methods can be used for more accurate analysis.

In vitro methods. The *in vitro* methods used to study PPIs are mainly chromatography based followed by analysis using mass spectrometry. Tandem Affinity Purification – Mass Spectrometry (TAP-MS) is a powerful method to determine composition of a protein complex. The protein of interest is fused to a tag and is expressed in cells and used for purification. When the protein is purified, other interacting proteins also get purified with it. Chromatography is used to isolate this protein complex from other proteins, cellular debris, etc. This complex is then analyzed for its constituent proteins using mass spectrometry (Kaiser *et al.*, 2008).

In vivo methods. The most widely used *in vivo* method is yeast two hybrid assay. This assay will be used for purposes of this dissertation.

Principle of the Two Hybrid Assay. The yeast two hybrid is the first developed two hybrid assay. Yeast two hybrid assay is based on the principle of the split protein assay where Gal4, a modular protein is split into two parts. The two parts of the protein can be non-covalently reconstituted to perform the function of a transcriptional activator. The split protein can be reconstituted by bait and prey (interacting proteins) attached to the split parts. A reporter gene is used for detection of the reconstitution of the split protein. Gal4 is made up of two domains; A DNA Binding Domain (DBD) is in involved in binding to specific DNA sequences and the Activation Domain (AD) functions in activating the process of transcription.

When proteins X and Y are expressed and interact with each other, Gal4 is reconstituted and expresses the reporter gene. If the proteins X and Y do not interact the reporter gene is not expressed. The HIS3 gene which is required for histidine biosynthesis is used as a reporter gene in yeast two hybrid assays. (Brückner, et al, 2009). The important requirements for a successful yeast two hybrid screen are expression vectors for bait and prey fusion proteins, mating compatible yeast strains, and an effective reporter gene (Mehla, Caufield, & Uetz, 2015). Array based high throughput screening has been used recently to create huge amount of interaction data within relatively short period of time. Array based techniques are used to study interactomes (interactions among an entire proteome of an organism).

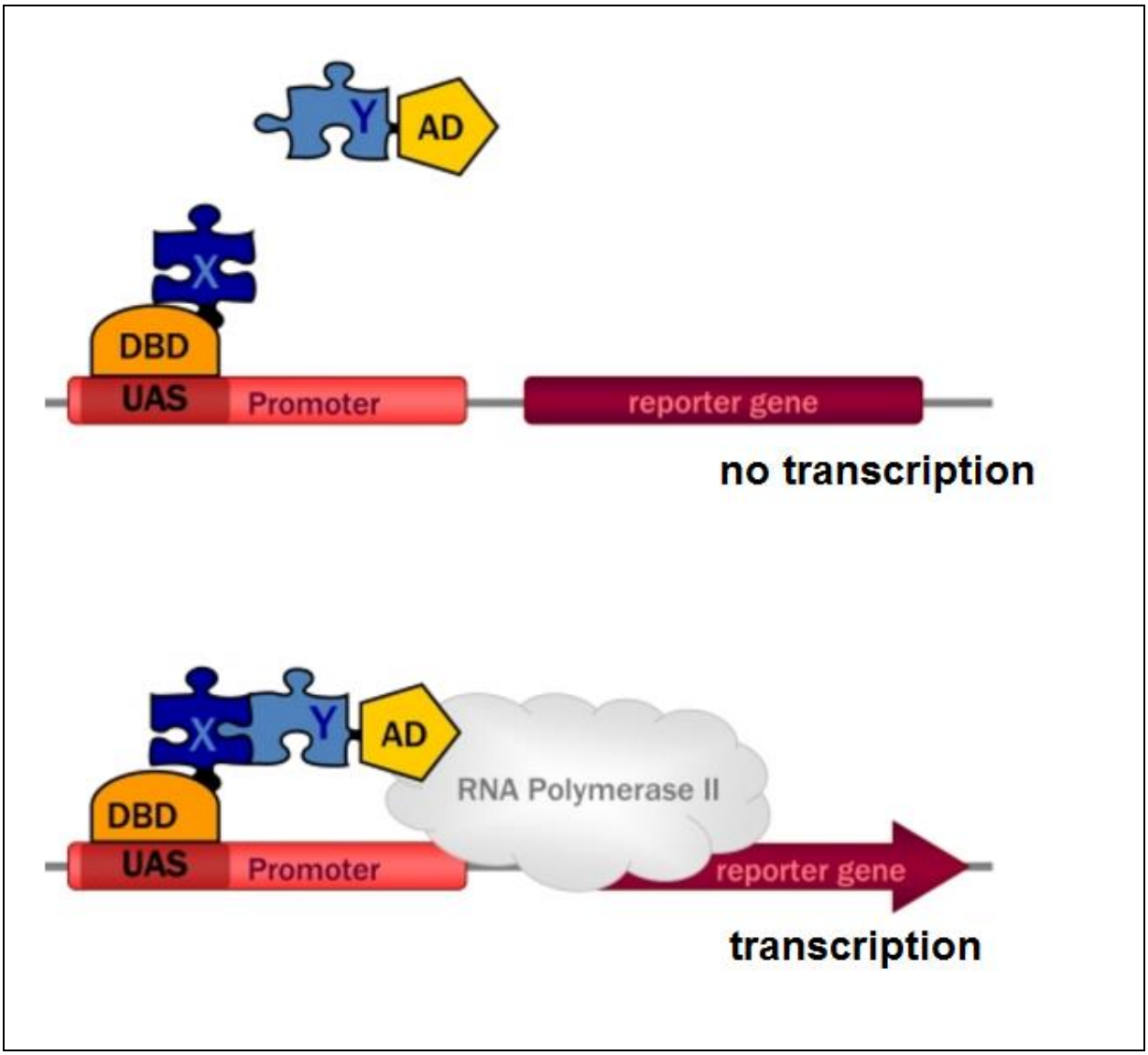


Fig. 1.1 – Principle of Yeast two hybrid assay (Bruckner, 2009)

Advantages of the assay. Yeast two hybrid assay has found widespread utility and popularity due to the following features. i) Two hybrid assay is *in vivo* and hence the proteins being screened for interactions should be folded correctly and may undergo post translational modifications as they would naturally do in a cell, only if these are yeast proteins. ii) The interacting proteins can be from the same or from different organisms. For example, phage – host interactions (Giles phage

and host *Mycobacterium smegmatis*) (Mehla, *et al*., 2015). iii) The proteins being screened do not have a size limit. Protein domains or fragments can also be used for testing interactions. iv) The assay is highly sensitive which is likely due to multiple factors including overproduction of hybrid proteins. v) The assay is relatively inexpensive and can be done in a high throughput manner.

Limitations of the Assay. Several features of the two-hybrid assay impose limitations on the type of the interactions that can be analyzed. i) One of the major drawbacks of the assay is its potential high rate of false positives. The number of false positives can be reduced by including stringent assay conditions and by filtering the interaction dataset. ii) The post translation modifications in prokaryotes are different from the ones in eukaryotes. Hence, interactions dependent on posttranslational modification may not be identified.

In silico methods. Protein modelling can be used to study protein-protein interactions. Various online as well as desktop based tools are available for protein docking, eg. ZDOCK, GRAMM-X, etc.

Interactomics. All protein-protein interactions across the proteome of an organism are the ***interactome*** of that organism. Entire interactome studies have been carried for a handful of bacterial species, and model organisms like baker's yeast (*Saccharomyces cerevisiae*) (Uetz et al., 2000), fruit fly (*Drosophila melanogaster*) (Yu, 2008), nematode (*Caenorhabditis elegans*) (S. Li et al., 2004), and humans (Rolland et al., 2014). The yeast interactome (Uetz et al., 2000) was the first to be studied extensively, followed by bacteria, including *Escherichia coli* (Rajagopala, Sikorski, Kumar, Mosca, Vlasblom, Arnold, Franca-koh, et al., 2014), *Helicobacter pylori* (Häuser et al., 2014), *Mycobacterium tuberculosis* (Yi Wang et al., 2010), *Mycoplasma pneumoniae* (Kühner et al., 2009), among several others (Häuser et al., 2014). These interactomes were studied using different permutations of the split protein assay (yeast two hybrid, bacterial two hybrid assays, etc) or using affinity purification and mass spectrometry (AP/MS). Most

studied interactomes include less than 80% of the proteome since some proteins are low in abundance in a cell or may not be expressed in standard laboratory growth conditions.

Interactomics finds a huge number of applications in basic as well as applied sciences. Studying protein-protein interactions in important cellular systems can help improve our understanding of basic cellular biology as well as disease. Proteomics can also be used to identify **biomarkers** for early and more efficient diagnosis of diseases. Using an interactomics approach in conjunction with biomarkers has the major advantage of being able to identify all possible genes and proteins that can act as biomarkers. A recent study used this approach to develop an algorithm to identify biomarkers for breast cancer. In this case, the researchers integrated gene expression data with PPI to build a classifier to which certain parameters were applied to obtain biomarkers for the disease (Chen, 2011). Since protein-protein interactions are required for all cellular functions, they can also be used as **drug targets** especially in cases where the interaction is essential for survival of the organism. One of the forthcoming areas of study for determining drug targets using interactomics is host – pathogen PPIs e.g. in human and bacterial pathogens, pathogenic bacteria and bacteriophage. This approach also makes it possible to target PPIs which are not required for normal cellular function of the human host and thus help reduce side effects caused by drugs (Feng, et al, 2015)(Skwarczynska & Ottmann, 2015). One other application of interactomics is towards determining **function of uncharacterized proteins** (Schwikowski, 2000). In most genomes 20-30% of all genes have no known function, and in many bacteria this number can exceed 50%. In fact, the genomes of most bacteriophages isolated from the environment encode on the order of 80% uncharacterized proteins. Even the human proteome encodes more than a 1000 uncharacterized proteins. Studying protein-protein interactions can help get a lead in determining the function based on the function of other proteins that the protein of interest interacts with. This approach is the focus of this dissertation.

CHAPTER II

CONSERVATION OF PPIs IN PARALOGS IN ESCHERICHIA COLI

Introduction

Gene duplication events are considered to be important events for evolution and adaptation of an organism. The general idea of gene duplication is to have more copies of a gene for phenotypic stability. If the parent gene accumulates mutations in subsequent generations, then the duplicated gene can perform its function. However the duplicated genes generally evolve to perform different functions and are a major source of gaining new functions by the organism (Espinosa-Cantú A, 2015). Functions of a gene may be retained after duplication or the function may be lost and become pseudogenes which will eventually be lost by the organism. By default duplicated genes (paralogs) have the even though that may change over time. This overlapping function of paralogs enables the organism to withstand selection pressure and maintain stability while undergoing evolution (Gavin Conant, 2008).

Paralogs are observed in simple organisms like microbes as well as very complex organisms i.e. mammals. Gene duplication in microorganisms can either be a duplication of the gene or duplication of a gene segment consisting multiple genes, such as duplication of an operon. A gene duplication study across 106 genomes was carried out. Genes belonging to 51 families were conserved across all species and performed housekeeping functions (Gevers, 2004). The

conserved genes included the ones involved in translation, transcription, metabolism, etc. (Marit Bratlie 2010).

Paralogs share sequence similarity with each other and hence we would expect that they would share interacting partners as well. The main aim of this study is to test the above hypothesis. To test this hypothesis, interactions will be tested in parallel for paralogs in *Escherichia coli* based on interactions observed in the *Escherichia coli* interactome Rajagopala *et al*., 2014 as well as PPI predictions based on previously published interactions (Fig. 2.1).

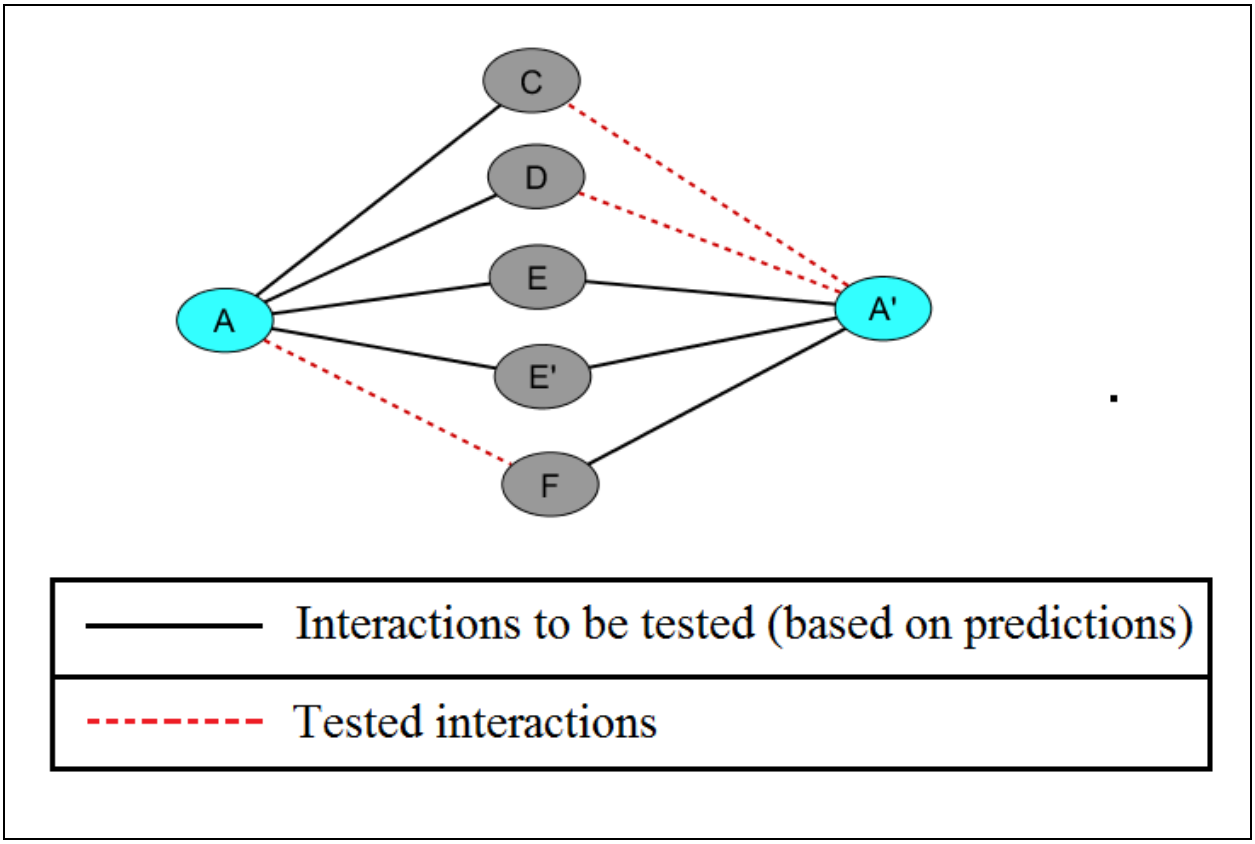


Fig. 2.1 – Approach for study of paralogs in *Escherichia coli* denoting tested and predicted interactions

Methods

Creating dataset. The *Escherichia coli K12* proteome was obtained from UniProt. This data was mapped to OG data from EggNOG database version 3.0 (Powell et al., 2012) such that each unique UniProt ID was mapped to an OG. One COG may be assigned to multiple proteins. If a COG had multiple proteins, then these proteins are considered paralogous to each other.

Binary PPI interaction data for *Escherichia coli* was obtained (Rajagopala *et al.*, 2014). This dataset included interactions studied using yeast two hybrid technique as well as literature curated interactions. Each interacting protein in this set was mapped to an OG. Some proteins couldn't be assigned to an OG but were included in the analysis as long as the interacting protein could be mapped. In addition to the tested interactions, a dataset of ~ 40,000 predicted interactions (predictions were based on the presence of these interactions in other microbial species, obtained from unpublished data from Harry Caufield) was also used in this study.

Construction of bait and prey clones. See Appendix A.

Yeast transformation. See Appendix A2.

Bait auto activation test. See Appendix A3.

Yeast two hybrid assay. See Appendix A4.

Results

Dataset Analysis. The *Escherichia coli* proteome consists of ~4200 proteins (UniProt). Among these proteins, 70% of the proteins were categorized into COGs and ~840 proteins were singletons (proteins that couldn't be categorized into COGs by EggNOG 3.0) (Fig. 2.2 A). COGs were used to determine paralogous proteins from the proteome. Based on EggNOG 3.0 (Powell et al., 2012) *Escherichia coli* proteins have been categorized into 2,135 unique COGs. Among

these unique COGs, ~591 COGs coding for ~1,809 proteins had more than one protein categorized into the same COG i.e. they had paralogs (Fig. 2.2 A and 2.2 C). Further analysis of the *Eco* proteome indicated that ~30% of the proteins with paralogs have known protein structures (Fig. 2.2 B). Also, a very high percentage of the proteins have a known protein function (UniProt annotations). It can be seen in Fig. 2.2 D that as the number of paralogs increases the number of proteins with unknown functions decreases.

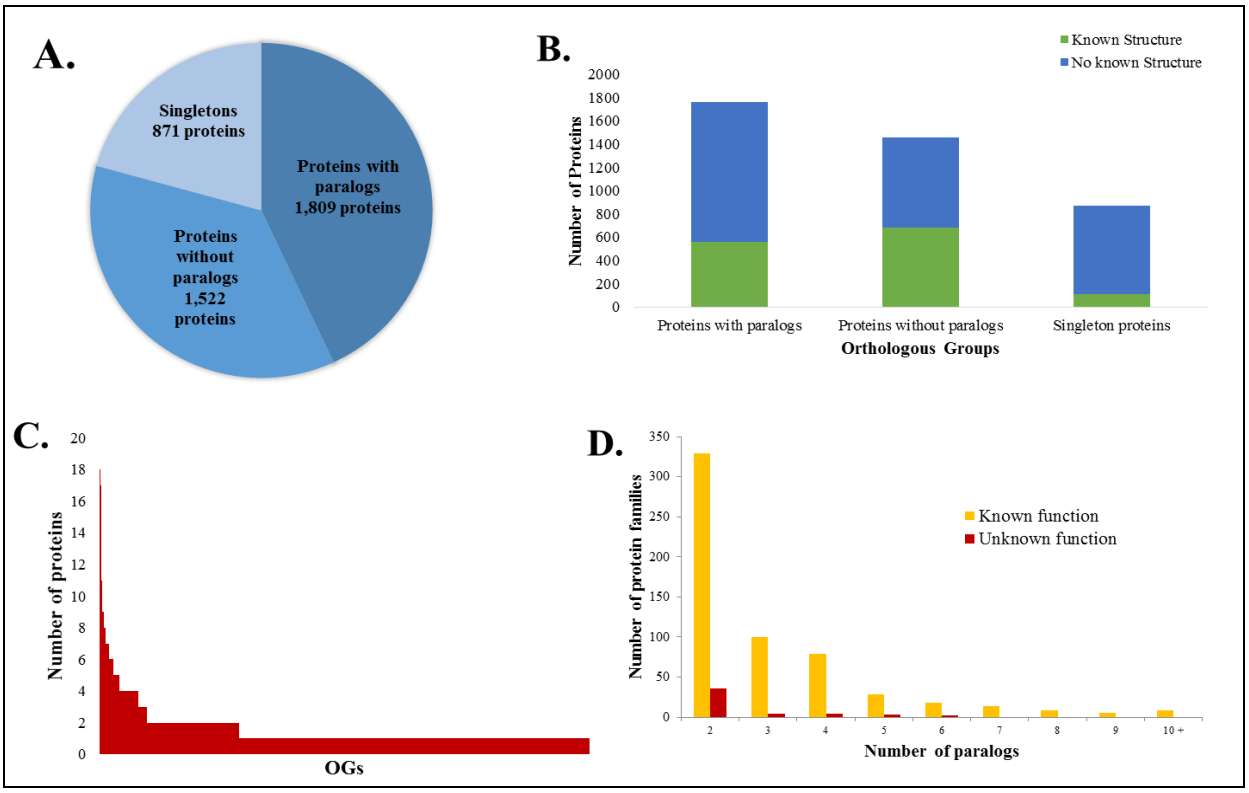


Fig. 2.2 – Paralogs in *Escherichia coli* (UniProt) A. *Escherichia coli* K12 proteome is made up of 30% singletons (not assigned any OGs), 50% OGs without paralogs and the remaining 20% OGs with paralogs. B. This graph shows that the overall PDB structures are available for less than 50% *Escherichia coli* proteins. Most paralogous proteins do not have a known structure. C. This graph depicts the number of OGs that may or may not have paralogs. Majority of *Escherichia coli* proteins do not have paralogs. D. This graph includes OGs with paralogs only.

The COGs in *Eco* were categorized into NCBI functional categories. This categorization revealed that the COGs with paralogs were distributed across all cellular functions and did not show clustering for any specific function like housekeeping genes.

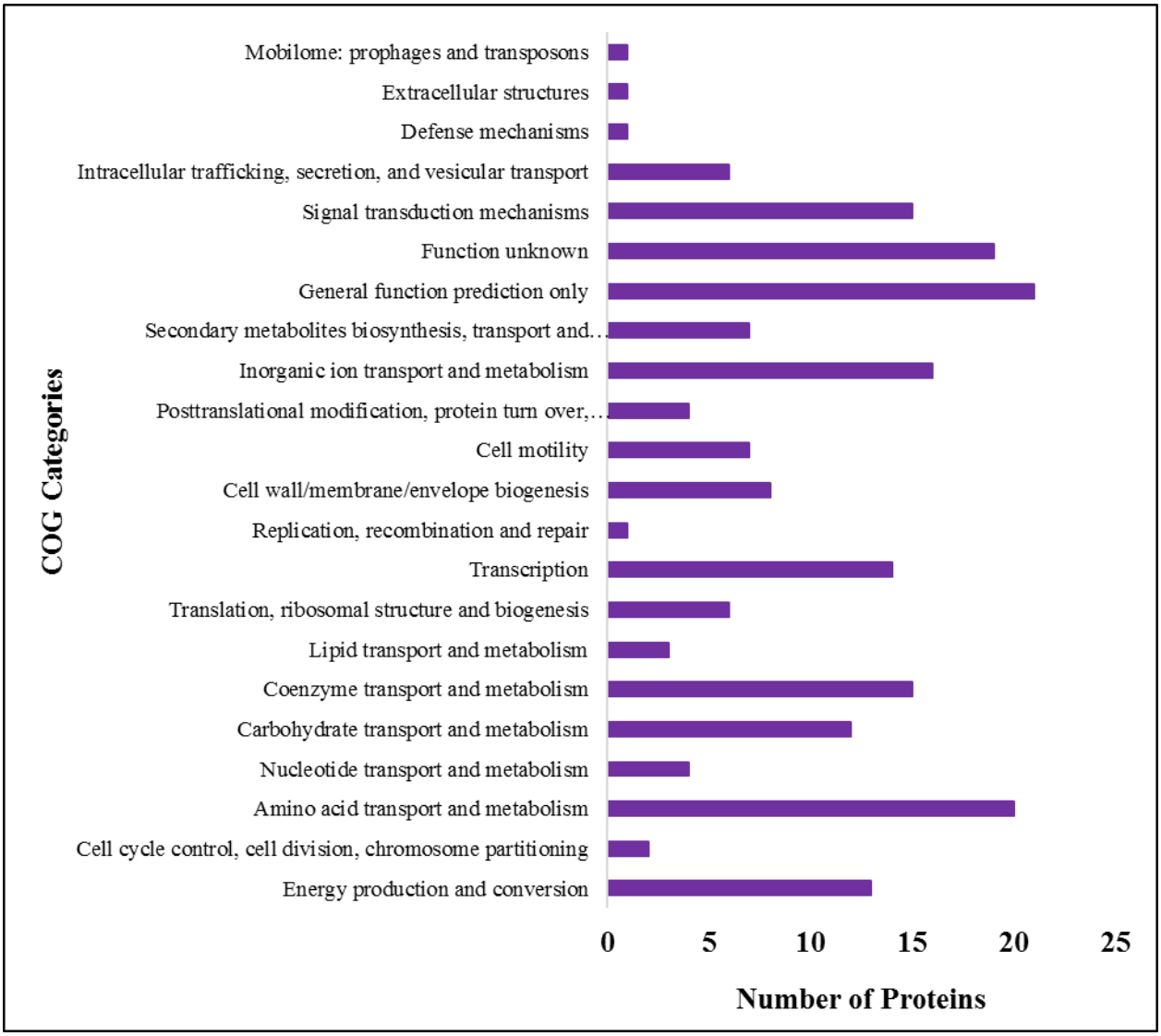


Fig. 2.3 – COGs organized into NCBI functional categories. Distribution of all COGs for paralogous proteins in *Escherichia coli*

Experimental Analysis. Two sets of paralogous proteins and their interactions were tested using yeast two hybrid system. The bait proteins (paralogous proteins) were tested for interactions at various concentrations of 3-AT. The paralogous proteins to be tested were determined based on two criteria of at least one of the paralogs was annotated as uncharacterized protein and at least one of the paralogs had a known protein structure in *Escherichia coli*. These two paralog pairs

were selected for experimental analysis since they shared more interactions (previously tested/predicted) with each other.

Paralogous proteins: AllD, DlgD and YbiC

Proteins AllD, DlgD and YbiC are paralogous proteins and belong to COG2055. COG2055 includes proteins involved in energy production and conversion specifically malate/lactate/ureidoglycolate dehydrogenases. Protein sequence alignment does not reveal high sequence similarity (Fig. 2.4). The protein sequence similarity might be low because the proteins have evolved to perform different functions. Protein structures have been studied in using X-ray diffraction in *Escherichia coli* and are available for all three paralogs. Comparison of the three structures revealed a high level of structural similarity. Protein AllD is annotated as ureidoglycolate dehydrogenase. AllD is involved in the anaerobic utilization of allantoin. Allantoin is a metabolic intermediate in plants, animals and bacteria and is produced as a degradation product of purine nucleobases from uric acid. Under anaerobic conditions, allantoin activates promoters at the 5' end of AllD. The activated gene leads to expression of ureidoglycolate dehydrogenase (Cusa *et al.*, 1999). Protein DlgD is 2,3-diketo-L-gulonate reductase and catalyzes the reduction of 2,3-diketo-L-gulonate in the presence of NADH, to form 3-keto-L-gulonate (Yew, *et al.*, 2002). YbiC has been annotated as uncharacterized oxidoreductase. All the proteins belong to Pfam PF02615, Ldh 2 protein family which has proteins that code for malate/L-lactate dehydrogenase.

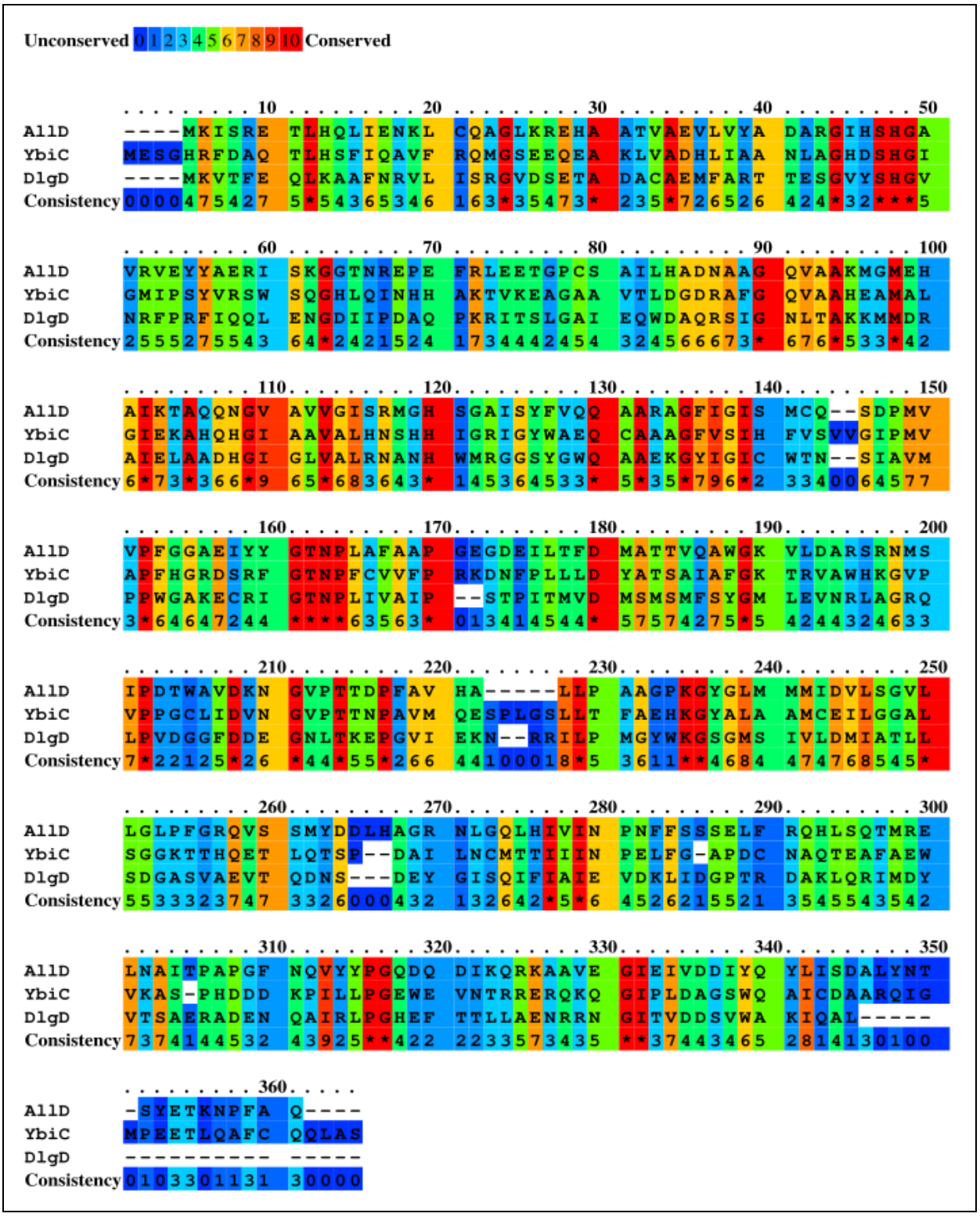


Fig. 2.4 – Protein sequence alignment for paralogous proteins AllD, DlgD and YbiC

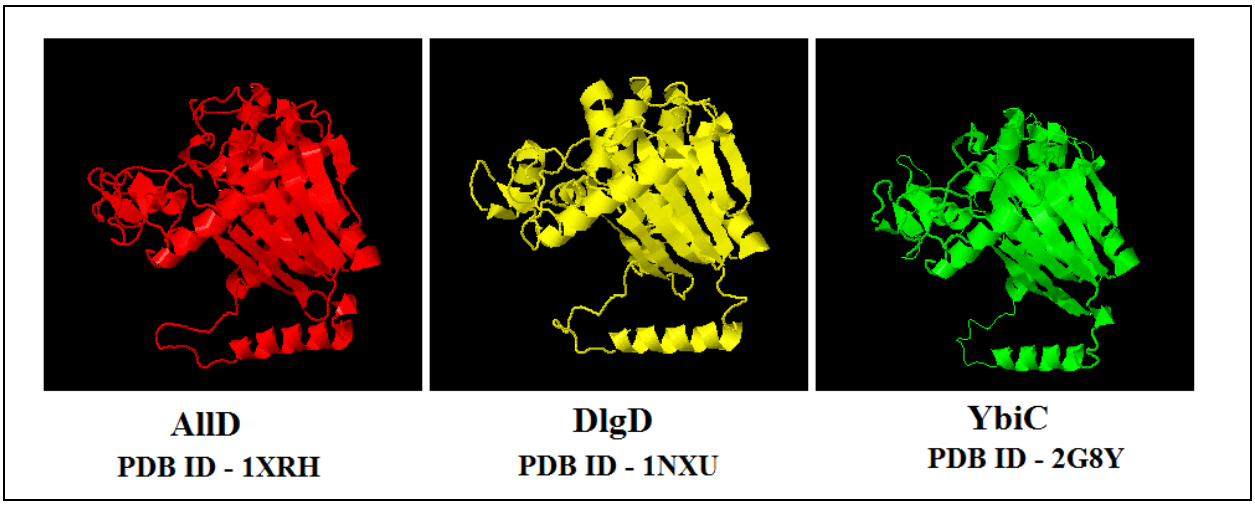
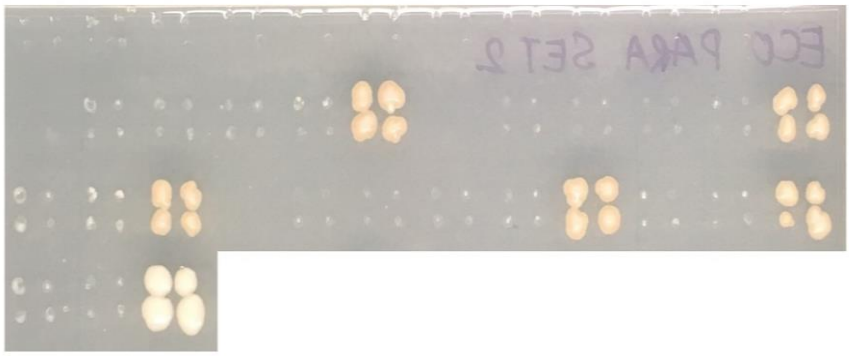


Fig. 2.5 – Structures of paralogous proteins AllD, DlgD and YbiC

The interactions (denoted with red dotted line) in the figure (Fig. 2.7 A) were tested using different protein – protein interaction methods, including yeast two hybrid assay, bacterial two hybrid assay, mass spectrometry, etc. in previous interactome studies. The predicted interactions (denoted with solid black lines) were obtained from unpublished Harry Caufield dataset based on interactions detected positive in organisms other than *Escherichia coli*. The yeast two hybrid assay carried out to test this set of already tested and predicted interactions detected very few interactions. The interactions detected positive in *Escherichia coli* by other methods were not detected in the Y2H screens. However, various interactions for YbiC were detected positive in this screen. However, no interactions were detected for proteins AllD and DlgD. To get a better coverage for the proteins, different vector combinations can be used. Also, since the previously tested and predicted interactions have not necessarily been tested using the current method, other methods can be used to get a full picture of the interactions.



BAIT	1	2	3	4	5	6	7	8	9	10	11	12
A	allD	dlgD	ybiC	allD	dlgD	ybiC	allD	dlgD	ybiC	allD	dlgD	ybiC
B	allD	dlgD	ybiC	allD	dlgD	ybiC	allD	dlgD	ybiC	allD	dlgD	ybiC
C	allD	dlgD	ybiC									

PREY	1	2	3	4	5	6	7	8	9	10	11	12
A	allD	allD	allD	bga2	bga2	bga2	bgaL	bgaL	bgaL	bglR	bglR	bglR
B	ch60	ch60	ch60	dlgD	dlgD	dlgD	nadE	nadE	nadE	uvrA	uvrA	uvrA
C	ybiC	ybiC	ybiC									

Fig. 2.6 – Yeast two hybrid screen for testing interactions for paralogous proteins AllD, DlgD and YbiC at 10mM 3-AT

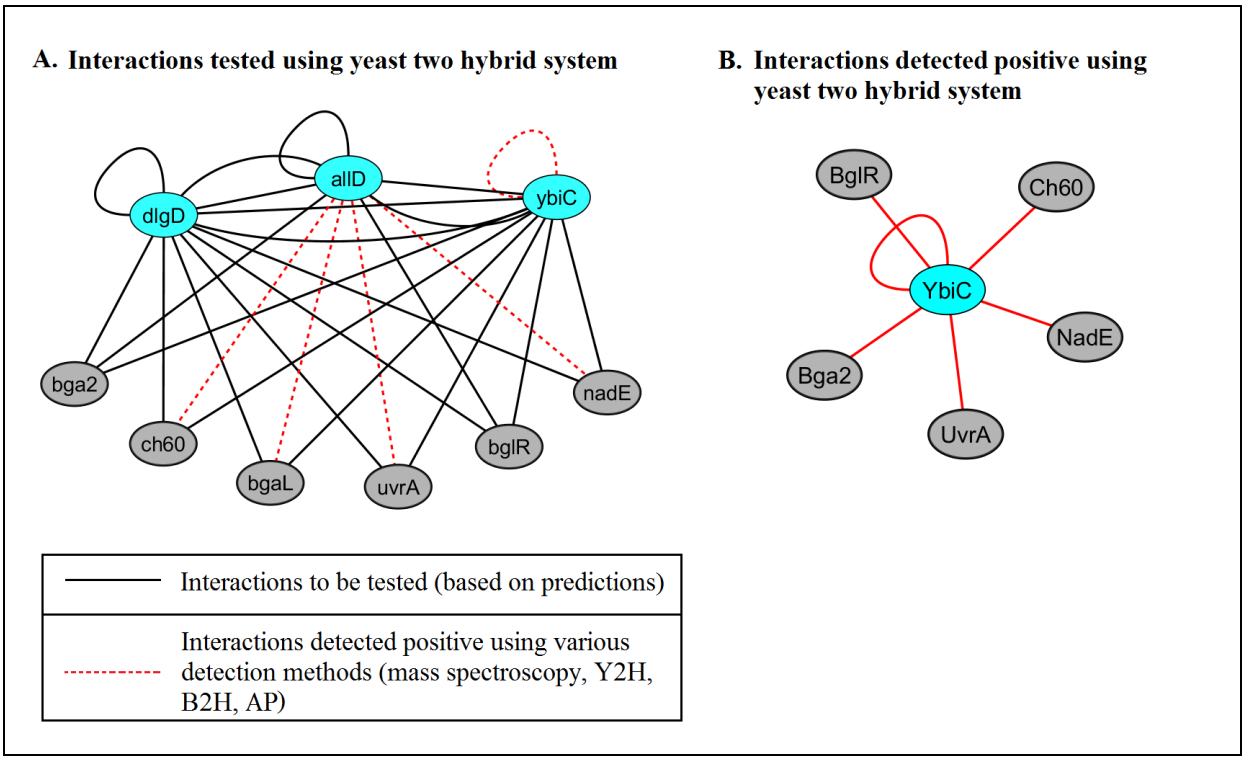


Fig. 2.7 – Protein – protein interaction network for paralogous proteins AllD, DlgD and YbiC

Paralogous proteins: SmrA and YfcN

Proteins SmrA and YfcN are paralogous proteins and are categorized under COG2840 which includes proteins involved in replication, recombination, and repair specifically DNA nicking endonucleases. Protein SmrA is annotated as probable DNA endonuclease SmrA based on sequence homology. No experimental or bioinformatics analysis has been carried out to determine the function of this protein. However, a partial structure of SmrA is available (PDB ID - 3QD7) (Gui et al., 2011). On the other hand, protein YfcN has been annotated as uncharacterized protein. There is no annotation based or experimental data available for this protein. The protein alignment shows low conservation of amino acids, however, since these proteins are paralogous both of them have a similar domain structure and belong to the same Pfam (PF01713).

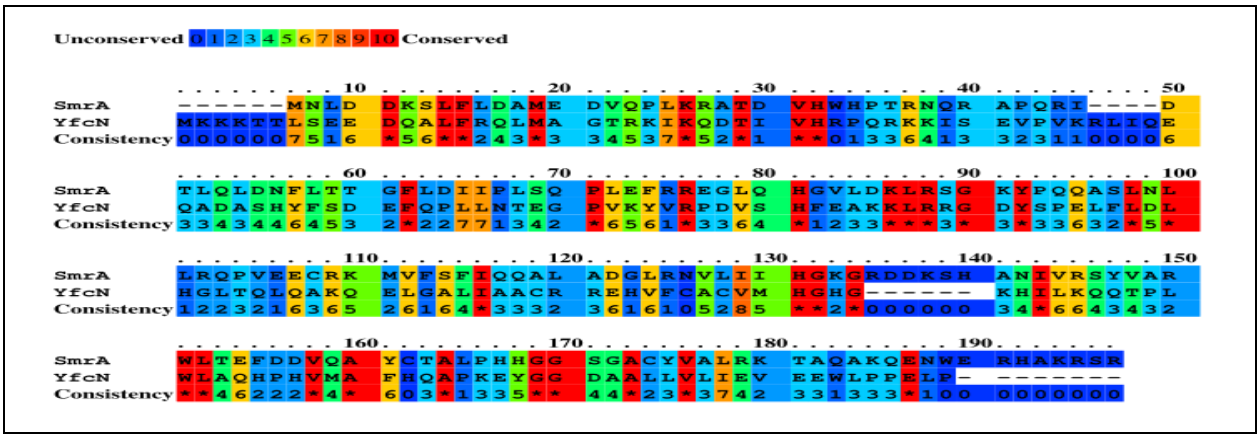
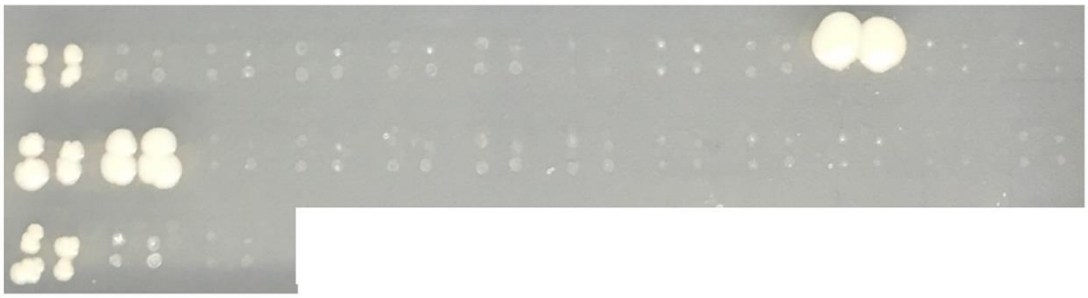


Fig. 2.8 – Protein sequence alignment for paralogous proteins SmrA and YfcN

The Y2H screen showed a combination of previously tested and predicted interactions. Both SmrA and YfcN interacted with protein MinC. Protein MinC is a septum site determining protein. It is a cell division inhibitor that block the formation of polar Z ring septum (G. Li & Young, 2012). SmrA and YfcN interact with other proteins as well, however, they share only one interaction. In this case as well, the screen is not very conclusive. Since the interaction set that was used for the screen included interactions tested by different methods (mass spectrometry, B2H, TAP-MS, etc.), only one screen cannot be used to get a full coverage. To get a full coverage the interactions need to be tested using different PPI detection methods.



BAIT	1	2	3	4	5	6	7	8	9	10	11	12
A	alsA	alsA	araG	araG	glcC	glcC	kefG	kefG	lsrA	lsrA	mglA	mglA
B	minC	minC	rbsA	rbsA	sygA	sygA	uspC	xylG	xylG	yijO	yijO	ytfR
C	ytfR											

PREY	1	2	3	4	5	6	7	8	9	10	11	12
A	smrA	yfcN	smrA	yfcN	smrA	yfcN	smrA	yfcN	smrA	yfcN	smrA	yfcN
B	smrA	yfcN	smrA	yfcN	smrA	yfcN	smrA	smrA	yfcN	smrA	yfcN	smrA
C	yfcN											

Fig. 2.9 – Yeast two hybrid screen for testing interactions for paralogous proteins SmrA and YfcN at 10mM 3-AT

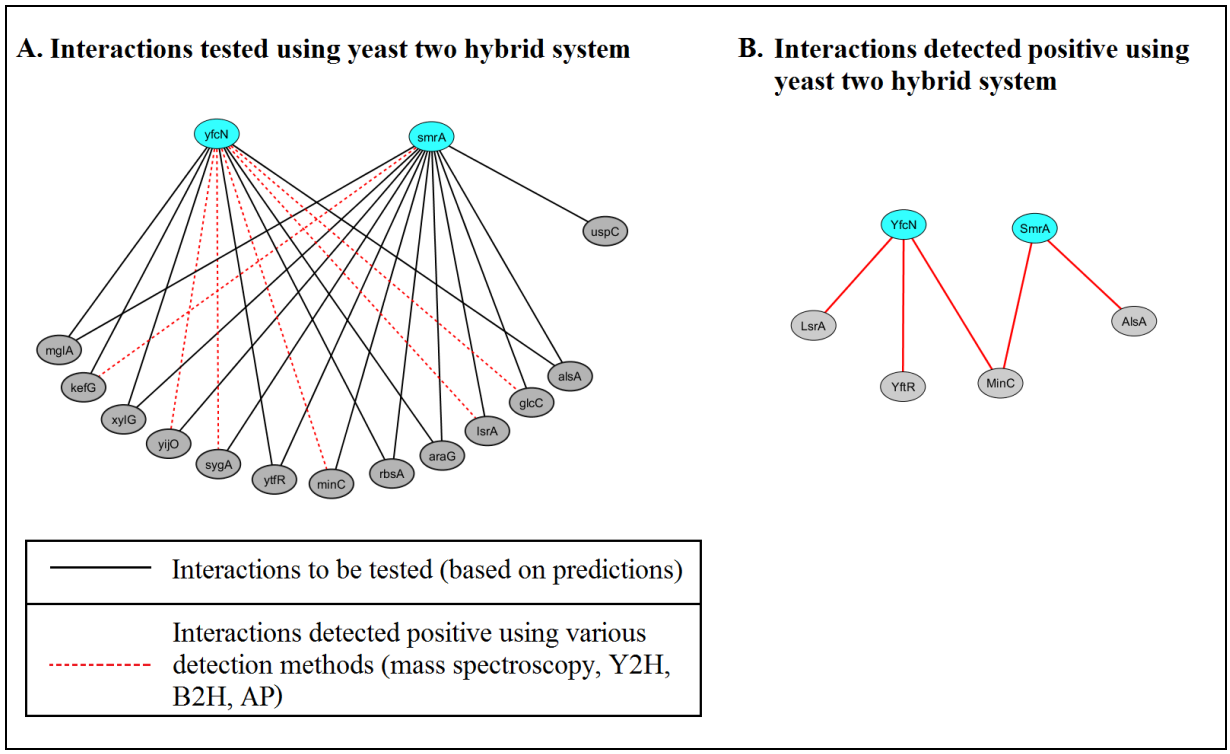


Fig. 2.10 – Protein – protein interaction network for paralogous proteins SmrA and YfcN

Discussion

~ 43% of the *Eco* proteome comprises of paralogs. A huge percentage of these paralogs do not have functional annotations. Paralogous proteins where at least one of the paralog was an uncharacterized protein and at least one protein had a known structure were included in this study while making sure they shared at least 4 shared interactions based on experimental and predicted data. Experimental PPI data was specifically for *Eco* interactions and was obtained from various literature sources that employed different experimental techniques. Predicted PPI data also involved interactions studied using different techniques however, it included all PPI from all organisms for which interaction data was available including *Eco*.

Interactions for paralogous proteins AllD, DldG and YbiC were tested using Y2H screens. All of these proteins have a well-studied structure available and the structure comparison revealed high

level of similarity though the sequence alignment showed an overall low conservation of amino acids. The Y2H screen showed interactions for YbiC, no interactions were observed for AllD and DldG in the screens. One of the reasons for this could be that the interactions with these proteins were weak or transient and could not be detected using Y2H screens, other methods could be used to test these interactions. Also in this study only one vector combination was used, using all four vector combinations (pGADT7g – pGBGT7g, pGADT7g – PGBKCg, pGADCg – PGBGT7g, pGADCg – PGBKCg) might assist in getting better coverage. In this case, paralogs show a high structural similarity and would be expected to share interaction partners.

Paralogs SmrA and YfcN were also tested in this chapter using Y2H assay. The Y2H screen detected one interaction shared by both the paralogs, however, remaining interactions were detected only in either of the proteins. In case the sequence similarity was low and since YfcN structure is not available, structure comparison could not be done. Similar approach as can be followed for this paralog pair as well. However, based on current data no predictions can be made about whether these proteins would share interacting partners or whether their function are very distinct.

CHAPTER III

COMPARATIVE INTERACTOMICS FOR CONSERVED PROTEINS

Introduction

Interactomes have been studied for handful of microbes as well as eukaryotic model organisms like *Saccharomyces cerevisiae* (H. Yu et al., 2008), *Drosophila melanogaster* (Guruharsha K G *et al*., 2011), *Caenorhabditis elegans* (S. Li et al., 2004) , and humans (Rolland et al., 2014). The yeast interactome was the first to be studied extensively and comprehensively. It was followed by *Escherichia coli* (Rajagopala, Sikorski, Kumar, Mosca, Vlasblom, Arnold, Franca-Koh, et al., 2014) (Arifuzzaman et al., 2006)(Hu et al., 2009), *Helicobacter pylori* (Häuser et al., 2014), *Mycobacterium tuberculosis* (Y Wang et al., 2010), *Treponema pallidum* (Titz et al., 2008) among several others. These interactome datasets can be used as resource for studying new signaling pathways, determining new drug targets, studying pathogenicity related genes, annotating proteins, etc. Most of these interactome studies cover ~70-80% of the proteome and focus on a particular metabolic pathway or a cellular pathway to validate the interactions based on their biological roles. The interactomes have been studied using various techniques including yeast two hybrid assay, bacterial two hybrid assay, affinity purification and mass spectrometry (AP-MS), pull down assay, etc. Among these high confidence interactions, ~40-45% interactions involve uncharacterized or unknown function proteins.

The interactome of *Treponema pallidum*, a syphilis spirochete was studied using yeast two hybrid assay. *Treponema pallidum* encodes for 1,039 proteins. ~70% of its proteome was cloned and 3,649 protein-protein interactions were detected. To derive biological insights from the interactions, a subnetwork was created for proteins involved in DNA metabolism. As a result of this protein subnetwork in combination with literature, ~20 proteins were more precisely annotated (Titz et al., 2008). The *Helicobacter pylori* interactome was studied using yeast two hybrid assay. This interactome study covered ~70% of the proteome and revealed a core network with 908 high confidence interactions. *Mycobacterium tuberculosis* interactome was studied using a high throughput bacterial two hybrid assay. The interactome was studied comprehensively to yield ~60% high confidence interactions among 3,989 ORFs (Y Wang et al., 2010). Most of the interactomes show very low PPI overlap with each other. One example includes a comparison between the interactomes of *Treponema pallidum* with *Escherichia coli*, *Helicobacter pylori* and *Campylobacter jejuni*. The comparison showed as low as 5 interactions conserved across *Treponema pallidum* and *Helicobacter pylori*, 23 interactions conserved between *Treponema pallidum* and *Escherichia coli*, and 26 interactions conserved across *Treponema pallidum* and *Campylobacter jejuni* (Titz et al., 2008). The overlap between interactions was lower than 5%. The low overlap could be due to different methodologies used to study the interactomes, larger phylogenetic distance between the species compared, as well as incomplete proteome coverage in case of these interactomes.

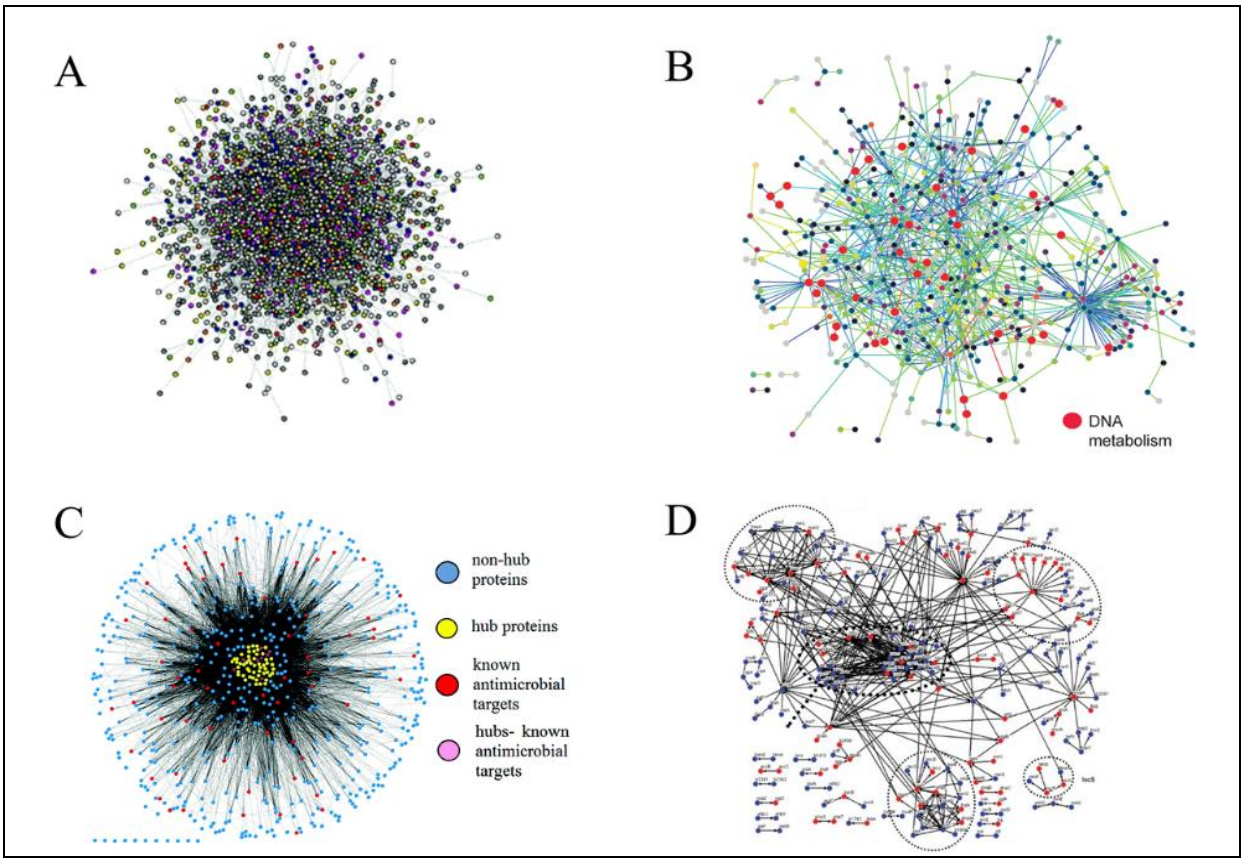


Fig. 3.1 – Interactomes of bacterial species studied using different techniques. A. *Mycobacterium tuberculosis* interactome studied using B2H(Mehla, 2015) B. *Treponema pallidum* interactome studied using Y2H (Titz, 2008) C. *Staphylococcus aureus* MRSA using Co-Immunoprecipitation (Cherkasov, 2011) D. Escherichia coli interactome studied using TAP-MS (Arifuzzaman, 2006)

The yeast interactome was studied using yeast two hybrid assay. The interactions to be tested were obtained from the Yeast Proteome Database (YPD) which included a set of 2,709 published interactions. The yeast two hybrid assay detected 2,358 (~90%) of these interactions among 1548 proteins. The interacting proteins had a diverse range of functions and were located in different locations in a cell (Uetz, 2000). Subsequently, the yeast interactome was studied using the same technique (yeast two hybrid assay) but with a different approach. In this case, interactions were

tested across the entire ORFeome (~95%) of yeast. This study yielded a total of 4,549 high quality interactions among 3,278 proteins. Though the two interactomes were studied using the same technique, there was only ~17% overlap between the two (Ito et al., 2001). A representative subset of the yeast proteome was also studied using immune affinity purification - mass spectrometry. This subset included proteins performing various cellular functions. This study detected 3,617 interactions across 1,578 proteins covering ~25% of the yeast proteome (Krogan et al., 2006). The interactions from this study showed a very low overlap with the interactions observed in the two yeast two hybrid screens. *Escherichia coli* interactomes was also studied using yeast two hybrid technique (Rajagopala, *et al*., 2014)as well as mass spectrometry (Arifuzzaman et al., 2006) (Hu et al., 2009). The results are similar to that observed in yeast. There is a very low overlap (~5%) between interactions across the mass spectrometry and one yeast two hybrid screens.

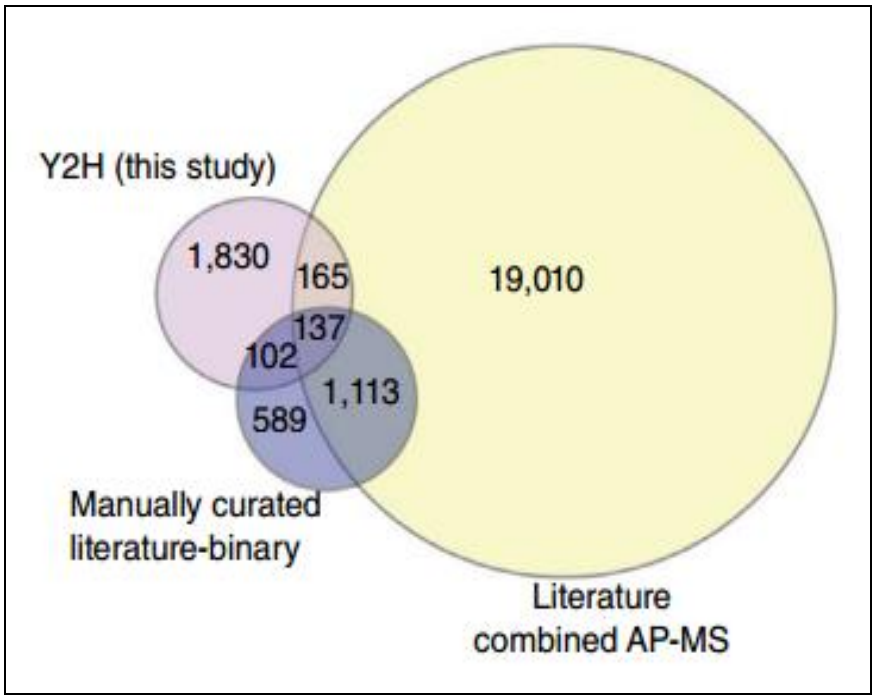


Fig 3.2 – Overlap between PPIs in *Escherichia coli* using different detection methods

Conserved proteins will be studied in this chapter. The interactions will be studied in parallel to enable direct comparison between species. Our aim is to create a dataset of binary PPIs where both of the interacting proteins will be conserved. Orthologs for these protein pairs will be determined in other species and all interactions will be tested experimentally by yeast two hybrid assay. The interactions conserved across distantly as well as closely related species will help predict function for conserved proteins without a known function due to 'guilt by association' assuming that more than one interaction with an unknown protein will be conserved. This approach will also yield a set of high confidence interactions (reducing the number of false positives).

Selection of species. Protein clone sets are publicly available for only ~ 15 bacterial species. Among these, five pathogenic species were chosen based on phylogenetic distance. The reason for low overlap across interactomes of the previously studied bacterial species was attributed to larger phylogenetic distance. To determine whether phylogenetic distance has an impact, two closely related (*Yersinia pestis, Vibrio cholera*) and two distantly related (*Staphylococcus aureus, Rickettsia prowazekii*) organisms were selected. *Eschecrichia coli* is the only organism in this dataset that has previous interaction data using Y2H. Hence the *Eco* interactome was used as a reference dataset and all analysis were based on that dataset. *Escherichia coli* is a gram-negative gamma proteobacterium that colonizes the gastrointestinal tract of eukaryotic organisms. Most of the strains of *Eco* are harmless but some can be pathogenic causing food poisoning (Kaper, Nataro, & Mobley, 2004). *Eco* is a model organism and hence is well studied to be used as a reference organism. *Yersinia pestis* is also a gram negative gamma proteobacterium which causes bubonic and pneumonic plague causing deaths due to recurring pandemics (Deng et al., 2002). *Vibrio cholerae* is closely related to *Eco* and is gram negative

gamma proteobacterium as well. It can potentially cause an epidemic and life-threatening secretory diarrhea (Finkelstein R A, 1996). *Staphylococcus aureus* on the other hand is distantly related to *Eco* and is a gram positive bacterium which is generally harmless and is found on skin and respiratory tracts of humans. However, some serotypes can cause respiratory infections. *Rickettsia prowazekii* is more distantly related to *Eco* and also has a reduced genome (834 protein coding genes) (Andersson et al., 1998). It is a gram positive bacteria and causes louse born typhus in humans which can reach epidemic levels.

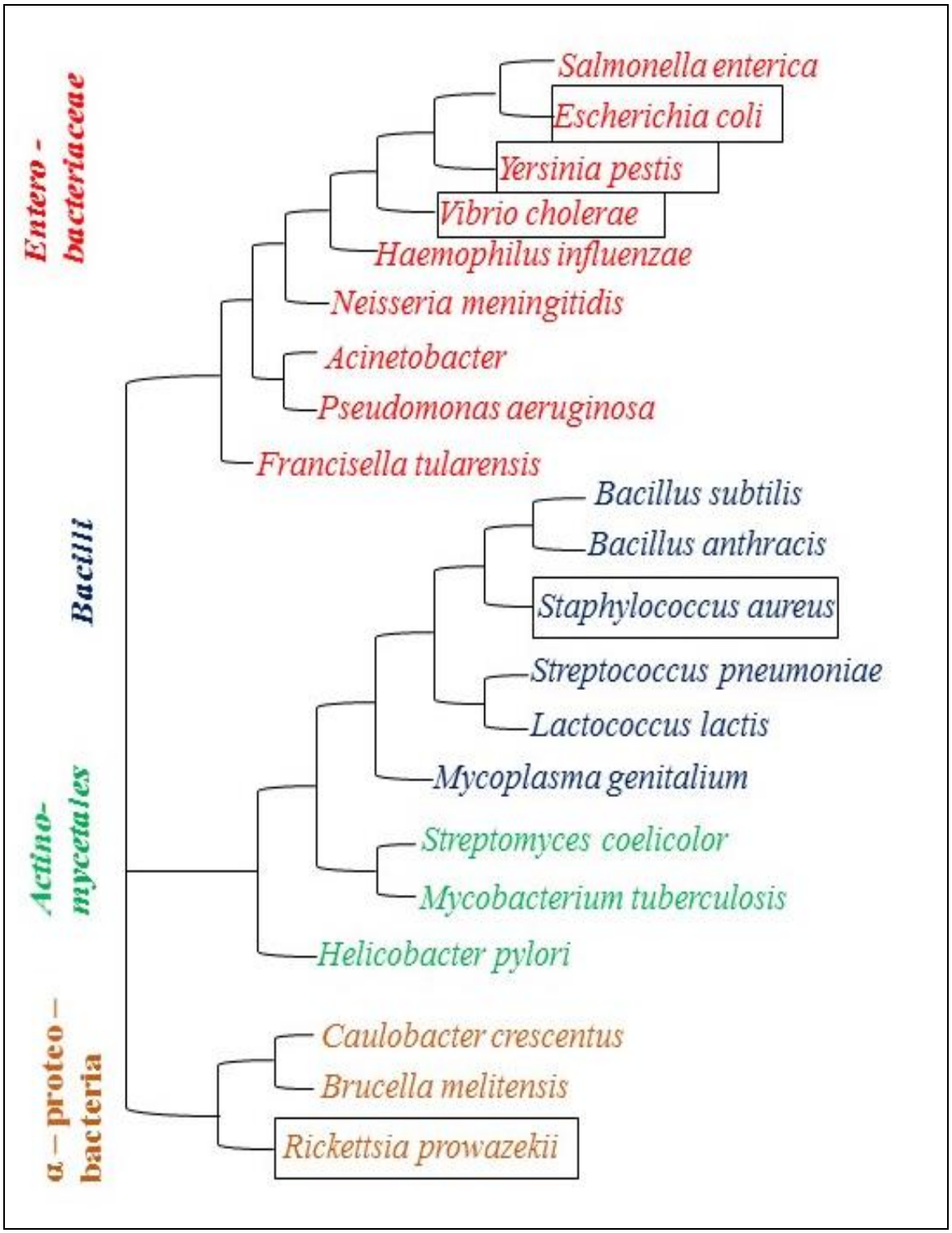


Fig. 3.3 - Phylogenetic tree highlighting the four organisms included in the study

SECTION A – Comparative Interactomics across conserved proteins in pathogenic organisms

A handful of bacterial interactomes have been studied. However, the interactions obtained from these screens cannot be directly compared against each other due to various reasons. Our aim was to create a dataset such that the interactions observed can be compared directly across species. Also, to avoid the issue of low overlap in spite of using the same technique, these interactions among distinct species will be studied in parallel. Thus, we aim to get a dataset of directly comparable interactions conserved across the various species included in the study. The study of conserved interactions will aid in increasing our understanding of the biological function of proteins in the cell as well as help in identifying interactions essential for the organism. These conserved and essential interactions can also be used as drug targets.

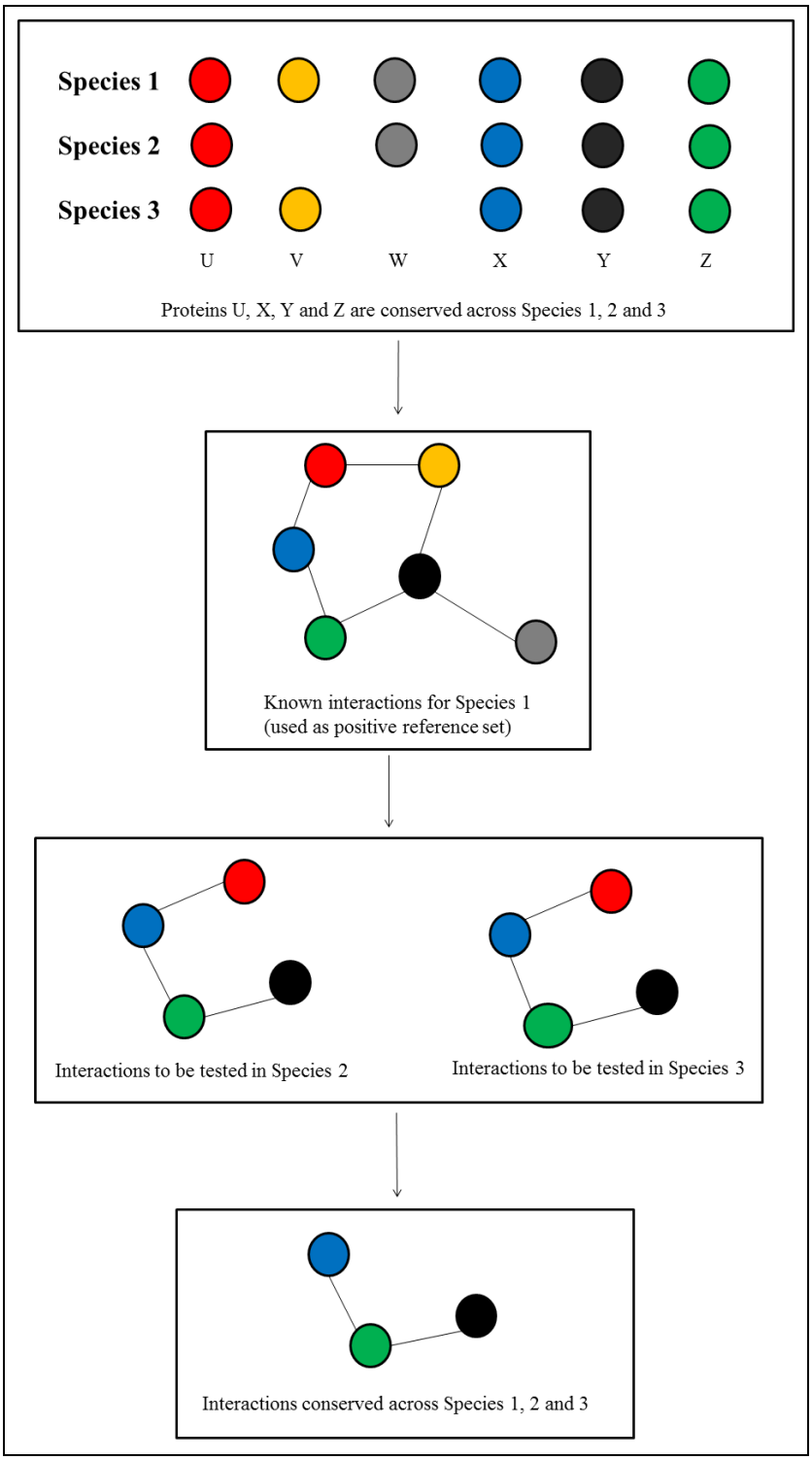


Fig. 3.4 - Concept of comparative interactomics to determine conserved interactions for proteins conserved across *Escherichia coli*, *Yersinia pestis*, *Vibrio cholera*, *Staphylococcus aureus*, and *Rickettsia prowazekii*

Methods

Creating dataset. (Fig. 3.5) Orthologous Groups from EggNOG database version 3.0 (Powell et al., 2012) were used to determine orthologs among the five species. The proteome data for all organisms was obtained from UniProt. All initial data analysis was done using a combination of Microsoft Excel and SAS.

OG data for *Eco, Ype, Vch, Sau* and *Rpr* was obtained from EggNOG database version 3.0

COG ID	COG Name	Eco	Ype	Vch	Sau	Rpr
COG1	description	+				
COG2	description	+	+	+	+	+
COG3	description	+	+	+	+	+
COG4	description	+	+	+		
COGn	description	+	+			

Data was filtered to obtain 395 OGs conserved across all five species (Appendix B Table 3.1)

COG ID	COG Name	Eco	Ype	Vch	Sau	Rpr
COG2	description	+	+	+	+	+
COG3	description	+	+	+	+	+

Binary reference interaction data was obtained for *Eco* from the comprehensive yeast two hybrid assay (Rajagopala et al, 2014)

Bait Uniprot ID	Prey UniProt ID	Bait Protein name	Prey protein Name
P1	P10	A	X
P2	Q8	B	Y
Q1	P10	C	X
Q2	Q6	D	Z
P3	Q7	E	Z'

OGs were mapped to the *Eco* reference dataset resulting in an OG-OG binary interaction dataset

Bait Uniprot ID	Prey UniProt ID	Bait Protein name	Prey protein Name	Bait COG	Prey COG
P1	P10	A	X	COG5	COG3
P2	Q8	B	Y	COG2	COG7
Q1	P10	C	X	COG6	COG3
Q2	Q6	D	Z	COG8	COG9
P3	Q7	E	Z'	COG10	COG9

The OGs were mapped to corresponding proteins in *Ype, Vch, Sau* and *Rpr* (Appendix B Table 3.2)

Bait COG	Prey COG	Ype Bait Uniprot ID	Ype Prey UniProt ID	Vch Bait UniProt ID	Vch Prey UniProt ID	Sau Bait UniProt ID	Sau Prey UniProt ID	Rpr Bait UniProt ID	Rpr Prey UniProt ID
COG1	COG3	V1	V6	Y1	Y6	S1	S6	R1	R6
COG2	COG7	V2	V7	Y2	Y7	S2	S7	R2	R7
COG6	COG3	V3	V6	Y3	Y6	S3	S6	R3	R8
COG4	COG9	V4	V9	Y4	Y9	S4	S9	R4	R9
COG5	COG9	V5	V10	Y5	Y10	S5	S10	R5	R10

Fig. 3.5 – Flowchart demonstrating the process of dataset creation for proteins in *Escherichia coli, Yersinia pestis, Vibrio cholera, Staphylococcus aureus* and *Rickettsia prowazekii*

Selection of protein pairs for experimental testing. Protein pairs were selected based on the presence of their orthologs in all five organisms as well as their presence in the positive interaction dataset for *Eco* interactions by (Rajagopala, *et al*., 2014). All proteins as a result of these two filters were selected and tested.

Construction of bait and prey clones. See Appendix A1.

Yeast transformation. Y2H compatible yeast strains were used for yeast transformation. AH109 (MATa, trp1-901, leu2-3, 112, ura3-52, his3-200, gal4Δ, gal80Δ, LYS2::GAL1UAS-GAL1TATAHIS3, GAL2UAS-GAL2TATA-ADE2, URA3::MEL1 UASMEL1TATA-lacZ, MEL1) and Y187 (MATα, ura3-52, his3-200, ade2-101, trp1-901, leu2-3, 112, gal4Δ, met–, gal80Δ, MEL1, URA3:GAL1UAS -GAL1TATA-lacZ) were used. See Appendix A.

Bait auto activation test. See Appendix A3.

Yeast two hybrid assay. See Appendix A4.

Results

Dataset Analysis. The data set included OG data for five species, *Escherichia coli, Yersinia pestis, Vibrio cholera, Staphylococcus aureus* and *Rickettsia prowazekii*. 395 Clusters of Orthologous Groups (COGs) were conserved across all the species.

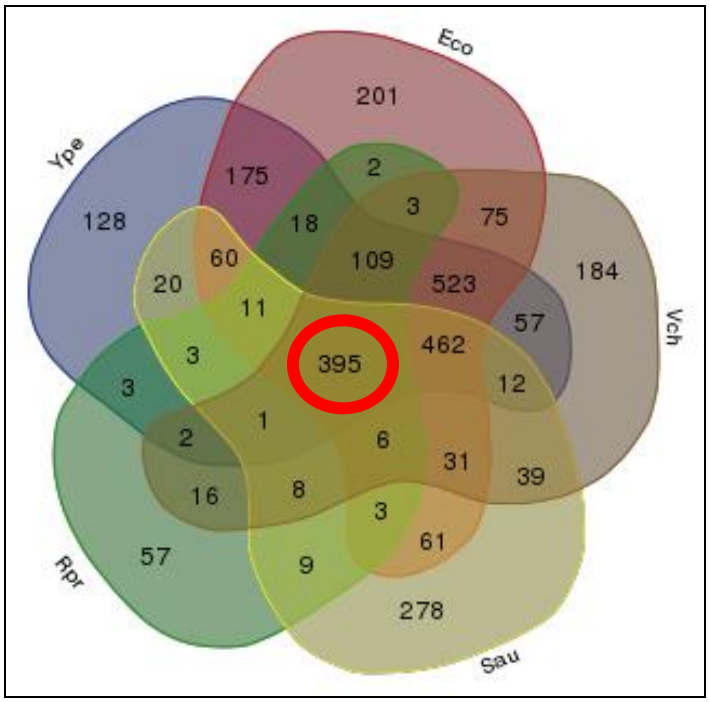


Fig. 3.6 – Number of shared COGs across *Eco*, *Sau*, *Rpr*, *Vch* and *Ype* (Appendix B - Table 3.1)

The reason for low number of consensus OGs was mainly due to inclusion of *Rpr* in the dataset since it is distantly related and also has a reduced genome. The conserved OGs include proteins performing basic functions like transcription, translation, replication, cell division, acyl carrier proteins, ribosomal proteins, etc. More than ~45% of conserved OGs belong to translation, ribosomal structure and biogenesis protein families (Fig. 3.7). These conserved OGs yielded 76 testable interactions in *Eco* based on the Rajagopala, *et al.*, 2014 dataset. The orthologs of these proteins were determined in other organisms and tested using the yeast two hybrid assay.

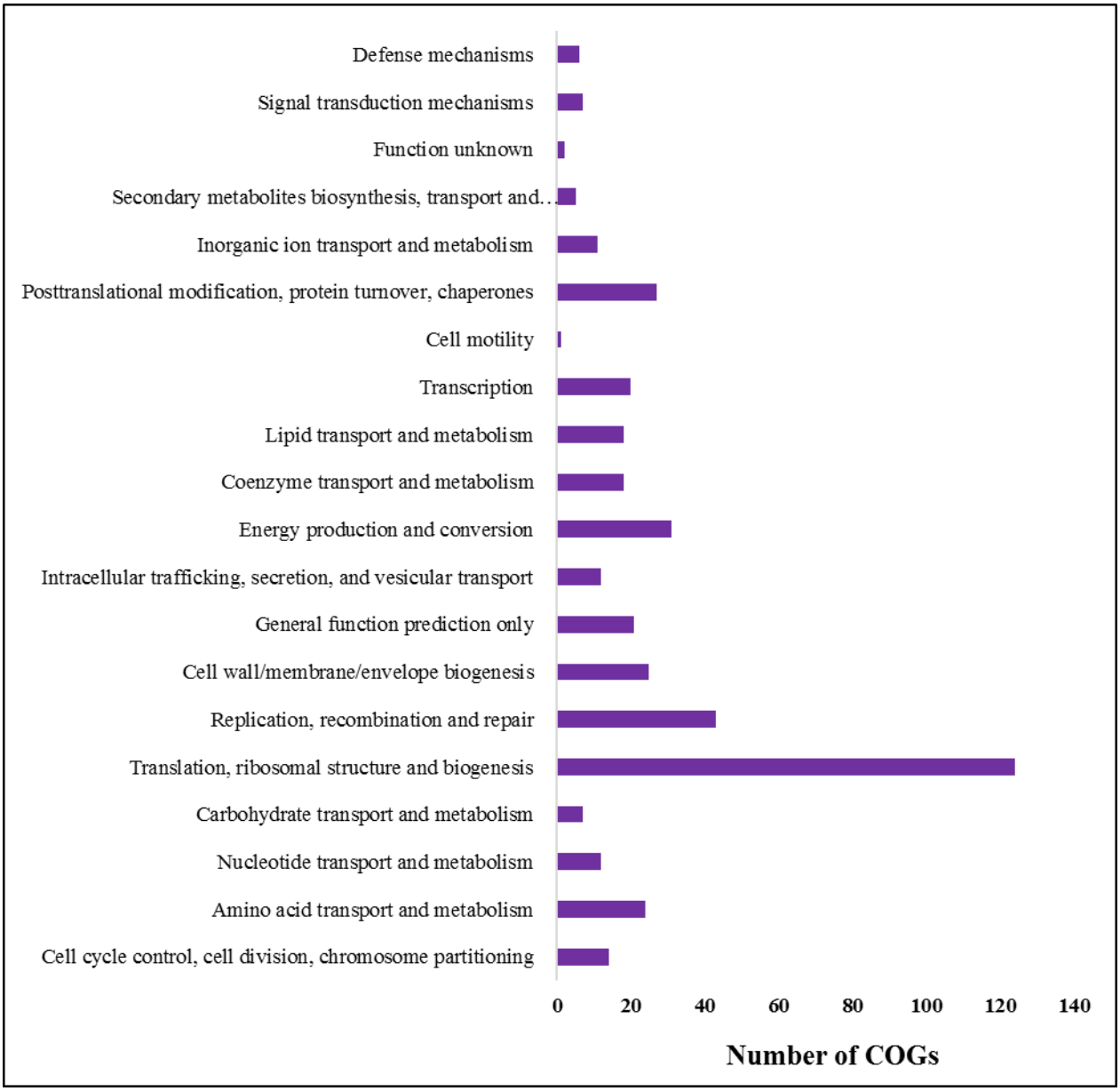
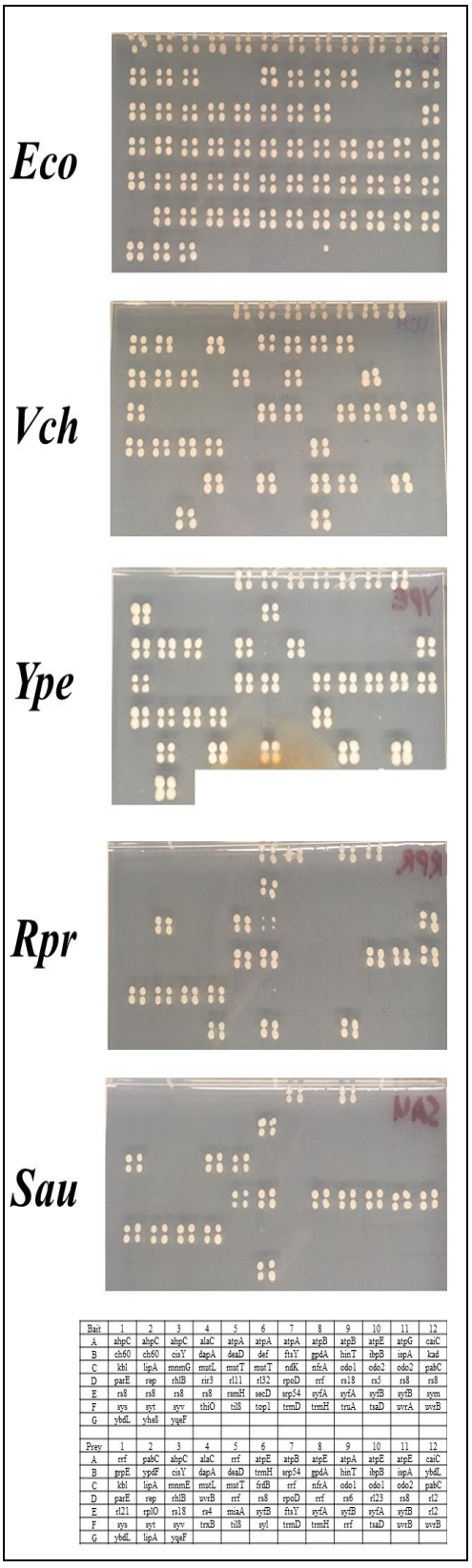


Fig. 3.7 – Distribution of 397 COGs conserved across *Escherichia coli, Yersinia pestis, Vibrio cholerae, Staphylococcus aureus* and *Rickettsia prowazekii* categorized into NCBI functional categories

Experimental Data Analysis. The interactions were tested in parallel using a yeast two hybrid assay. The working concentration of 3-AT was determined by carrying out a bait self-activation

test. Background interactions were observed for baits at 0mM and 1mM 3-AT concentrations. At 3mM 3-AT concentration, there was no background growth and hence 3mM 3-AT was used for screening for interactions in the yeast two hybrid assay.



Eco	1	2	3	4	5	6	7	8	9	10	11	12
A	ahpC	ahpC	ahpC	alaC	atpA	atpA	atpA	atpB	atpB	atpE	atpG	caiC
B	cls60	cls60	csiY	dapA	deaD	def	ftsY	gpdA	hinT	ibpB	ispA	kdsB
C	kbl	lipA	mnmG	murL	murT	murT	ndK	nfrA	odo1	odo2	odo2	pabC
D	parE	rep	rhlB	rsr3	rl11	rl32	rpoD	rrf	rs118	rs5	rs8	rs8
E	rs8	rs8	rs8	rs8	rsmH	secD	srp54	syfA	syfA	syfB	syfB	sym
F	sya	syt	syv	thiO	tilS	top1	trmD	trmH	truA	tsaD	uvrA	uvrB
G	ybdL	yhs8	yqeF									
Prev	1	2	3	4	5	6	7	8	9	10	11	12
A	rrf	pabC	ahpC	alaC	rrf	atpE	atpB	atpE	atpA	atpE	atpE	caiC
B	grpE	ypdF	csiY	dapA	deaD	trmH	srp54	gpdA	hinT	ibpB	ispA	ybdL
C	kbl	lipA	mnmE	murL	murT	frdB	rrf	nfrA	odo1	odo1	odo2	pabC
D	parE	rep	rhlB	uvrB	rrf	rs8	rpoD	rrf	rs6	rl23	rs8	rl2
E	rl21	rpl0	rs118	rs4	miaA	syfB	ftsY	syfA	syfB	syfA	syfB	rl2
F	sya	syt	syv	trmB	tilS	syl	trmD	trmH	rrf	tsaD	uvrB	uvrB
G	ybdL	lipA	yqeF									

Fig. 3.8 – Yeast two hybrid assay at 3mM 3-AT concentration for *Eco*, *Ype*, *Vch*, *Sau*, and *Rpr*

The majority of the proteins in the protein-protein interactions belonged to translational, ribosomal structure and biogenesis functional category. This was an expected outcome since *Rickettsia prowazekii* is very distantly related to the other organisms included in the study. Also, *Rickettsia prowazekii* has a reduced genome and hence protein-protein interactions occurring between proteins required for housekeeping or survival of a cell were more likely to be conserved.

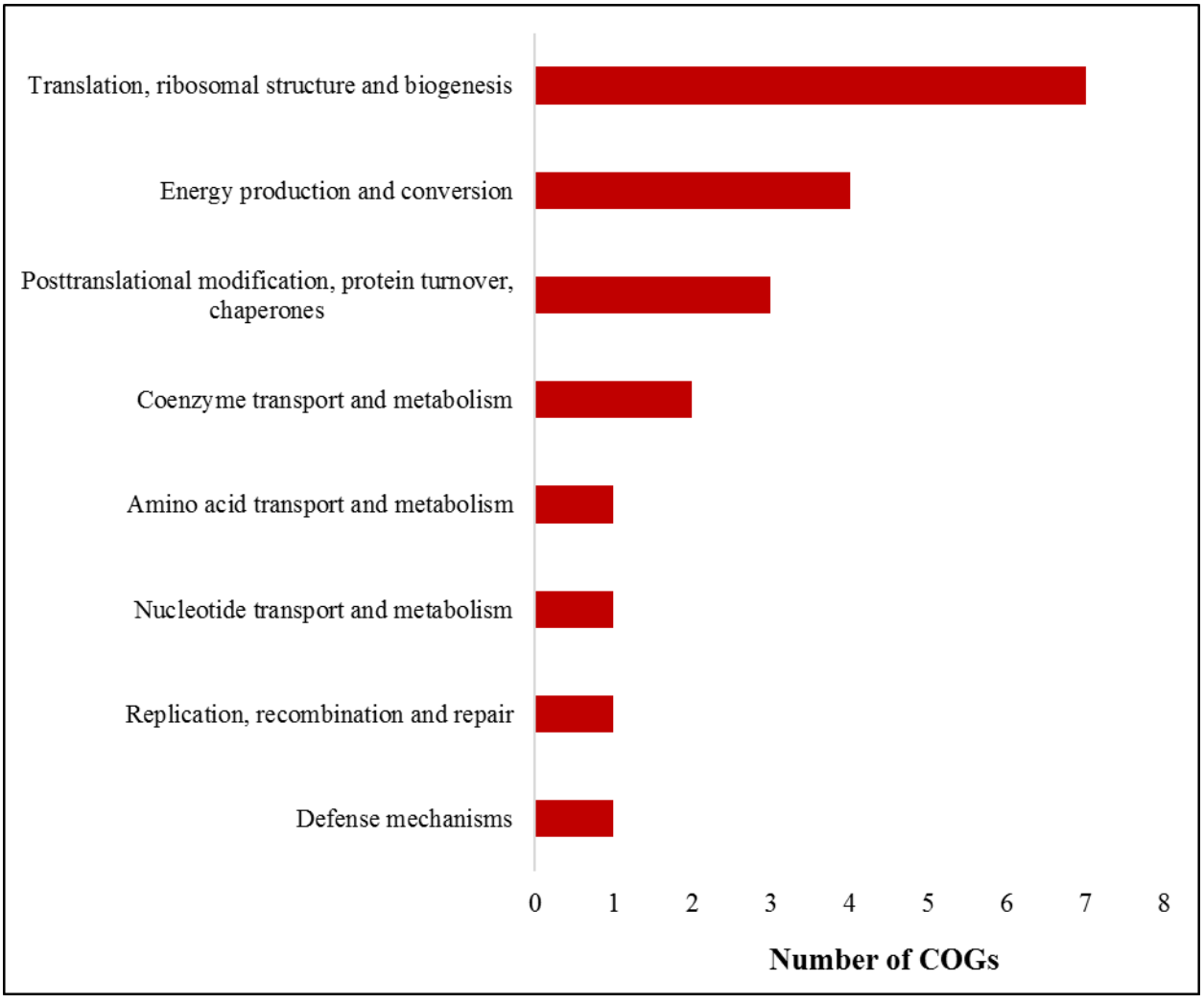


Fig. 3.9 - COGs for interacting proteins tested using yeast two hybrid assay categorized into NCBI functional categories (Appendix B – Table 3.2)

The network in the figure 3.6 includes interactions that can be referred to as high confidence interactions since its includes only those interactions which are observed in at least three of the five species. Some interactions were tested using a single protein once as a bait and once as a prey and hence multiple lines connect the two nodes. The dataset included a few self-interacting proteins as well. Certain proteins stand out as hub proteins in this network including Rs8 (30S ribosomal protein S8), Frr (ribosomal-recycling factor), AtpA (ATP synthase subunit alpha), and AtpE (ATP synthase epsilon chain).

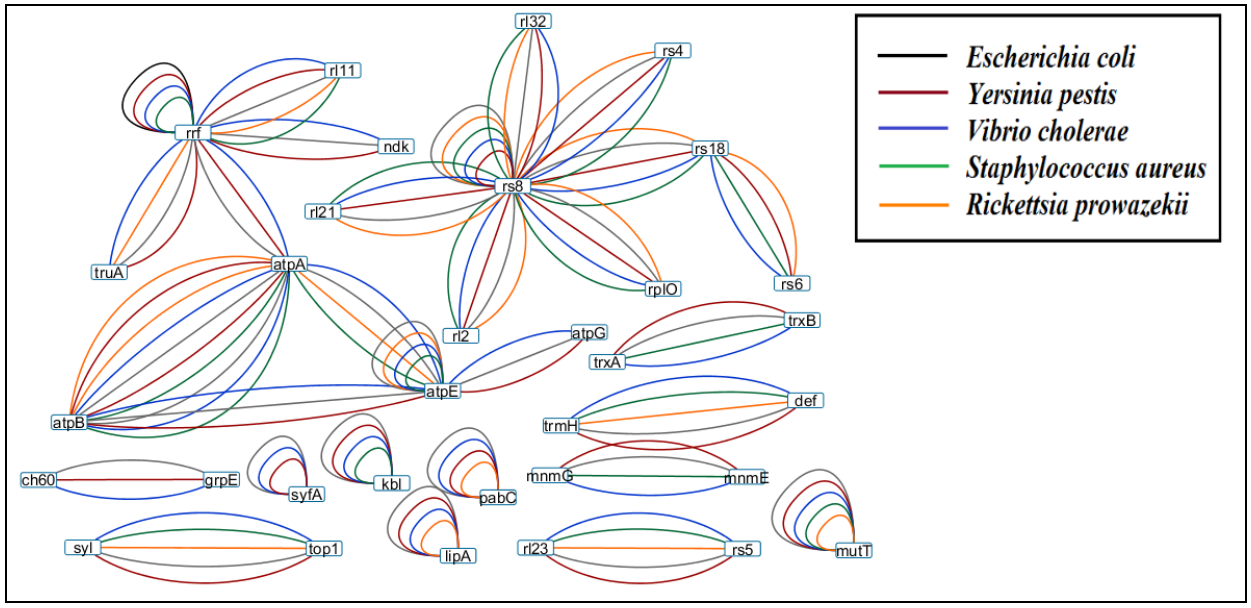


Fig. 3.10 – Network of interactions tested using yeast two hybrid assay across *Eco*, *Vch*, *Ype*, *Sau* and *Rpr* (Appendix B – Table 3.3)

SECTION B – Comparative Interactomics across conserved proteins with unknown function

About 1% of the bacterial proteins have function homology annotations based on experiments, and ~64% have annotations based on sequence, structure or 3D templates. About 35% of these

proteins do not have any functional annotations and are annotated as uncharacterized, unknown, probable or putative. Table 3.1 shows the genome size and percentage of uncharacterized proteins for the organisms used in this study. Different approaches have been used in the past to characterize individual or a family of proteins. One such commonly used approach is network approach which helps in narrowing down function of a protein by 'guilt by association'. Protein networks can be analyzed to get an idea about the function of unknown proteins based on the interactions (Erdin, Lisewski, & Lichtarge, 2011).

Organism	Genome Size (Mb)	Proteins	Uncharacterized Proteins
Escherichia coli W3110	4.64	4,226	1,168 (27.28%)
Yersinia pestis KIM 10+	4.7	4,257	2,189 (51.42%)
Vibrio cholerae O1 biovar El Tor str. N16961	4.03	3,782	1,472 (38.9%)
Staphylococcus aureus COL	2.8	2,775	971 (35%)
Rickettsia prowazekii Madrid E	1.1	842	307 (36.46%)

Table 3.1 – Genome size, proteome and uncharacterized proteins for the organisms used in this study (Ogata et al., 1999)(Kanehisa, *et al.*, 2016)(UniProt, *Retrieved:* September 2016)

The uncharacterized or predicted function proteins will be studied in this case. The interactions will be studied in parallel to enable direct comparison. Our aim is to create a dataset of binary PPIs where either of the interacting proteins will be uncharacterized. Orthologs for these protein pairs will be determined in other species and all interactions will be tested experimentally by yeast two hybrid assay. The interactions conserved across distantly as well as closely related

species will help predict functions for conserved proteins without a known function due to ‘guilt by association’.

Four pathogenic species were chosen based on phylogenetic distance. The reason for low overlap across interactomes of the previously studied bacterial species has been attributed to phylogenetic distance and methodological differences. To determine whether phylogenetic distance has an impact, two closely related (*Yersinia pestis, Vibrio cholerae*) and a distantly related organism (*Staphylococcus aureus)* were selected. *Escherichia coli* is the only organism in this dataset that has previous large scale interaction data using Y2H. Hence the *Escherichia coli* interactome was used as a reference dataset and all analysis were based on that dataset.

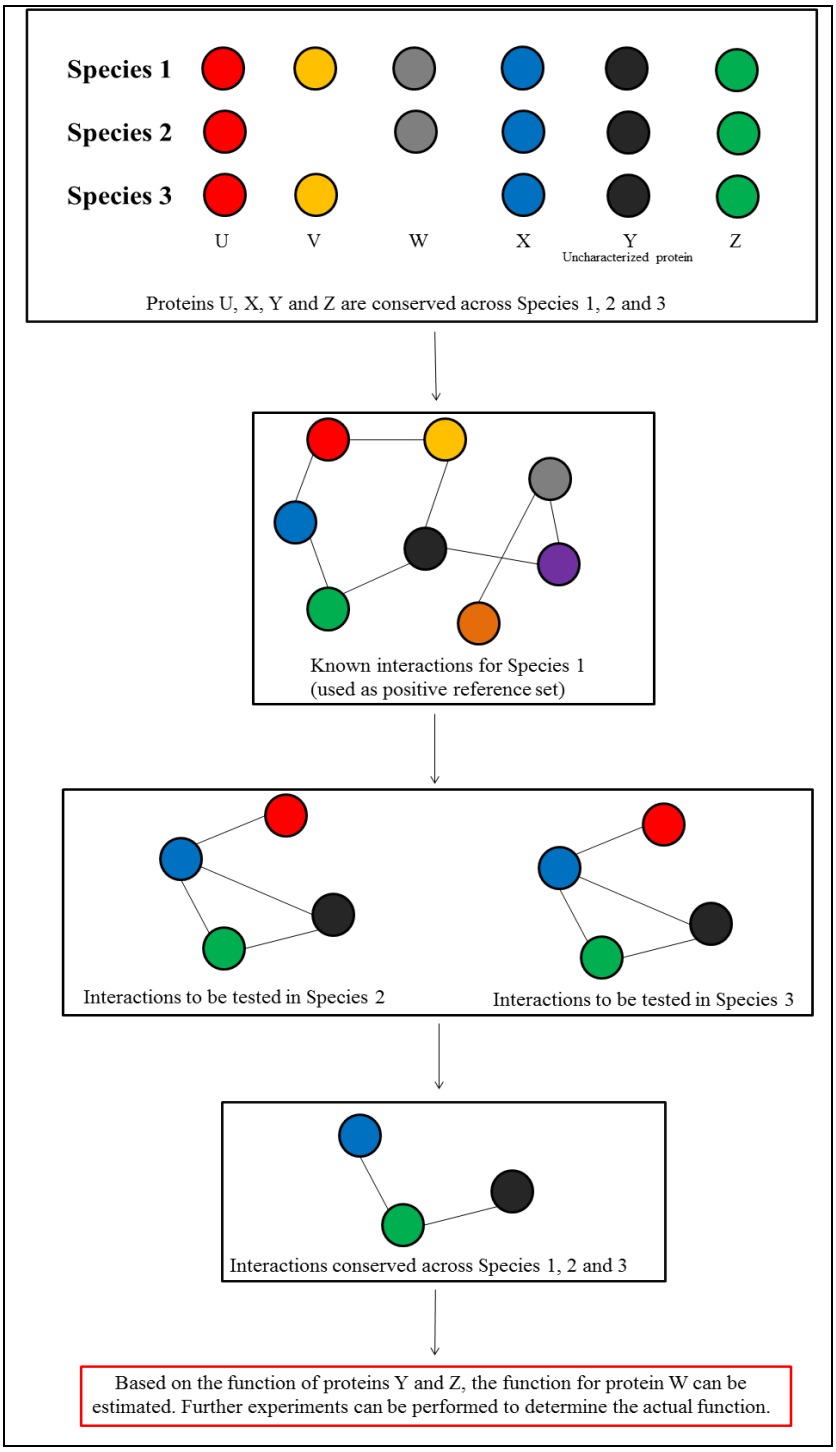


Fig. 3.11 - Concept of comparative interactomics to determine conserved interactions for unknown function proteins in *Escherichia coli, Yersinia pestis, Vibrio cholerae,* and *Staphylococcus aureus*

Methods

Creating dataset. (Fig. 3.12) Orthologous Groups from EggNOG database version 3.0 *(Powell et al.*, 2012) were used to determine orthologs among the four species. These OGs were filtered to create a list of OGs that were annotated as 'uncharacterized', 'predicted', 'unknown' or 'probable'. The proteome data for all organisms was obtained from UniProt. All initial data analysis was done using a combination of Microsoft Excel and SAS.

OG data for Eco, Ype, Vch and Sau was obtained from Eggnog version 3.0

COG ID	COG Name	*Eco*	*Ype*	*Vch*	*Sau*
COG1	Uncharacterized	+	+	+	+
COG2	Predicted XYZ	+	+	-	-
COG3	Transcription factor	+	+	+	+
COG4	Probable ABC	+	+	-	-
COG5	Predicted protein	+	+	+	+
COGn	Ribosomal protein	+	+	+	+

↓

Data was filtered to obtain 857 OGs present in all four species, among which 167 OG's belonged to uncharacterized proteins (Appendix B - Table 3.4)

COG ID	COG Name	*Eco*	*Ype*	*Vch*	*Sau*
COG1	Uncharacterized	+	+	+	+
COG5	Predicted protein	+	+	+	+

↓

Binary reference interaction data was obtained for *Eco* from the comprehensive yeast two hybrid assay (Rajagopala *et al,* 2014) and proteins annotated as uncharacterized/predicted were selected resulting in 94 testable interactions

Bait UniProt ID	Prey UniProt ID	Bait Protein Name	Prey Protein Name
P1	P3	A	D
P2	Q2	B	D'
P4	P5	C	E

↓

The UniProt ID were mapped to OGs resulting in OG-OG binary interaction dataset

Bait UniProt ID	Prey UniProt ID	Bait Protein Name	Prey Protein Name	Bait COG	Prey COG
P1	P3	A	D	COG1	COG3
P2	Q2	B	D'	COG5	COG3
P4	P5	C	E	COG10	COG8

↓

OGs were mapped to corresponding proteins in *Ype, Vch* and *Sau* (Appendix B Table 3.5)

Bait COG	Prey COG	Ype Bait UniProt ID	Ype Prey UniProt ID	Vch Bait UniProt ID	Vch Prey UniProt ID	Sau Bait UniProt ID	Sau Prey UniProt ID
COG1	COG3	Y1	Y4	V1	V4	S1	S4
COG5	COG3	Y2	Y5	V2	V5	S2	S5
COG10	COG8	Y3	Y6	V3	V6	S3	S6

Fig. 3.12 – Flowchart demonstrating the process of dataset creation for unknown function proteins in *Escherichia coli, Yersinia pestis, Vibrio cholera,* and *Staphylococcus aureus*

Selection of protein pairs for experimental testing. Protein pairs were selected based on the presence of orthologs in all four organisms as well as their presence in the positive interaction dataset for *Eco* interactions by Rajagopala *et al*. The above dataset was filtered for proteins annotated as 'uncharacterized', 'putative', 'predicted' or 'probable' in *Escherichia coli*. All proteins as a result of these two filtering steps were selected and tested.

Construction of bait and prey clones. See Appendix A1.

Yeast transformation. Y2H compatible yeast strains were used for yeast transformation. AH109 (MATa, trp1-901, leu2-3, 112, ura3-52, his3-200, gal4Δ, gal80Δ, LYS2::GAL1UAS-GAL1TATAHIS3, GAL2UAS-GAL2TATA-ADE2, URA3::MEL1 UASMEL1TATA-lacZ, MEL1) and Y187 (MATα, ura3-52, his3-200, ade2-101, trp1-901, leu2-3, 112, gal4Δ, met–, gal80Δ, MEL1, URA3:GAL1UAS -GAL1TATA-lacZ) were used. See Appendix A2.

Bait auto activation test. See Appendix A3.

Yeast two hybrid assay. See Appendix A4.

Results

Dataset Analysis. The data set included OG data for four species, *Escherichia coli, Yersinia pestis, Vibrio cholera,* and *Staphylococcus aureus*. 857 Cluster Orthologous Groups (COGs) were conserved across these species.

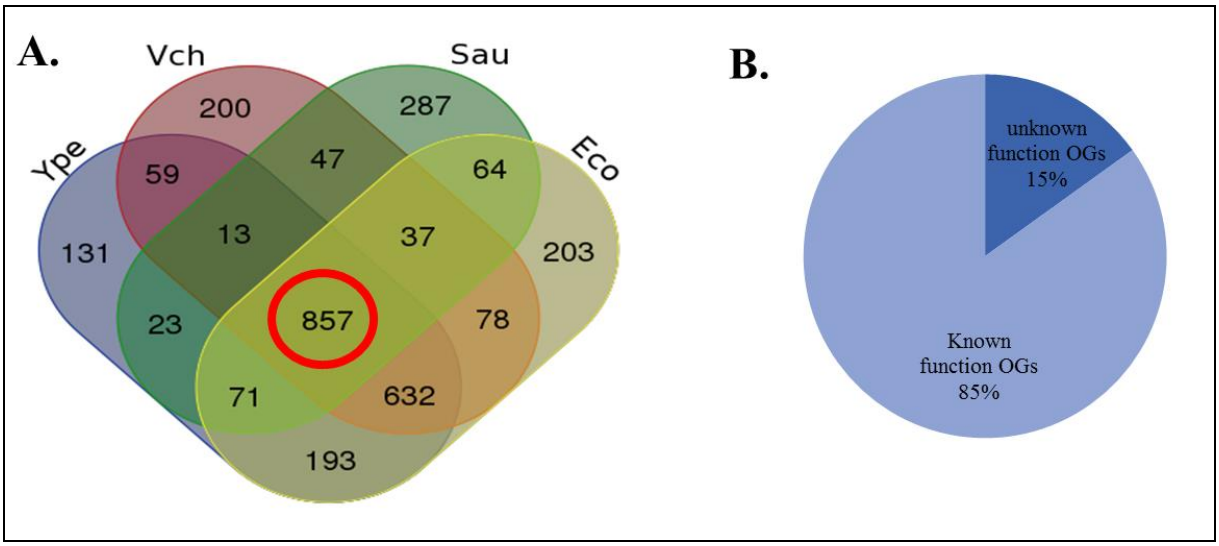


Fig. 3.13 – A. Venn diagram denotes 857 COGs are conserved across *Escherichia coli, Yersinia pestis, Vibrio cholera,* and *Staphylococcus aureus.* B. 15% of conserved OGs were annotated as unknown while the rest of the conserved OGs belonged to proteins with a known function

Among the conserved OGs, 15% OGs belonged to proteins annotated as unknown, uncharacterized, putative or probable in *Escherichia coli*. For this study, *Eco* was considered as the reference organism and hence unknown function proteins from *Eco* were used for testing interactions. Among the uncharacterized proteins in *Eco* 505 proteins were singleton proteins not assigned a COG) and the remaining 715 were categorized in 540 OGs. (Fig. 3.13) Among these 540 OGs, 167 OGs were conserved across all the organisms included in this study. These 167 OGs yielded 94 protein – protein interactions in *Eco* based on the Rajagopala *et al.*, 2014 dataset. This dataset included proteins such that at least one of the interacting proteins was annotated as unknown function protein.

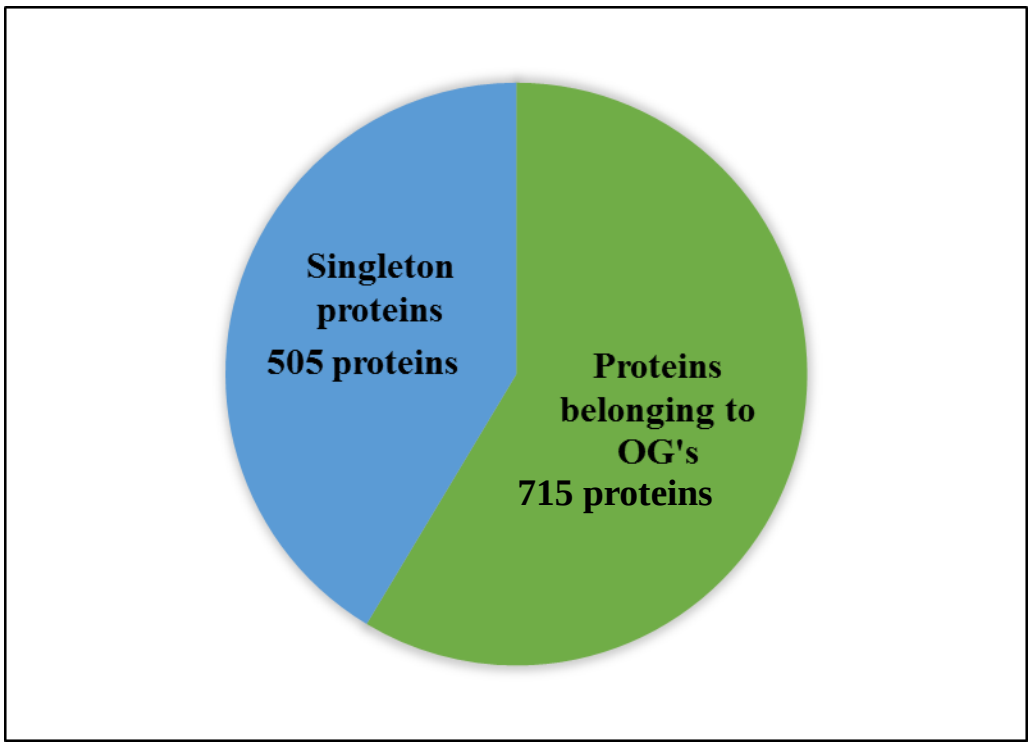


Fig. 3.14 – Uncharacterized proteins in *Escherichia coli K12* proteome categorized into singleton proteins and proteins categorized into OGs

All interactions from the dataset could not be tested due to various reasons. One of the main reason for low testable interactions was the fact that the protein was not present in the clone set (Fig. 3.16) though the OG was present in the organism. Another reason for lower number of testable interactions in organisms other than *Eco* could be due to absence or lesser number of paralogs in these organisms, i.e. A and A' are paralogs and hence are categorized under the same OG. A and A' interact with protein F in *Eco*, however A' is not present in *Ype* and hence the interaction cannot be tested.

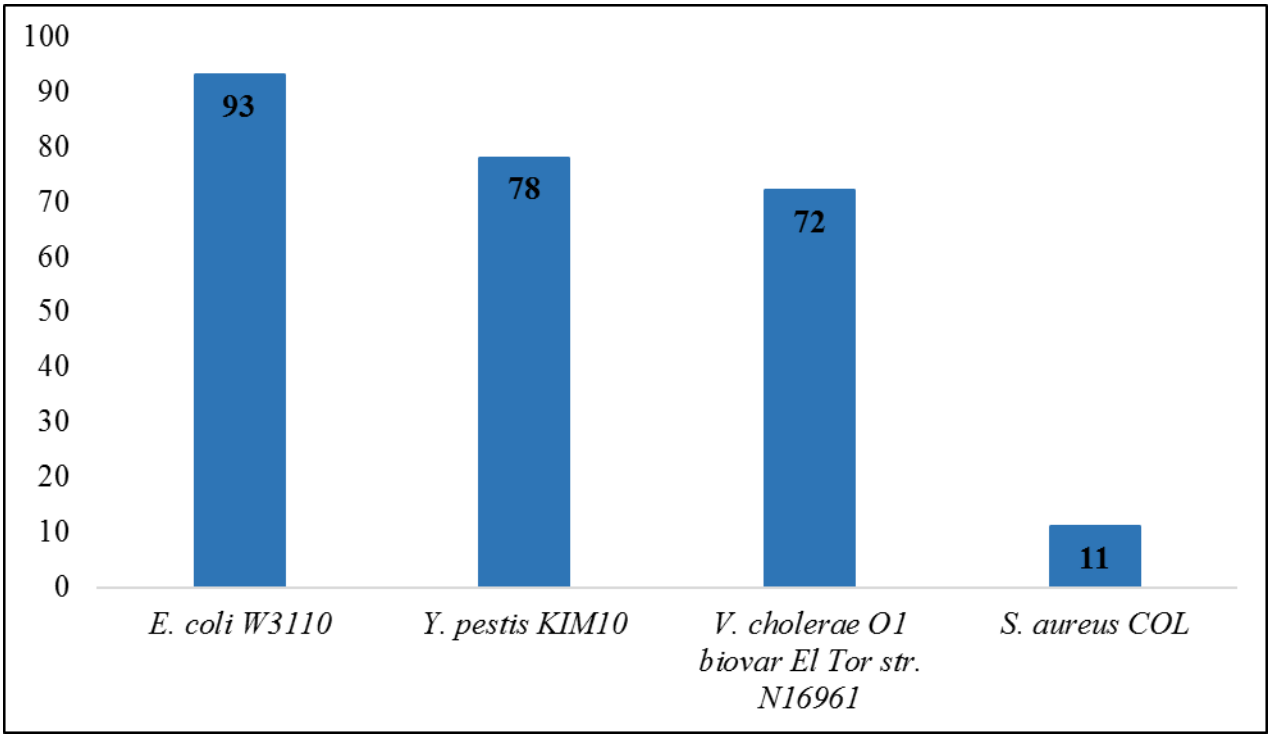


Fig. 3.15 – Predicted protein – protein interactions based on OGs for proteins present in clone set for unknown/ uncharacterized proteins

All conserved COGs were categorized into NCBI COG categories. These OG were classified into proteins with known function and proteins of unknown function. The majority of the known function conserved COGs performed functions related to translation, ribosomal structure, and biogenesis followed by COGs representing proteins having functions in different metabolic processes. The unknown function COGs mainly belonged to function unknown and general function prediction only. However, some of the COGs coding for uncharacterized proteins belonged to functionally annotated COG categories including translation, ribosomal structure, and biogenesis as well as transcription. For example, proteins YafY and YfjR are annotated as lipoprotein YafY and uncharacterized transcriptional regulator YfjR. Both proteins belong to COG2378 which is annotated as COG with proteins involved in Transcription function.

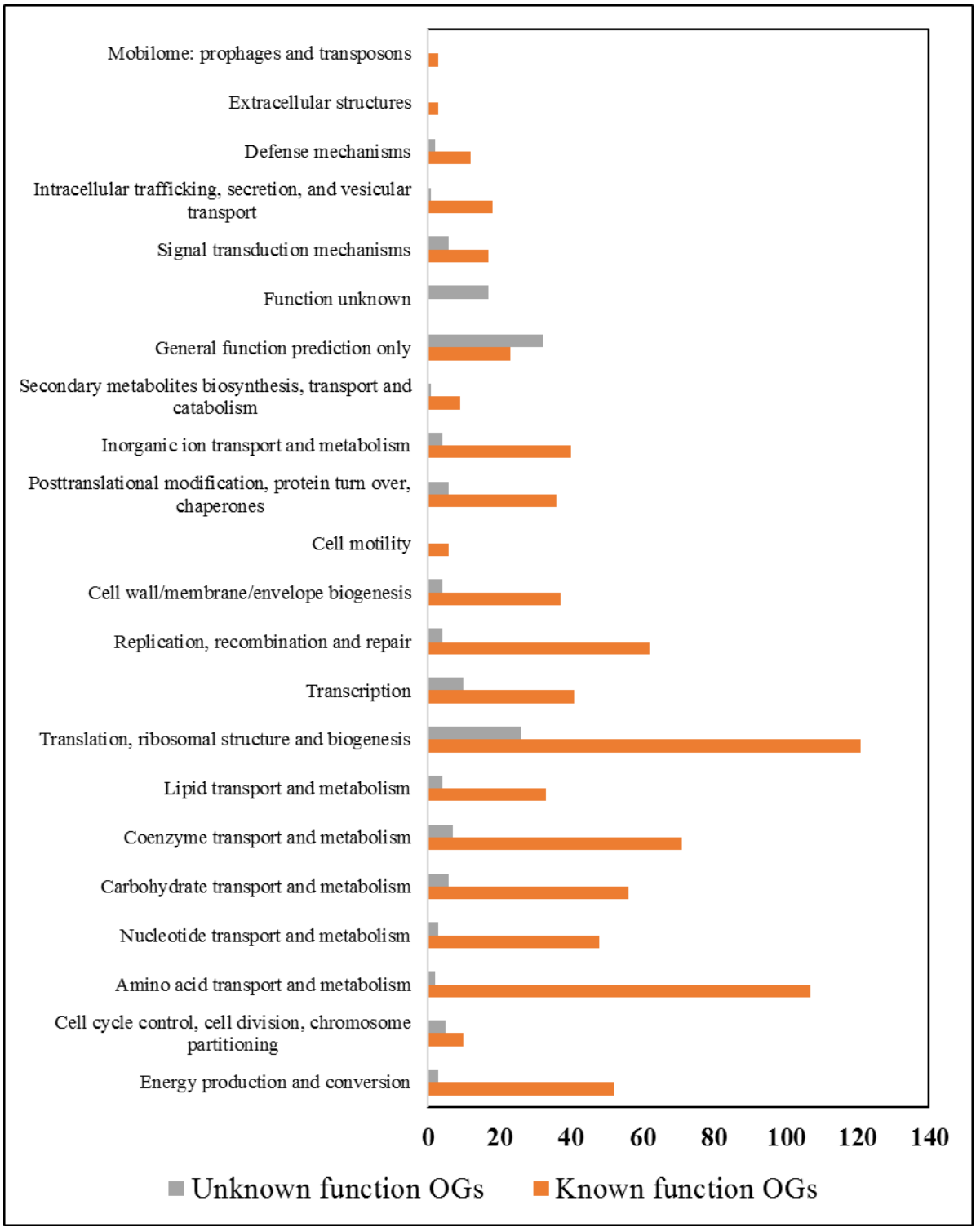
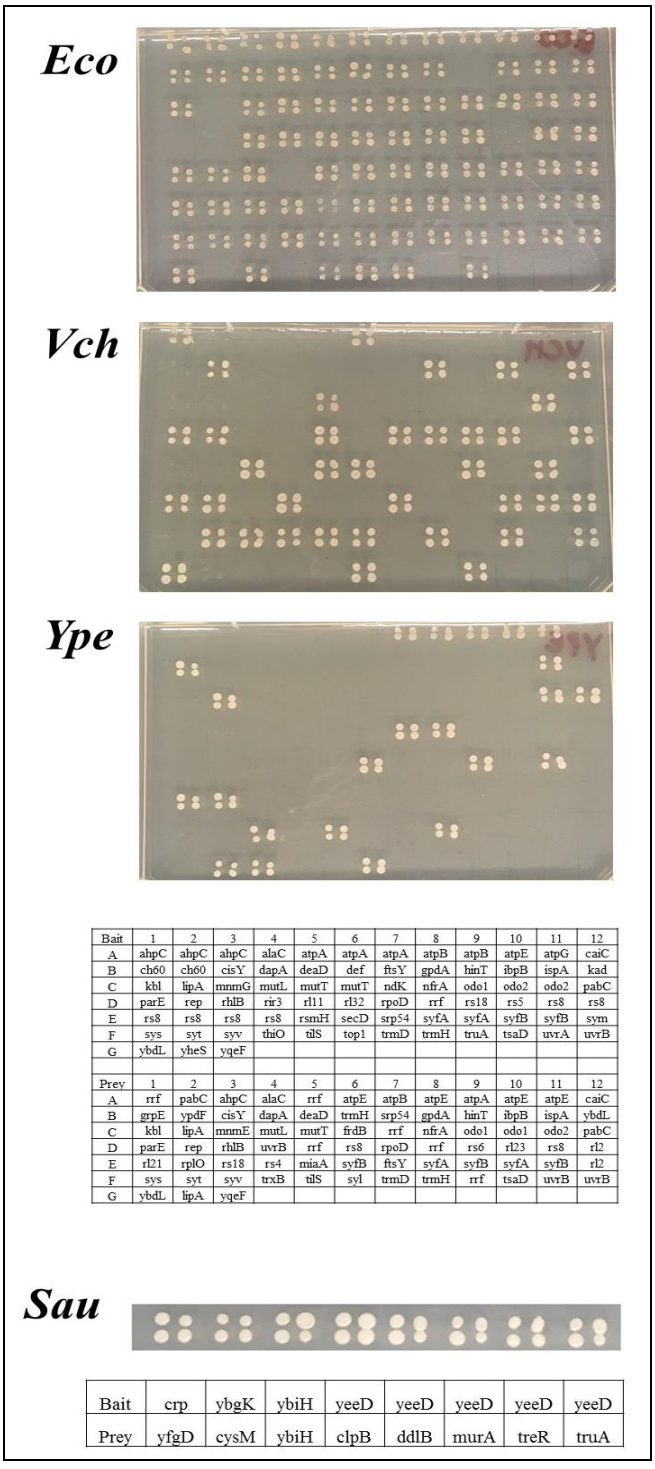


Fig. 3.16 - OGs organized into NCBI functional categories. Distribution of all OGs conserved across the four species categorized into functional categories for OGs of known and unknown function.

Experimental Analysis. The interactions were tested in parallel using yeast two hybrid system. The concentration for 3-AT was optimized during bait self-activation test. A clear background signal for self-activation was observed at 3mM 3-AT concentration for all baits (Fig. 3.17). However, 10mM 3-AT showed no self-activation for baits for either of the four species and hence was used for the yeast two hybrid assay.



Bait	1	2	3	4	5	6	7	8	9	10	11	12
A	ahpC	ahpC	ahpC	alaC	atpA	atpA	atpA	atpB	atpB	atpE	atpG	caiC
B	ch60	ch60	cisY	dapA	deaD	def	ftsY	gpdA	hinT	ibpB	ispA	kad
C	kbl	lipA	mnmG	mutL	mutT	mutT	ndK	nfrA	odo1	odo2	odo2	pabC
D	parE	rep	rhlB	rir3	rl11	rl32	rpoD	rrf	rs18	rs5	rs8	rs8
E	rs8	rs8	rs8	rs8	rsmH	secD	srp54	syfA	syfA	syfB	syfB	sym
F	sys	syt	syv	thiO	tilS	top1	trmD	trmH	truA	tsaD	uvrA	uvrB
G	ybdL	yheS	yqeF									

Prey	1	2	3	4	5	6	7	8	9	10	11	12
A	rrf	pabC	ahpC	alaC	rrf	atpE	atpB	atpE	atpA	atpE	atpE	caiC
B	grpE	ypdF	cisY	dapA	deaD	trmH	srp54	gpdA	hinT	ibpB	ispA	ybdL
C	kbl	lipA	mnmE	mutL	mutT	frdB	rrf	nfrA	odo1	odo1	odo2	pabC
D	parE	rep	rhlB	uvrB	rrf	rs8	rpoD	rrf	rs6	rl23	rs8	rl2
E	rl21	rplO	rs18	rs4	miaA	syfB	ftsY	syfA	syfB	syfA	syfB	rl2
F	sys	syt	syv	trxB	tilS	syl	trmD	trmH	rrf	tsaD	uvrB	uvrB
G	ybdL	lipA	yqeF									

Bait	crp	ybgK	ybiH	yeeD	yeeD	yeeD	yeeD	yeeD
Prey	yfgD	cysM	ybiH	clpB	ddlB	murA	treR	truA

Fig. 3.17 – Yeast two hybrid assay at 10mM 3-AT concentration for

A. *Eco* B. *Ype* C. *Vch* D. *Sau*

Majority of the proteins in the protein-protein interactions tested positive using yeast two hybrid assay belonged to transcription, followed by posttranslational modifications functional category. This was an expected outcome since many proteins are indirectly involved in the process of transcription which is required for cell survival. It was also observed that most of the functional categories observed for conserved interactions belonged to housekeeping functions for a cell that are required for cell survival.

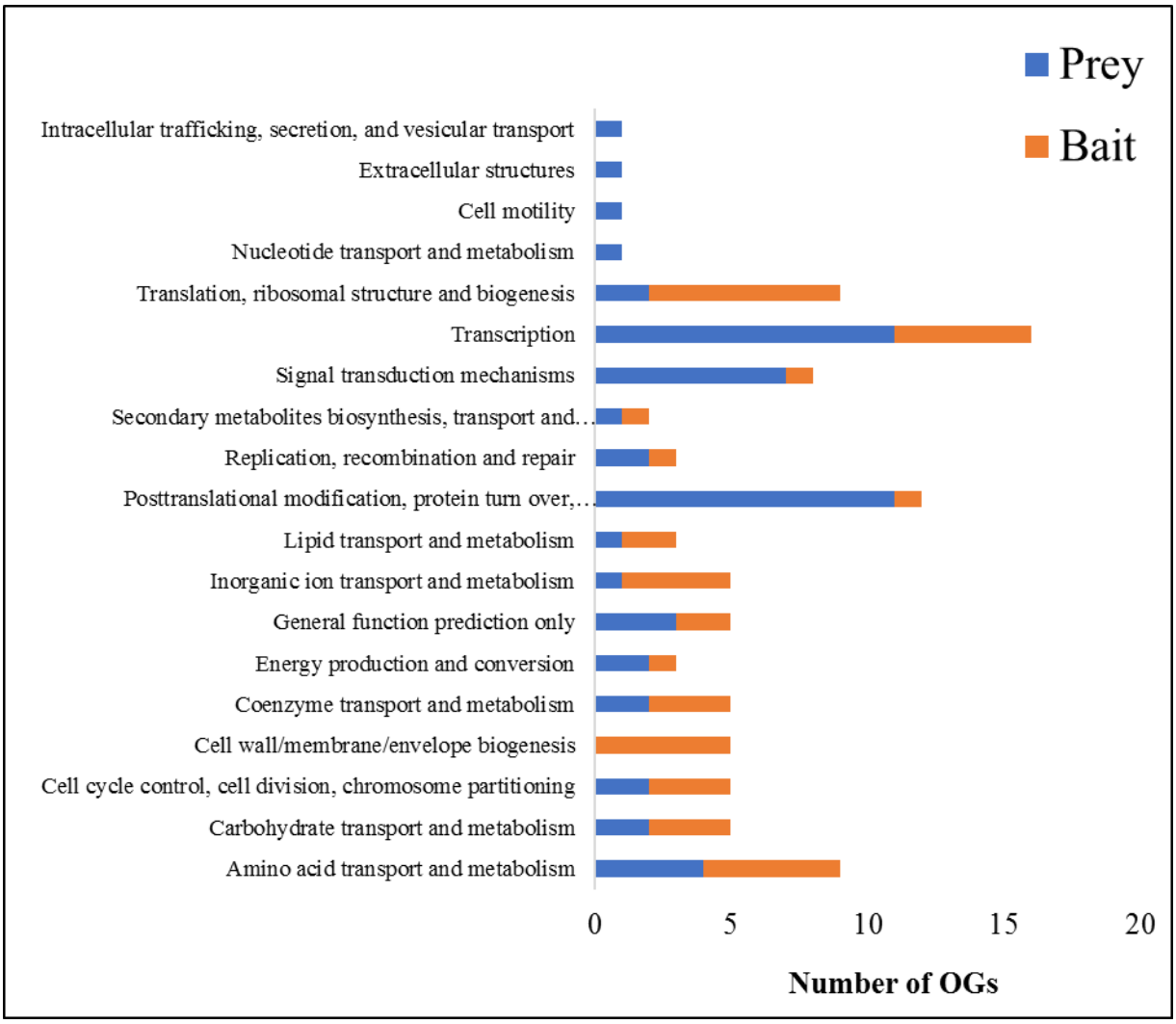


Fig. 3.18 – Unknown function bait and prey OGs to be screened by yeast two hybrid assay categorized into NCBI functional categories

The network in the figure 3.19 includes all interactions observed in the assay. It can be observed that most (82 interactions among 93 tested interactions) of interactions were reproducible in *Eco*. However, as few as 17 (<30%) interactions were detected positive in *Ype*, 16 (<30%) interactions were found positive in *Vch* and 3 interactions were observed in *Sau*. The network is also not a very well connected network and hence protein function predictions are impossible to make in this case. However, the network does show a few proteins that can be denoted as hub

proteins for this network, YbdM and YkgN. YkgN protein could not be obtained in *Sau* and hence was tested only in *Eco, Ype* and *Vch*.

YbdM has been annotated as uncharacterized protein YbdM and belongs to NCBI functional category for cell cycle control, cell division, chromosome partitioning. YbdM shows higher confidence interactions with EutD (Ethanolamine utilization protein EutD), AcrR (HTH-type transcriptional regulator AcrR) and Hcr (NADH oxidoreductase HCR) since these interactions were detected in more than one organisms in the yeast two hybrid screens. Hcr and AcrR have functions related to sequence specific DNA binding. EutD on the other hand is involved in ethanolamine catabolic process and response to heat. Protein orthologous to YbdM in *Bacillus subtilis* (PrkD) has been annotated as a protein kinase that phosphorylates proteins with crucial roles in DNA metabolism (Shi et al., 2014). Based on the interaction data as well as orthology data this protein can be said to have a role that assists in the process of DNA metabolism and/or transcription. Further investigation of function for this protein can be done based on this predicted function.

YkgN has been annotated as a putative transposase based on its domain structure (pfam PF01527). It belongs to NCBI COG category mobilome: prophages, transposons. This protein has orthologs in *Citrobacter spp, Enterobacter spp, Yersinia spp, Vibrio spp* and some other closely related species. Protein alignment in *Citrobacter spp, Enterobacter spp, Yersinia spp, Vibrio spp* shows greater 95% similarity. However, protein function has not been studied experimentally for any of the orthologs. For all orthologs a probable transposase function has been assigned. YkgN shows conserved interactions with few other proteins having very distinct functions. The protein interactions in this case cannot be used to predict function of YkgN.

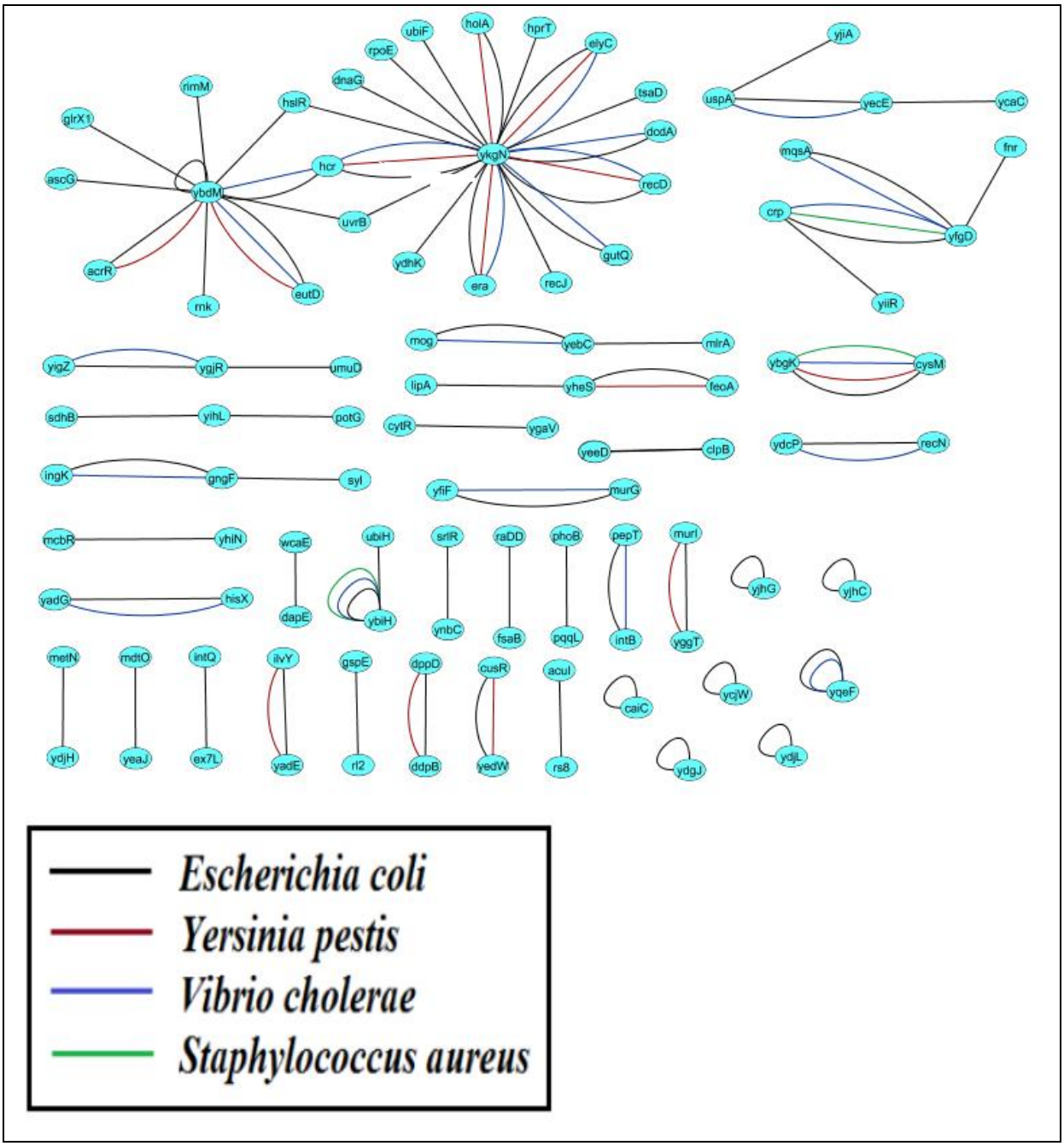


Fig. 3.19 – Network of interactions tested using yeast two hybrid assay across *Eco*, *Vch*, *Ype*, and *Sau* (Appendix B – Table 3.6)

Discussion

This chapter includes the study of two independent datasets. One of the datasets (Section A, Appendix B Table 3.2) includes proteins conserved across five pathogenic bacterial species and the other dataset (Section B, Appendix B Table 3.5) is focused on uncharacterized proteins across four more closely related bacterial species. The PPIs for these two datasets were screened in parallel using yeast two hybrid assay to yield a high confidence interaction set and to be able to predict functions in case of uncharacterized proteins.

Section A - Comparative Interactomics across conserved proteins in pathogenic organisms

Selection of Species. The species used for the study were selected based on availability of clone sets as well as their phylogenetic distance in comparison with *Escherichia coli.* The species used in this case were *Escherichia coli, Yersinia pestis, Vibrio cholerae, Staphylococcus aureus* and *Rickettsia prowazekii*. Though *Escherichia coli, Yersinia pestis, Vibrio cholerae,* are very closely related, *Staphylococcus aureus,* and *Rickettsia prowazekii* are quite distantly related to *Eco.* Furthermore, *Rickettsia prowazekii* also has a reduced genome (~800 genes). Hence, the testable interaction dataset in this case was very small.

Proteins involved in the interactions. All the proteins in the interaction dataset were known function proteins. The yeast interaction screens detected a high number of interactions (~ 50%) conserved in at least three of the five species included in the study. The conserved interactions involved proteins responsible for translation, ribosomal function and biogenesis; energy production and conversion; transcription; etc. Since these functions are required for survival and basic functioning of a cell it was an expected outcome.

SECTION B – Comparative Interactomics across conserved proteins with unknown function

Selection of Species. The dataset included only uncharacterized proteins. The species used in this case were *Escherichia coli, Yersinia pestis, Vibrio cholerae,* and *Staphylococcus aureus*. *Rickettsia prowazekii* is a reduced genome and hence no testable interactions involving unknown function proteins could be yielded based on *Escherichia coli* yeast two hybrid screens. Hence, *Rickettsia prowazekii* was not included in this dataset.

Proteins involved in the interactions. Majority of these proteins could be revived and obtained from clonesets, however, though the COG for unknown function protein YkgN was present in *Sau,* the protein was not present in the cloneset and hence could not be tested for interactions.

Protein – protein interactions. In this case, very few PPIs (<30%) were conserved. *Sau* showed <5% PPIs in comparison to *Eco*. The major reason for this low overlap can be attributed to absence of proteins in the *Sau* clone set. This dataset also showed conserved interaction proteins belonging to translation, ribosomal function and biogenesis. This dataset also showed conserved proteins having function related to posttranslational modifications and transcription. The network was not very well connected and hence based on one or two interactions function predictions for proteins could not be made.

Unknown function protein. YkgN appeared to be a major hub protein in the interaction network for uncharacterized proteins. However, proteins YbdM and YkgN interacted with numerous other proteins and showed s few conserved interactions. Based on literature and interactions observed, YbdM can be predicted to have a function as a kinase in DNA binding related cell activities. YkgN could not be predicted for a function based on interactions, but based on Pfam (PF01527) and sequence analysis it can be annotated as a transposase.

CHAPTER IV

COMPARATIVE INTERACTOMICS FOR UNCHARACTERIZED HUB PROTEINS

Introduction

Proteins interact with each other to perform complex cellular functions. These interactions have been studied in the form of a network/graph for carrying out big picture analysis. In network theory, the proteins that interact with multiple proteins are called as 'hub' proteins.

Scientists have made various attempts at defining hub proteins to enable their easy identification in networks and to derive biological insights. Two criteria that are generally used to define or identify hub proteins are their connectivity and conservation. Hub proteins interact with numerous other proteins but they rarely interact with each other, i.e. they have lower connectivity with each other. Hub proteins are generally essential to the organism and have a high probability to be conserved across species (Vallabhajosyula, *et al.*, 2009). This may be due to their interaction with numerous other proteins making them physiologically more important thus leading to a slow evolution rate (Batada, *et al.*, 2006).

Generally high throughput interaction data leads to huge networks where some proteins fit the above criteria to be identified at hub proteins based on network analysis. Interactomes for organisms can be studied using various methods. This chapter will focus on interactomes studied using the yeast two hybrid technique. Interactomes for only a handful of bacterial species have been studied using yeast two hybrid techniques, namely, *Escherichia coli*, *Treponema pallidum*,

Helicobacter pylori, and *Campylobacter jejuni*. These interactomes include ~80-85% of the proteome. Network analysis of these interactomes reveals that a majority of the proteins interact with one of two proteins. Also, very small percentage of proteins have more than 100 interacting partners. The second figure (Fig. 4.1 A, 4.1 B) shows that hub proteins (proteins having more than 8 interacting partners) mostly cluster together i.e. a majority of proteins have less than 30 interacting proteins.

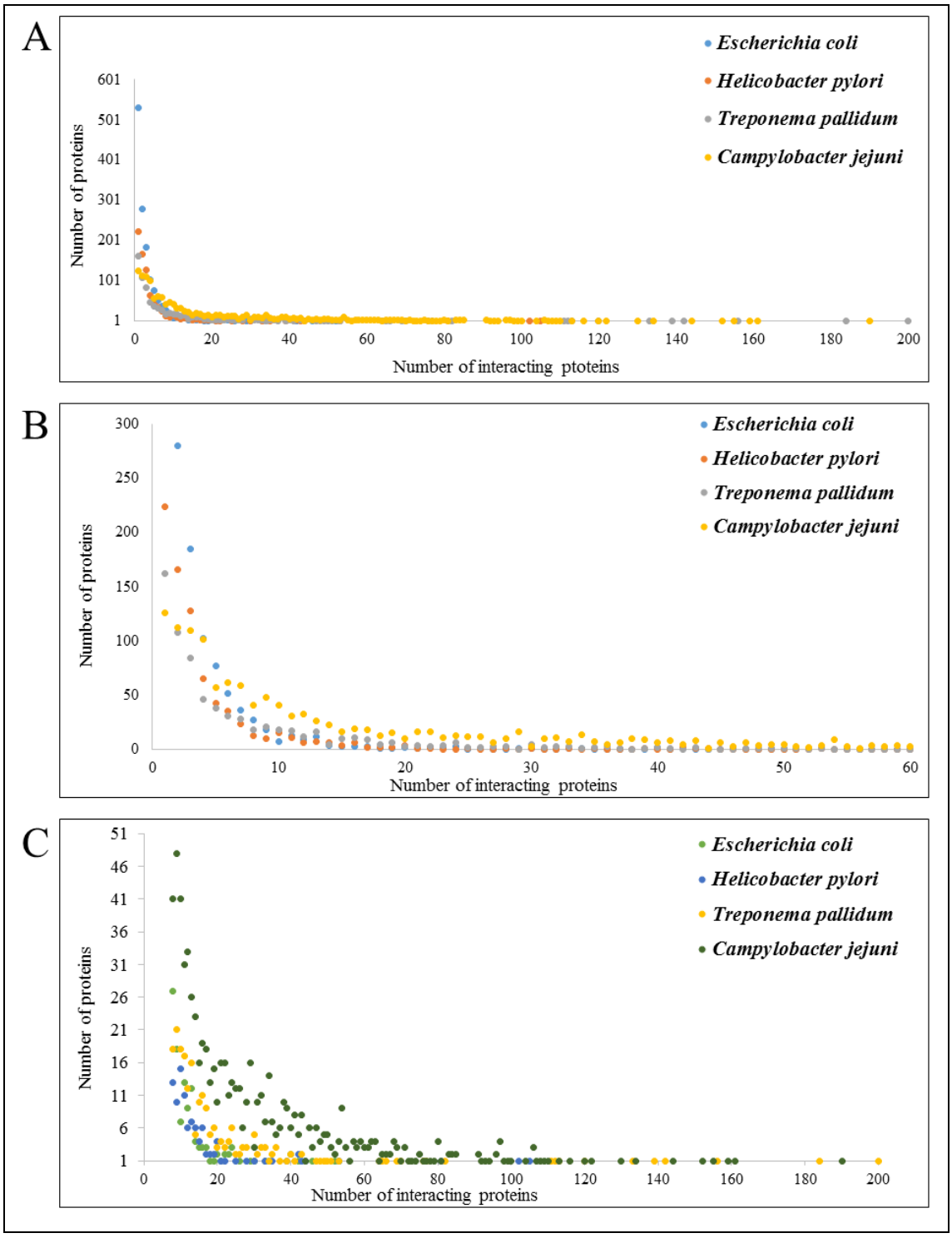


Fig. 4.1 – Node distribution for proteins across interactomes studied using yeast two hybrid system A. All proteins in the organisms B. All proteins that interact with <60 other proteins C. Only hub proteins in the organisms (hub protein has >8 interactions)

The aim of this study was to study protein – protein interactions for uncharacterized or unknown function hub proteins in *Escherichia coli* and to test them in other closely related species *Vibrio cholerae, Yersinia pestis,* and *Staphylococcus aureus*. The first objective was to get a set of high confidence interactions. The interactions that would be conserved in all of the species include in the study would be less likely to be false positives. The second objective was to draw inferences about the biological relevance of the interactions observed for hub proteins. Since *Sau* is not very closely related to *Eco*, conservation of an interaction in *Sau* would mean that the interaction is more likely to be conserved across a wide range of species having implications on the function of the protein being studied. The third objective was to get an insight into the functional role of these proteins based on the conserved interactions and to state hypothesis related to their functions.

Methods

Creating dataset. The hub proteins were selected based on the interactions observed in the binary interaction dataset for *Eco* (Rajagopala, Sikorski, Kumar, Mosca, Vlasblom, Arnold, Franca-Koh, et al., 2014). Proteins interacting with more than 7 proteins were considered to be hub proteins. Hub proteins annotated as uncharacterized were selected for further analysis. Homologs for these interacting protein pairs were identified in *Ype, Vch* and *Sau* using NCBI BLAST analysis.

Construction of bait and prey clones. See Appendix A.

Yeast transformation. See Appendix A.

Bait auto activation test. See Appendix A.

Yeast two hybrid assay. See Appendix A.

Results

Dataset Analysis. Approximately 1271 proteins are involved in protein-protein interactions (yeast two hybrid assay) in *Escherichia coli.* Among these proteins 112 proteins (9%) were detected to be hub proteins (Fig. 4.2). In this case hub proteins are defined as proteins that interact with more than or equal to 8 proteins. Among all the hub proteins, 23 proteins were uncharacterized/putative function proteins (Fig. 4.2). The hub proteins for *Ype, Vch* and *Sau* were determined based on sequence homology to the hub proteins in *Eco*. OGs were not used in this case since most of the uncharacterized hubs in *Vch, Ype* and *Sau* were not categorized into OGs.

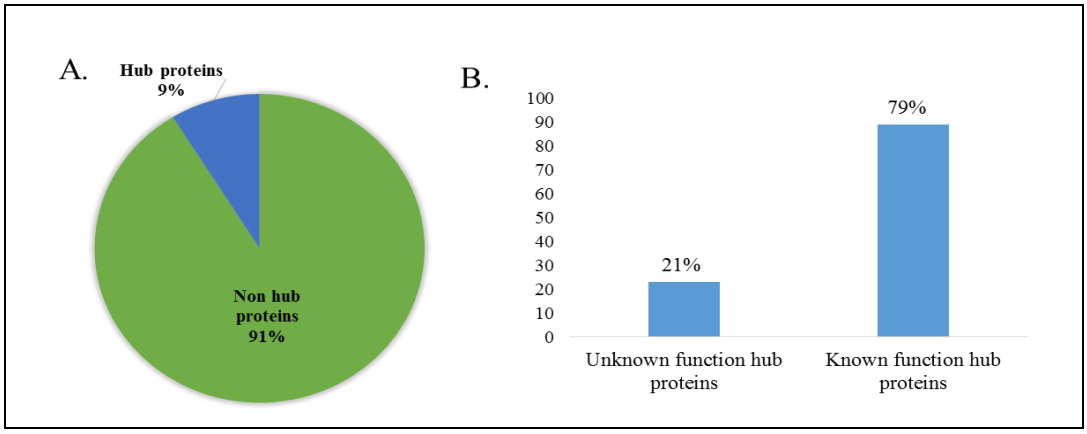


Fig. 4.2 – Hub proteins distribution in *Escherichia coli* interactome

Among the 23 hub proteins in *Escherichia coli*, only five hub proteins were conserved across all the other organisms included in this study, YjjW, YffB, YbhK, YpdC and YeeD. All proteins except YpdC are highly conserved across bacterial species. Protein YpdC is conserved in as few as ~ 300 organisms and hence was not included in this study.

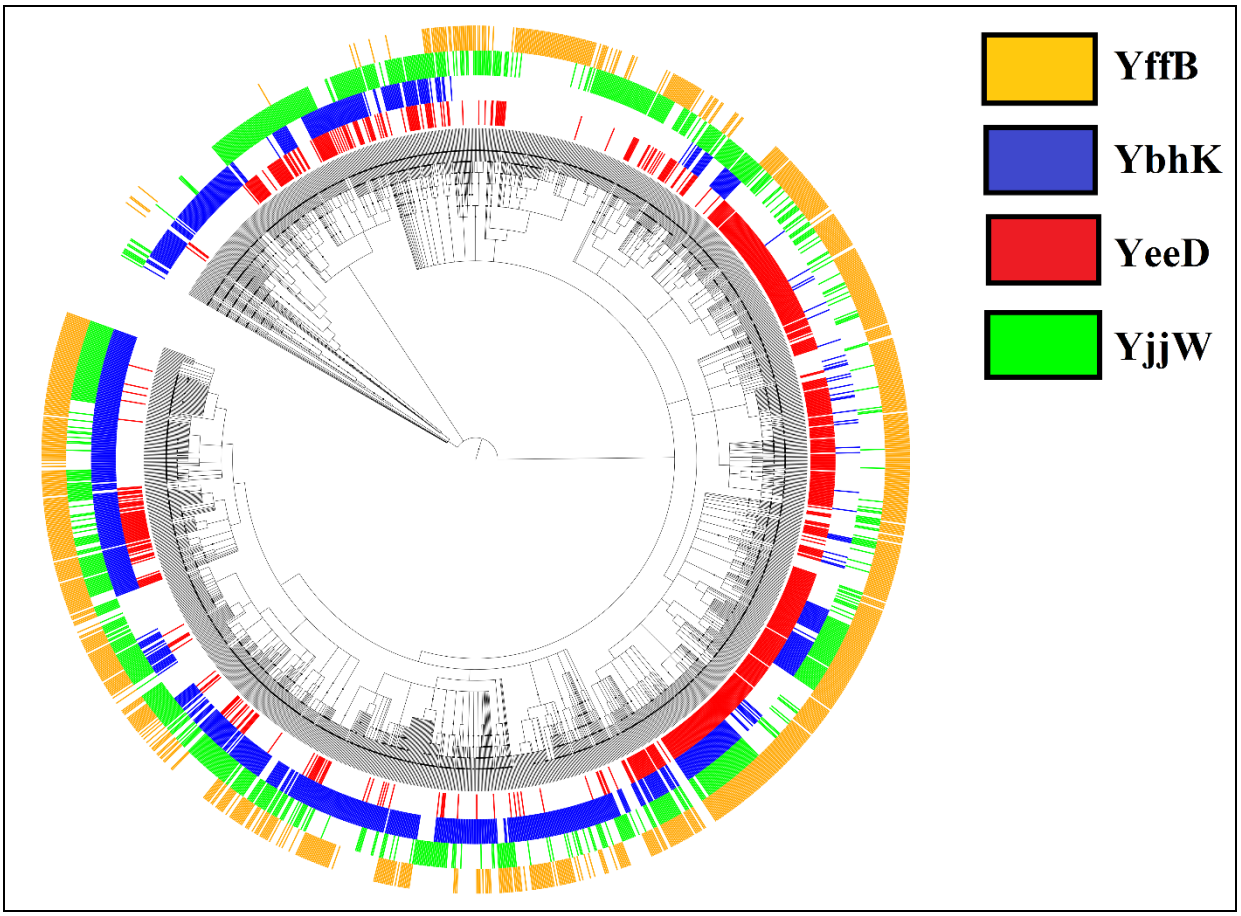


Fig. 4.3 – Phylogenetic tree showing presence of hub proteins YjjW, YeeD, YbhK and YffB in bacterial species (iTOL version 3.4)

The sequence identity was higher than 35% in all organisms for YffB and YbhK however; it seemed to have a low percent similarity for YeeD and YjjW. In all cases, the sequence coverage was greater than 80%, except for YeeD in *Ype* (Fig. 4.4).

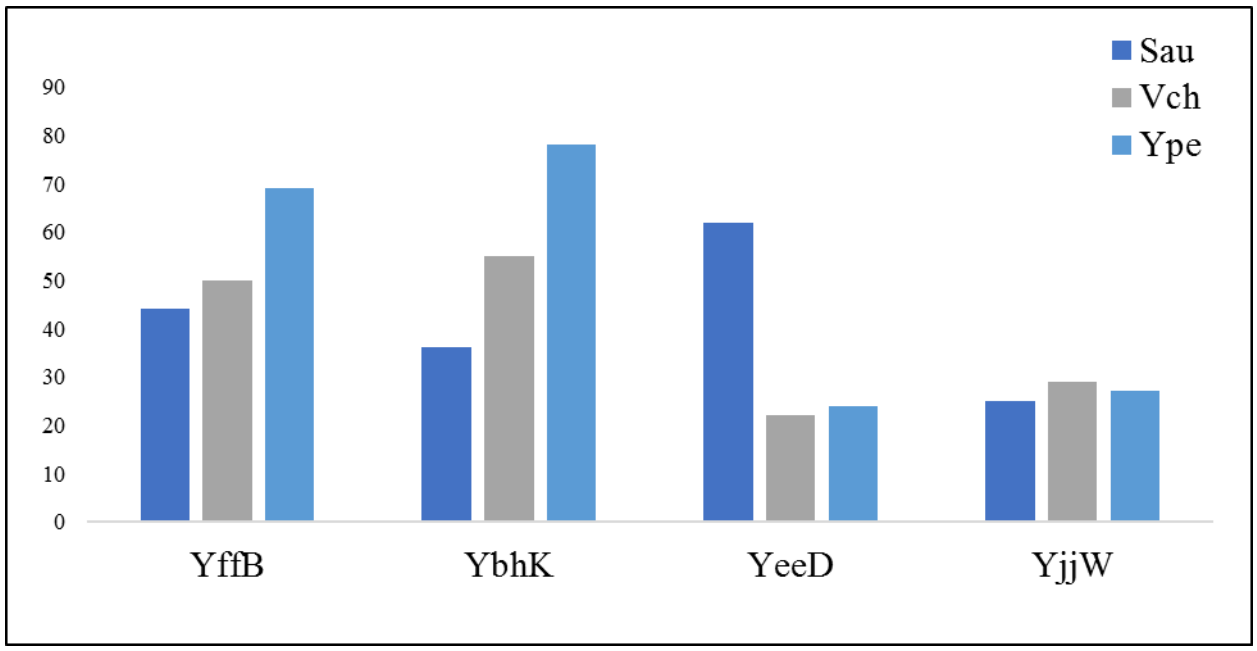


Fig. 4.4 – Hub proteins sequence identity based on NCBI BLAST analysis for *Sau, Vch,* and *Ype* in comparison with *Eco*

Homology was also used to determine presence of interacting proteins in *Ype, Vch* and *Sau*. For all the hub proteins, majority of the interactors were conserved across *Ype* and *Vch* since they are more closely related to *Eco* (Appendix B Tables 4.3, 4.4, 4.5, 4.6).

Hub protein - YeeD

This protein is annotated as UPF0033 protein YeeD in UniProt. It is composed of 75 amino acids. It is a cytoplasmic protein that has been annotated (Gene Ontology) as a binding protein and having sulfur transferase activity. YeeD belongs to pfam PF01206, a sulfurtransferase protein family. YeeD belongs to COG0425 in *Escherichia coli* which has been annotated as posttranslational modification, protein turnover, chaperones.

YeeD has two other paralogs in *Eco*, sulfur transferase TusA (81 amino acids) and unknown function YedF (77 amino acids). TusA protein interacts with IscS to activate its cysteine desulfurase activity. This interacting complex was eluted by affinity chromatography. The sulfurtransferase activity of TusA was tested by using labelled cysteine. It was observed that activated sulfur is then passed back to TusA from IscS. This sulfur is then used in the pathway for biosynthesis of 2-thiouridine. Structural analysis of IscS – TusA complex reveals the molecular basis of the sulfurtransferase activity of TusA. TusA has homologs in many bacterial species including α-, β-, and γ-proteobacteria and bacillus, as well as in archaeal species. TusA knockouts show filamentous cell shape indicating that this protein is also required for general physiology of *Escherichia coli*. (Ikeuchi, *et al.*, 2006).

Among the three paralogs, the structure of YedF and TusA in *Escherichia coli* are available. A comparison between the sequence alignments and structure alignments of TusA (PDB ID – 1DCJ) and YedF (PDB ID – 1JE3) shows that the structure is highly conserved though the sequence alignment does not show high level of similarity.

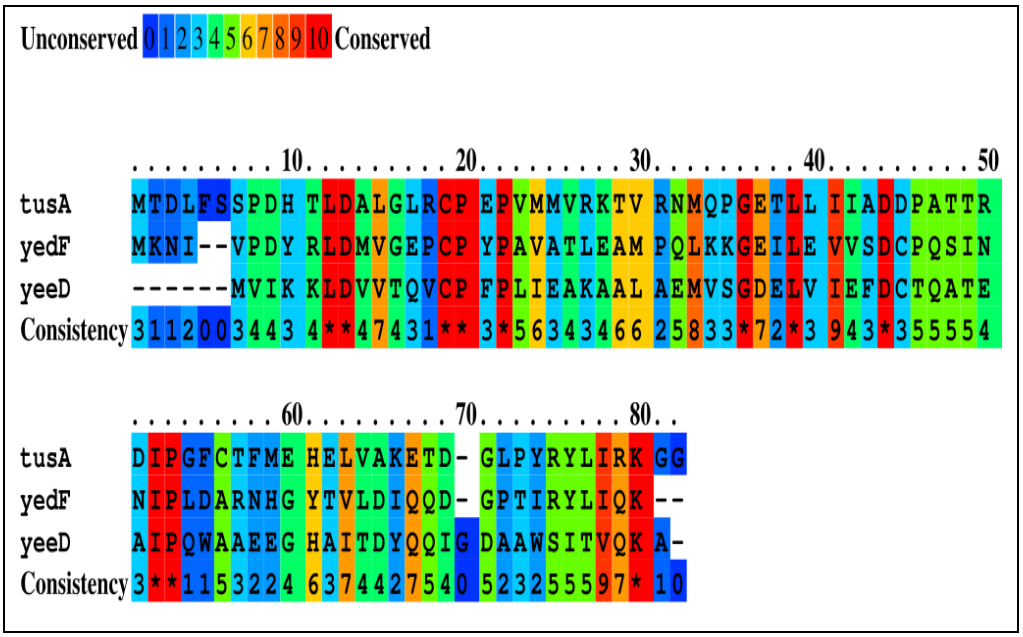


Fig. 4.5 – Protein sequence alignment for YeeD with its paralogs YedF and TusA

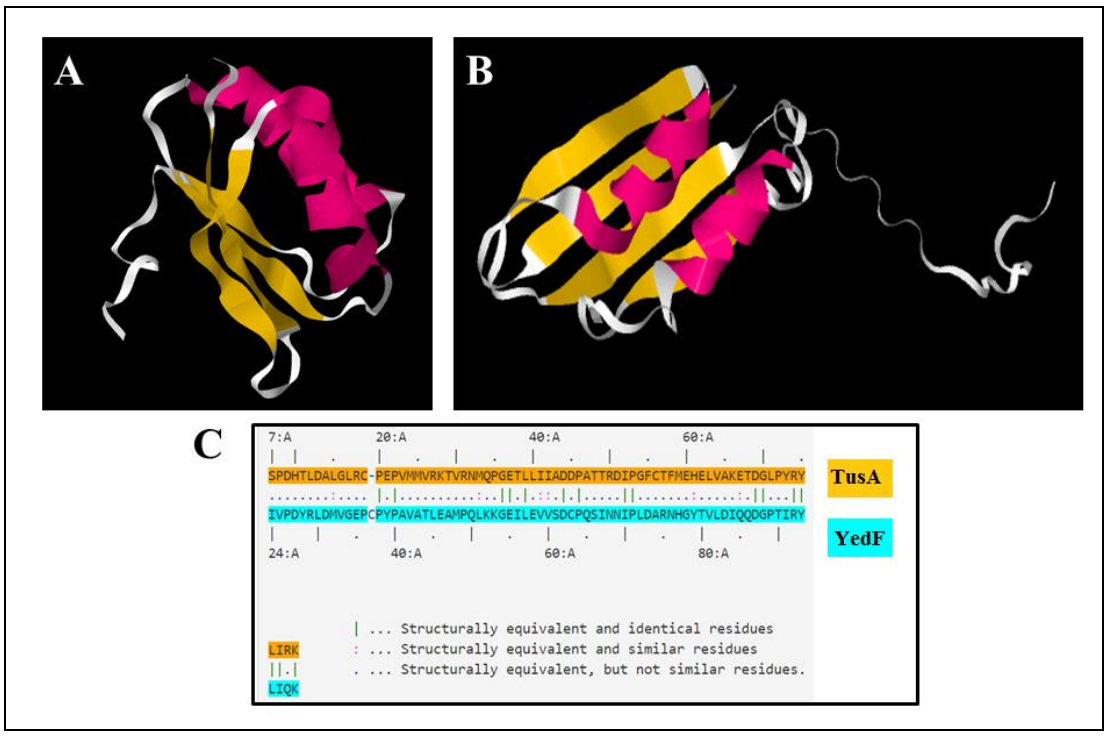


Fig. 4.6 – Structures and structure based alignment for YeeD paralogs. A. Structure of TusA B. Structure of YedF C. Structure based alignment for TusA and YedF

Conservation. Based on sequence homology, the YeeD homolog in *Sau* shows higher sequence similarity to *Eco* protein as compared to *Vch* and *Ype* (Fig. 4.4). A protein sequence alignment of YeeD homologs in *Eco, Sau, Vch* and *Ype* is shown in Fig. 4.7. Some regions of the proteins are highly conserved in all four species. These regions might be responsible for interactions with other proteins.

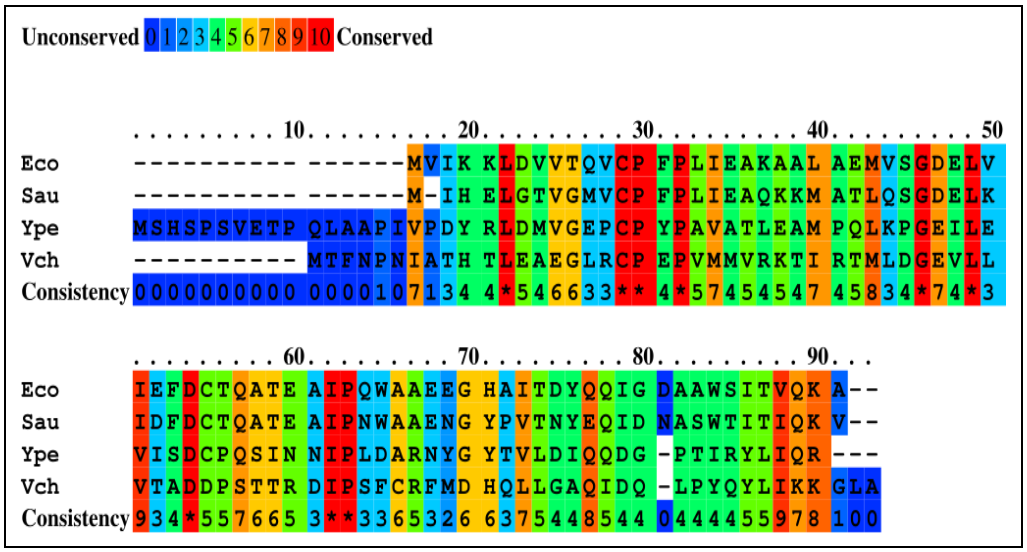
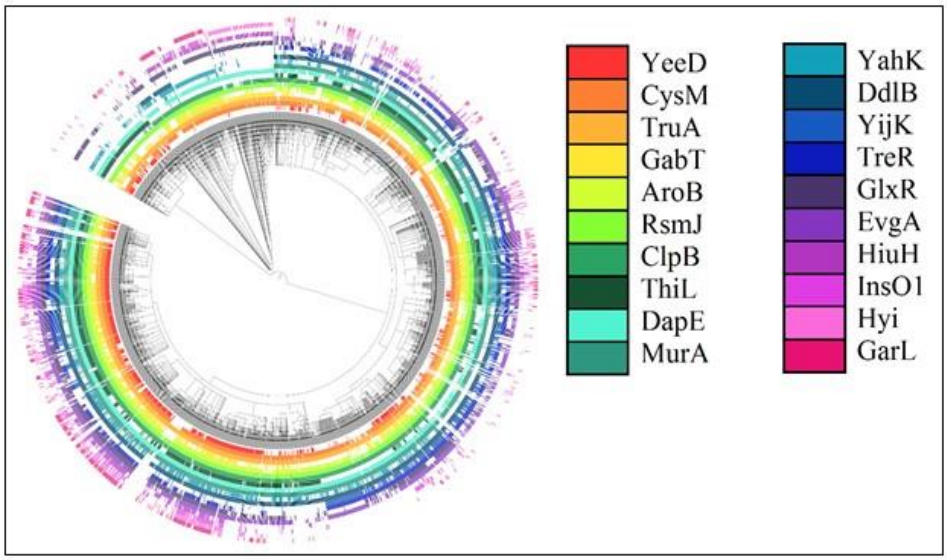


Fig. 4.7 – Protein sequence alignment for protein YeeD for the four species *Eco, Vch, Ype* and *Sau*

Phylogenetic Analysis. YeeD is conserved across more than ~1100 other species. A phylogenetic analysis of YeeD among closely related species revealed that YeeD is totally absent in *Mycobacterium subspecies* and is also absent in ~50% of the *Bacillus subspecies*. Proteins interacting with YeeD seem to be conserved across closely related organisms. This is specifically true about proteins that are conserved across all three organisms. Proteins TreR, MurB, TruA, DdlB and ClpB are conserved through all species. Protein GlxR is missing in *Staphylococcus subspecies.*

A.



B.

Protein	Number of species
AroB	674
ClpB	672
CysM	712
DapE	710
DdlB	671
EvgA	598
GabT	482
GarL	283
GlxR	562
HiuH	277
Hyi	295
InsO1	208
MurA	666
RsmJ	728
ThiL	574
TreR	480
TruA	722
YahK	434
YeeD	743
YijK	262

Fig. 4.8 – A. Phylogenetic tree representing the distribution of protein YeeD and its interacting proteins across in bacterial species B. Number of species the proteins interacting with the hub protein are present in relative to presence of YeeD represented in the phylogenetic tree

Sequence Identity Analysis for Interacting proteins. The orthologs of proteins interacting with hub protein YeeD in *Sau, Vch* and *Ype* were determined using sequence homology. It was observed that there was no correlation between sequence identity and conservation of interactions. For example, proteins DapE and AroB showed greater than 65% sequence homology for *Vch* and *Ype* however their interaction with YeeD was not conserved. On the other hand, proteins having a sequence identity of as low as 22% (DdlB in *Vch*) showed positive interactions with YeeD when screened using yeast two hybrid technique.

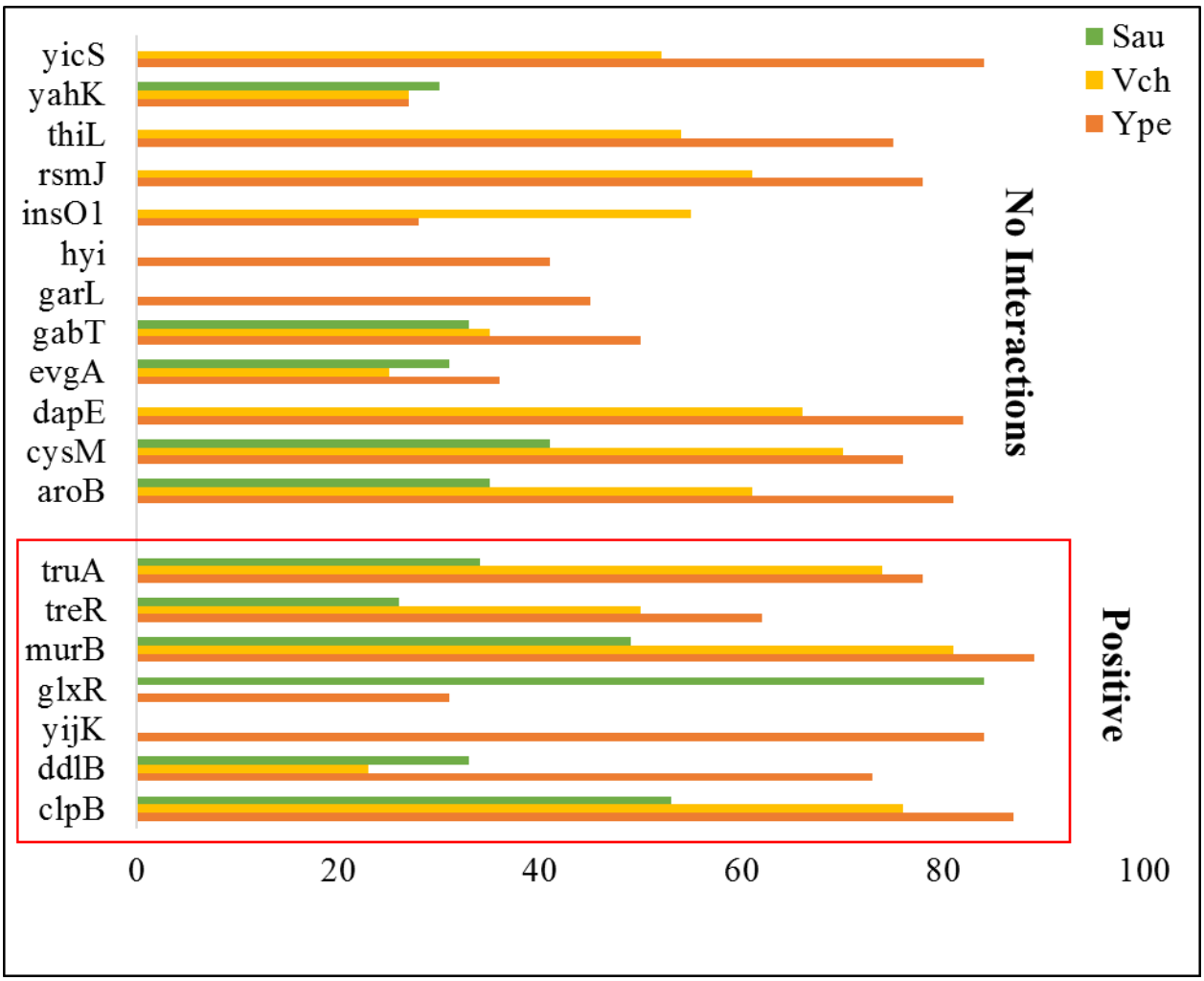


Fig. 4.9 – Sequence identity for proteins tested for interaction with hub protein YeeD in comparison with *Eco*. The proteins in the red box showed positive interactions when tested using yeast two hybrid assay in at least one of the three species used in this study

Network Analysis. YeeD has 22 interacting proteins in *Eco* and as few as 11 interacting proteins present in *Sau*. After testing these interactions individually for each organism in parallel, it was observed that ~12 interactions in *Eco* could be reproduced. Among the interactions conserved in *Eco*, 7 were observed in *Ype*, 6 in *Sau* and 5 in *Vch* (Fig. 4.11).

These conserved interactions were classified into functional categories using biological function gene ontology (GO). The proteins interacting with YeeD belonged to five GO categories, response to heat, metabolic process, unknown function, transcription, and biosynthetic processes. Majority of the interactions conserved across species involved proteins involved in transcription and biosynthetic processes.

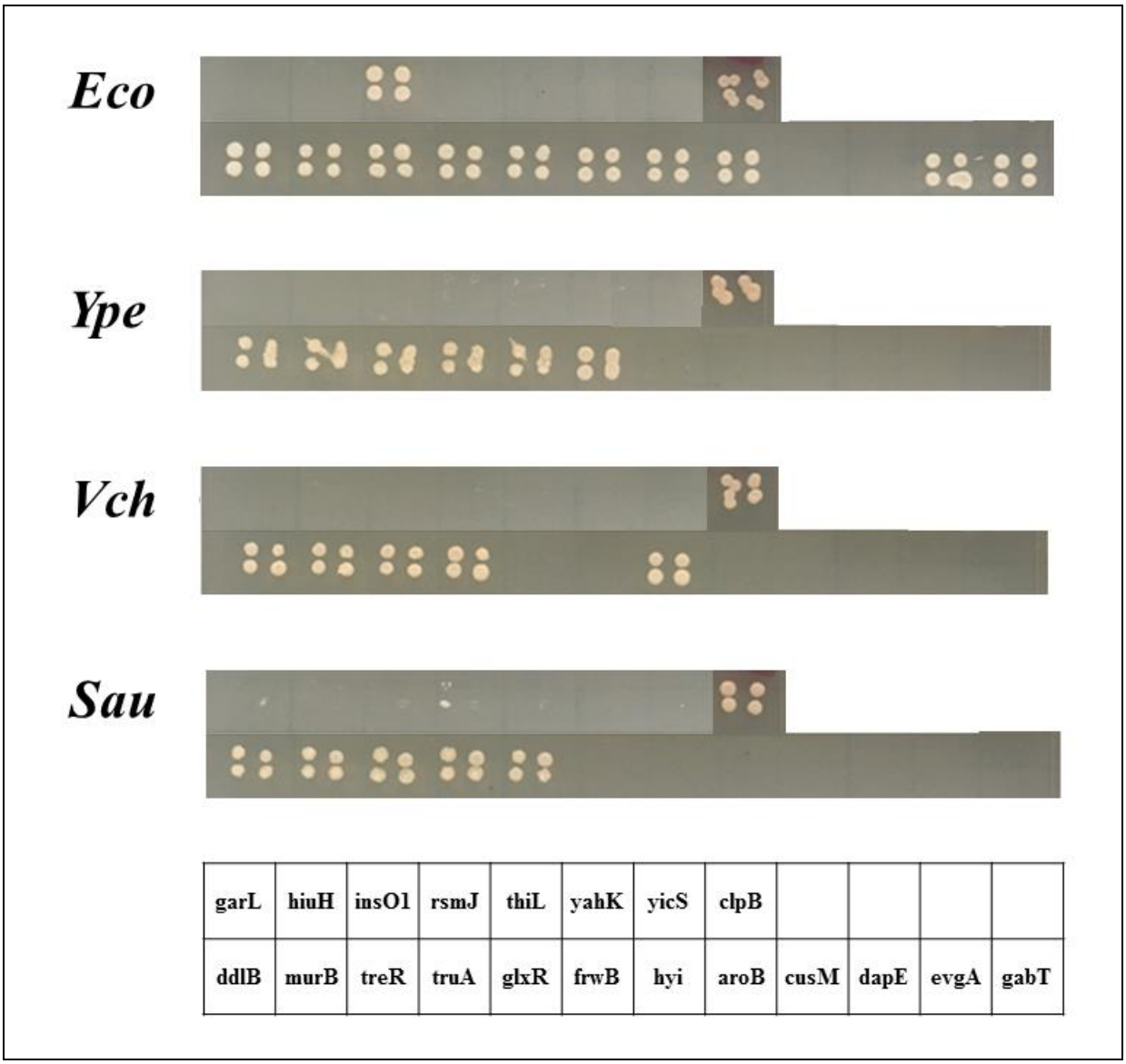


Fig. 4.10– Yeast two hybrid screen for hub protein YeeD at 10mM 3-AT

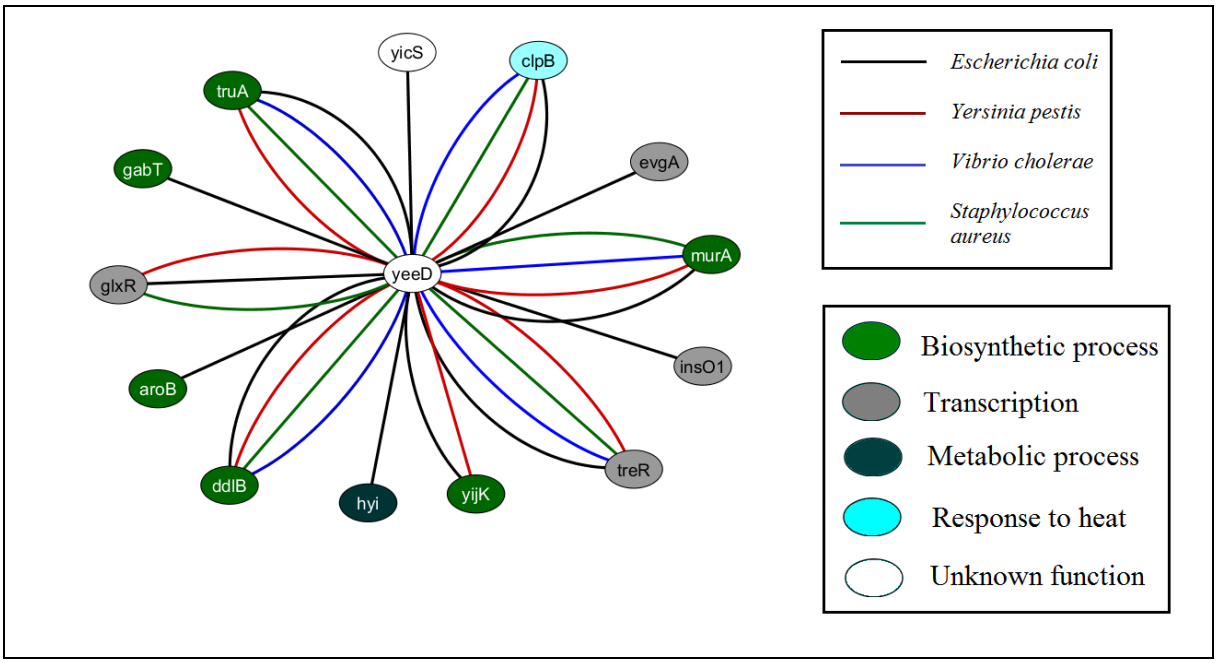


Fig. 4.11 – Protein – protein interaction network for hub protein YeeD

Hub protein – YjjW

YjjW is an unknown function protein in *Escherichia coli*. It has been recently annotated as putative glycyl-radical enzyme activating enzyme. It is a 287 amino acid long protein. YjjW has Gene Ontology (GO) function as metal binding protein. YjjW has been assigned to COG1180 which belong to NCBI functional category for proteins involved in posttranslational modification, protein turnover, chaperones. The protein has 11 sites annotated as Iron-Sulfur (metal) binding sites which are marked by presence of a cluster made of 3 cysteines and an exchangeable S-adenosyl-L-methionine.

YjjW has three other paralogs in *Escherichia coli*, PflA, PflC, and PflE. The paralogous proteins PflA and PflC are pyruvate formate-lyase activating enzymes. PflE is a putative pyruvate formate-lyase activating enzyme. All these proteins consist of a cluster made of 3 cysteines and an exchangeable S-adenosyl-L-methionine which is the site for metal binding, i.e. iron sulfur

binding. At biological function level PflA and PflC are also involved in glucose metabolic process. Knockout studies were carried out in *Escherichia coli* for proteins PflA and PflC. The mutants were cultured in media containing glucose as a carbon source. PflA mutants produced large amounts of D-lactate from glucose under the microaerobic condition. PflC codes for PFL activating enzyme II and is not a part of the PFL operon (Zhu & Shimizu, 2004). All proteins belong to Pfam PF04055 which has radical SAM (S-adenosyl-L-methionine) domain. Sequence analysis of the paralogs shows high conservation at a few locations, but overall the conservation is low (Fig. 4.12). Protein structure information is not available for any of the paralogous proteins.

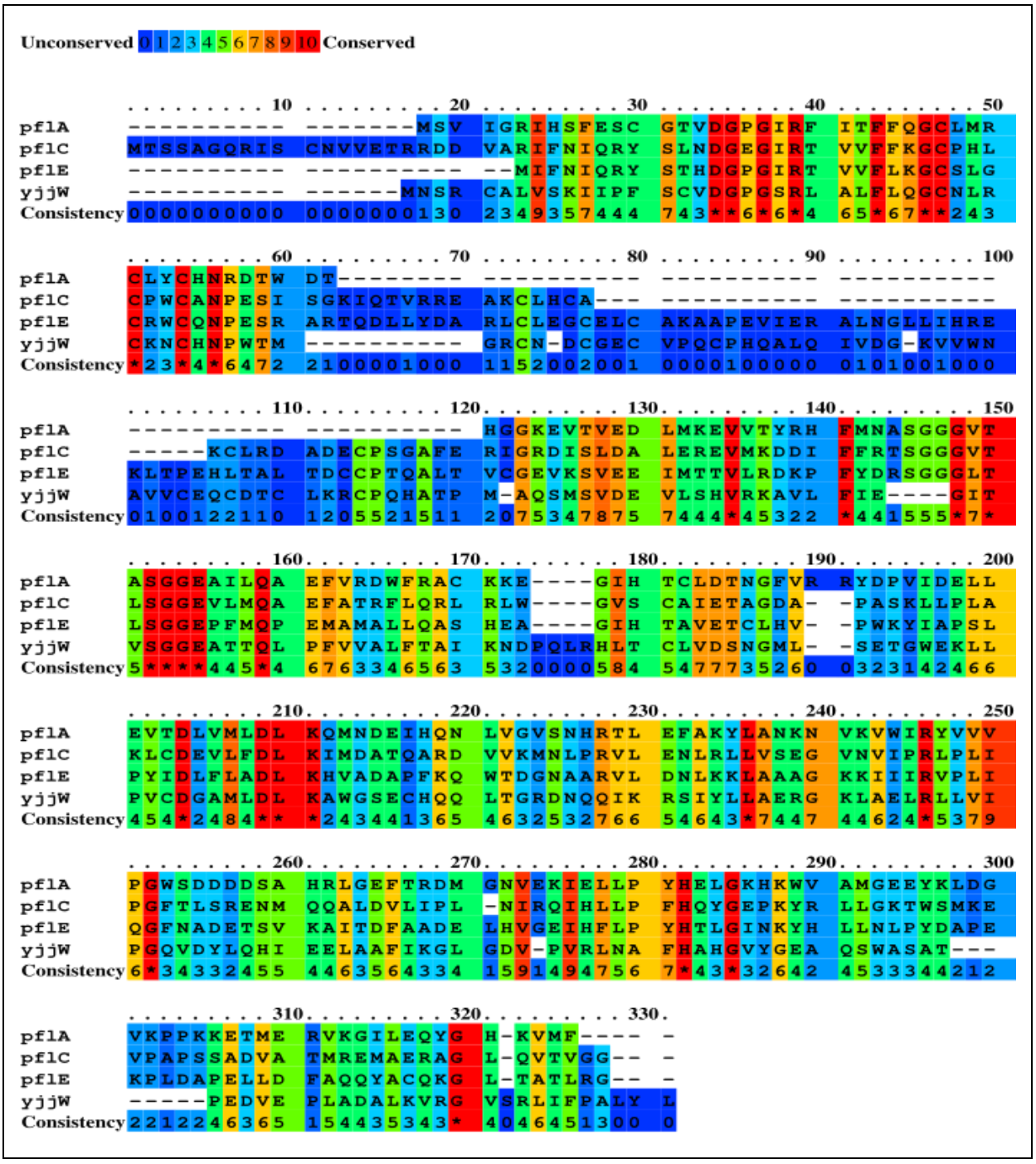


Fig. 4.12 – Protein sequence alignment for YjjW and its paralogs

Conservation. YjjW is conserved across more than ~1988 species. Protein sequence alignment of YjjW homologs in *Eco*, *Sau*, *Vch* and *Ype* is shown in Fig.4.13. Some regions of the proteins are

highly conserved in all four species. These regions might be responsible for interactions with other proteins.

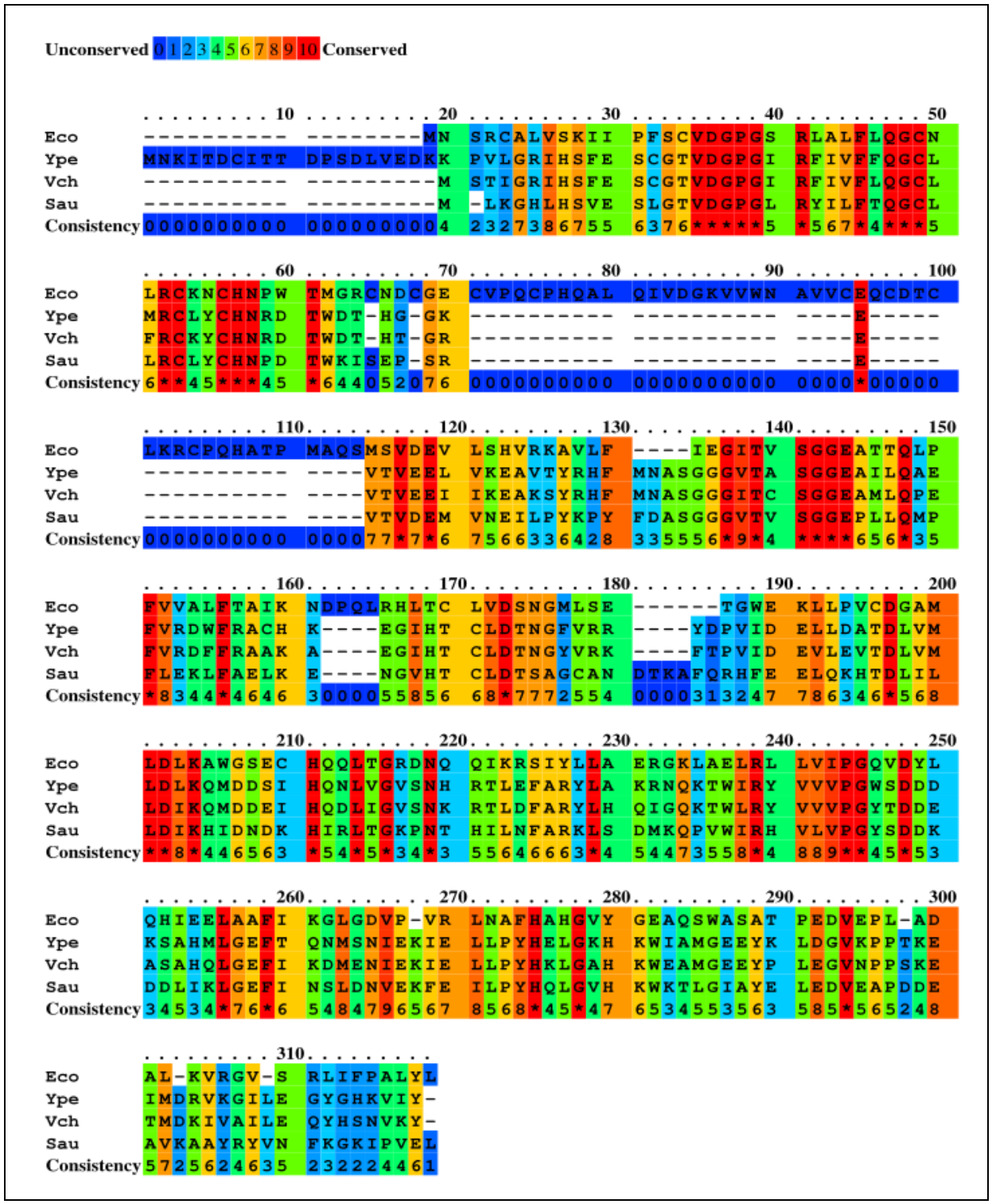
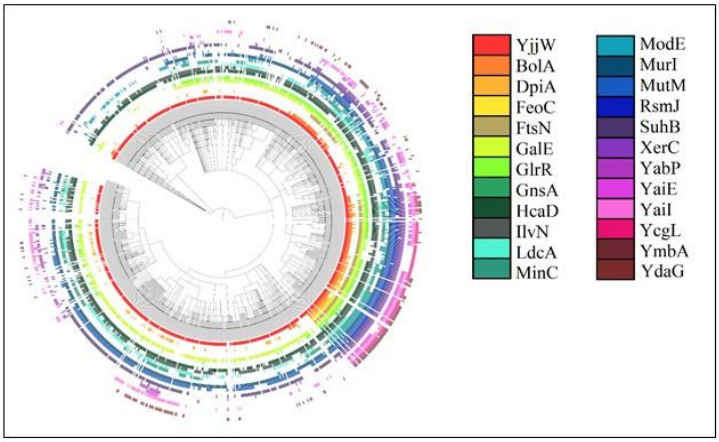


Fig. 4.13 – Protein sequence alignment for YjjW for the four species *Eco, Ype, Vch* and *Sau*

Phylogenetic Analysis. Phylogenetic analysis of YjjW among closely related species revealed that YjjW is present in most bacterial species, i.e. it is highly conserved (Fig. 4.14). However, most of the interactors in *Escherichia coli* dataset do not show very high conservation. There does not seem to be a relation between conservation of the protein and conservation of the interaction across the four species.

A.



B.

Protein	Number of Species
YjjW	961
BolA	238
DpiA	1385
feoC	47
FtsN	425
GalE	759
GlrR	1395
GnsA	12
HcaD	1295
IlvN	777
LdcA	1997
MinC	408
ModE	1066
MurI	775
MutM	1608
RsmJ	172
SuhB	1429
XerC	275
YabP	1090
YaiE	187
YaiI	662
YcgL	147
YmbA	615
YdaG	111

Fig. 4.14 – A. Phylogenetic tree representing the distribution of protein YjjW and its interacting proteins in bacterial species B. Number of species the proteins interacting with the hub protein are present in relative to presence of YjjW represented in the phylogenetic tree

Sequence Identity Analysis for Interacting proteins. The orthologs of proteins interacting with hub protein YjjW in *Sau, Vch* and *Ype* were determined using sequence homology. There was no correlation between sequence identity and conservation of interactions. For example, proteins RsmJ and SuhB showed greater than 65% sequence homology for *Vch* and *Ype* however their interaction with YjjW was not conserved. On the other hand, proteins having a low sequence showed positive interactions with YjjW when screened using yeast two hybrid technique. No correlation was observed with protein conservation across species, i.e. even though proteins MurI, FeoC, YaiI and DpiA were conserved across all four species their interaction was not conserved.

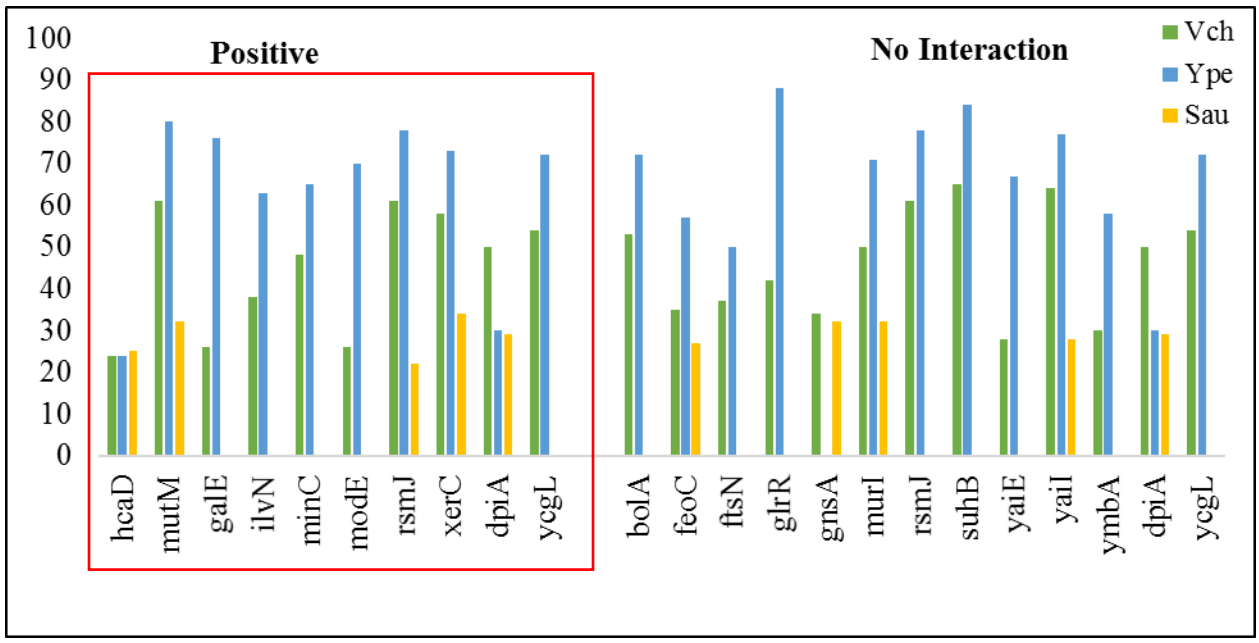
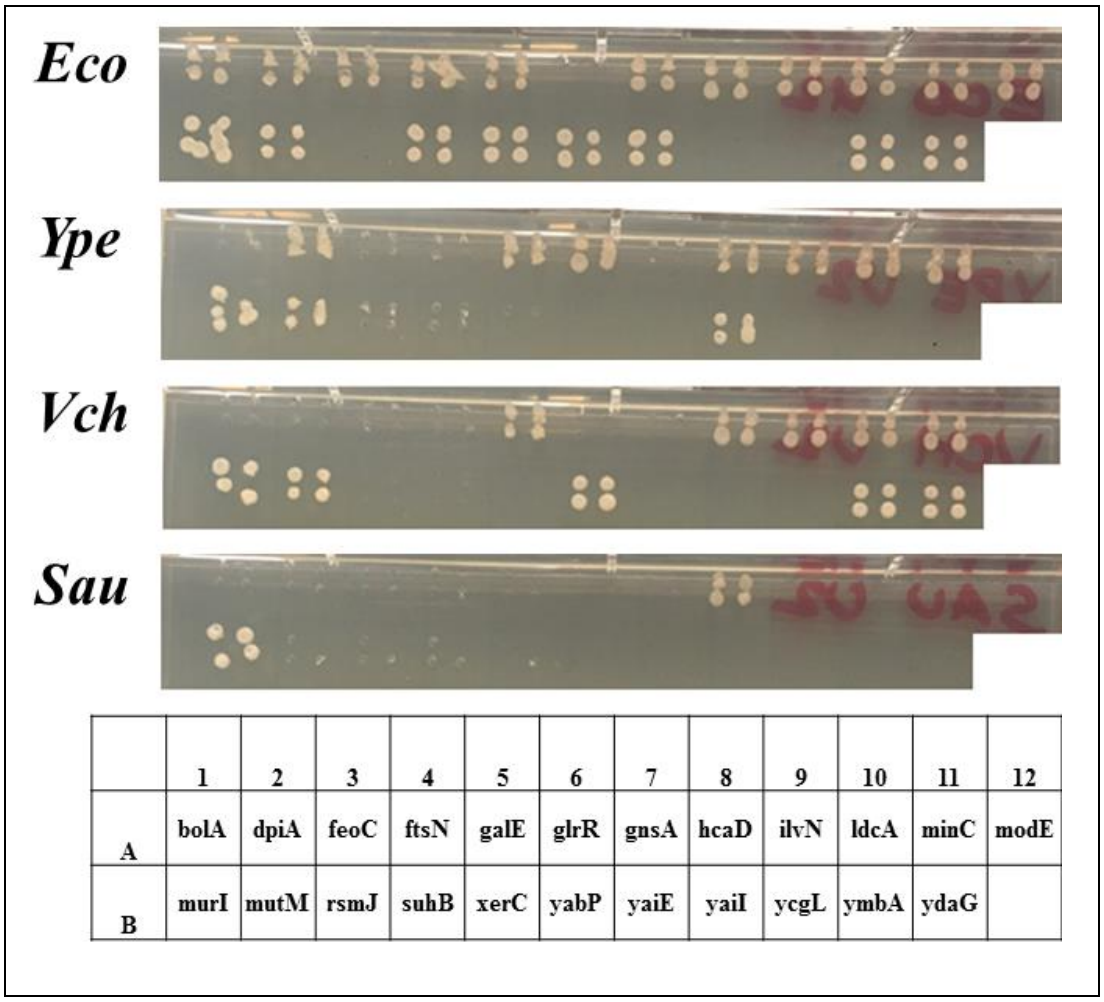


Fig. 4.15 – Sequence identity for proteins tested for interaction with hub protein YjjW in comparison with *Eco*. The proteins in the red box showed positive interactions when tested using yeast two hybrid assay in at least one of the three species used in this study

Network Analysis. YjjW has 23 interacting proteins in *Eco* and as few as 10 interacting proteins present in *Sau*. After testing these interactions individually for each organism in parallel, it was observed that ~19 interactions in *Eco* could be reproduced. Among the interactions conserved in *Eco*, 9 were observed in *Ype*, 8 in *Vch* and 2 in *Sau* (Fig. 4.17). These conserved interactions were classified into functional categories using biological function gene ontology (GO). The proteins interacting with YjjW belonged to four GO categories, ion binding, cell division, unknown function, and transcription. Majority of the interactions conserved across species involved proteins annotated as being related to ion binding and cell division. YjjW has metal binding iron-sulfur cluster sequences and hence interactions with ion binding proteins is expected. This iron-sulfur cluster activity might be required during certain stages of cell division and hence this protein must be interacting with cell division proteins as well.



	1	2	3	4	5	6	7	8	9	10	11	12
A	bolA	dpiA	feoC	ftsN	galE	glrR	gnsA	hcaD	ilvN	ldcA	minC	modE
B	murI	mutM	rsmJ	suhB	xerC	yabP	yaiE	yaiI	ycgL	ymbA	ydaG	

Fig. 4. 16 – Yeast two hybrid screen for hub protein YjjW at 10mM 3-AT

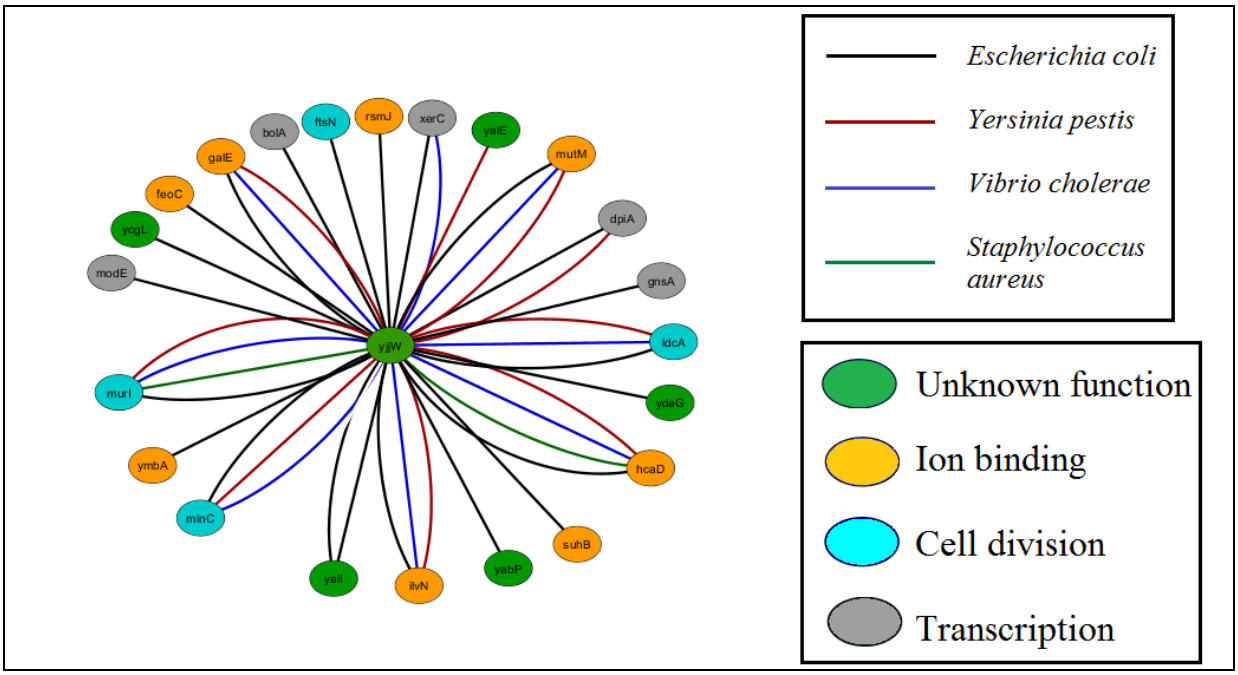


Fig. 4.17 – Protein – protein interaction network for hub protein YjjW

Hub protein – YbhK

YbhK is an unknown function protein in *Escherichia coli* annotated as putative gluconeogenesis factor. It is a 302 amino acid long cytoplasmic protein. YbhK belongs to pFam PF01933, uncharacterized protein family UPF0052. It belongs to COG0391 which has been annotated as having a function in coenzyme/carbohydrate transport and metabolism.

YbhK doesn't have paralogs. However, it has orthologs in *Bacillus subtilis* (YvcK) and *Mycobacterium smegmatis* (Rv1422) which have been studied extensively experimentally. YvcK has been studied for over a decade in *Bacillus subtilis.* YvcK knockout has been studied using different carbon sources in media. It was observed that YvcK mutants grew slow in some media and fast in some media condition. Based on this analysis, the authors concluded that YvcK is essential for carbon metabolism probably in gluconeogenesis required for synthesis of cell wall precursor molecules. Also, the cells lacking YvcK showed filamentous or L shaped cell shape.

(Gorke, Foulquier, & Galinier, 2005). With advance in technology over the last decade, the protein was studied using GFP tagging as well as in vivo co-localization but molecular function of the protein could not be determined (Foulquier et al., 2014). Rv1422 was studied in *Mycobacterium smegmatis.* Phenotypes similar to YvcK were observed, i.e. growth defects were observed under certain media conditions for Rv1422 knockouts. Also, the knockouts showed cell wall defects like shortening of the cell and bulging phenotypes. However, in this case as well, the authors were not able to pin point the molecular mechanism for the protein function (Mir *et al.*, 2014). These two studies might be the responsible for the current annotation of the protein in *Escherichia coli* as putative gluconeogenesis factor as well as the Biological function GO term annotation as 'regulation of cell shape'.

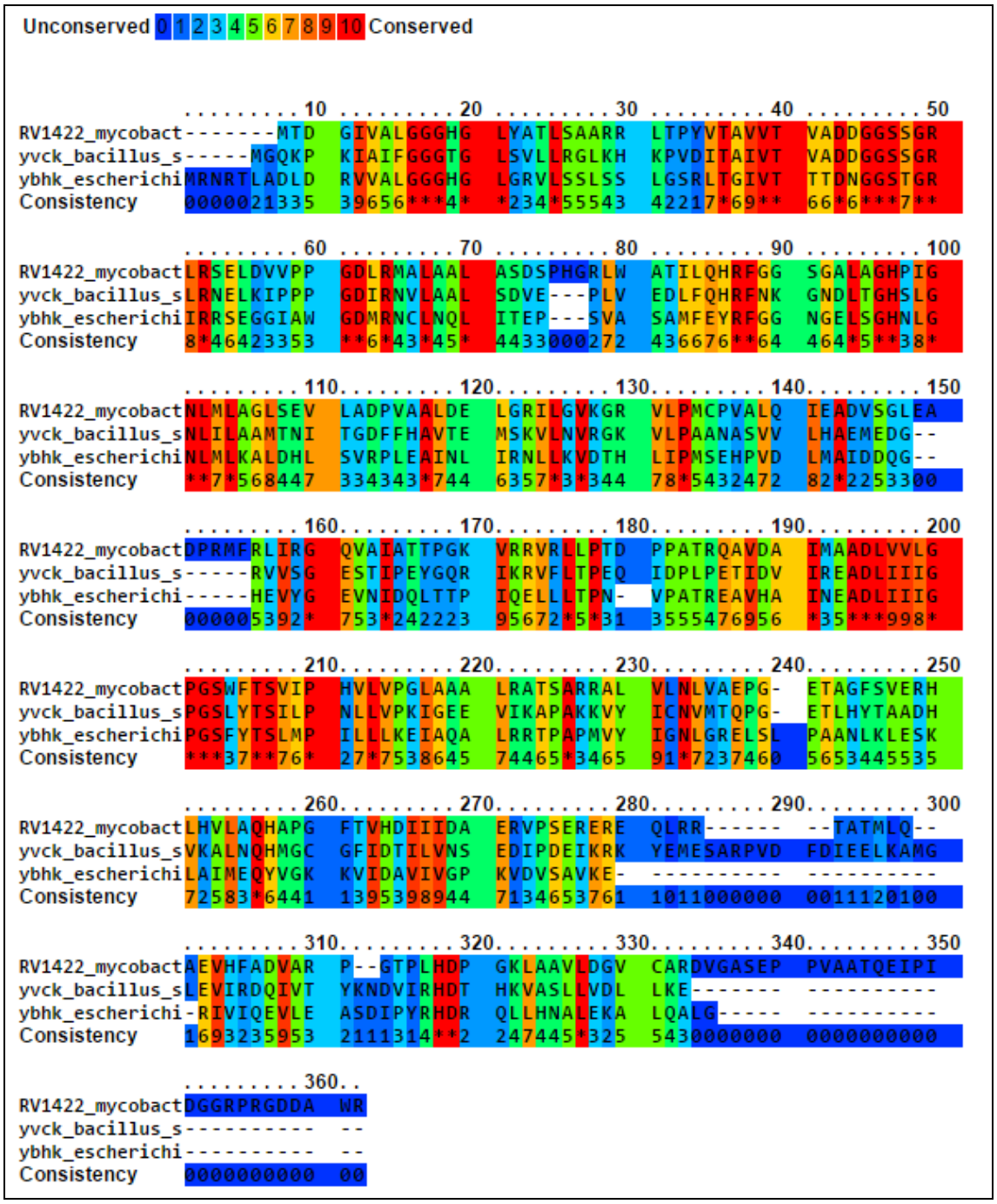


Fig. 4.18 - Protein sequence alignment for homologs of YbhK in *Bacillus subtilis* and *Mycobacterium smegmatis*

Conservation. YbhK is conserved across more than ~1121 species. Protein sequence alignment of YbhK homologs in *Eco, Sau, Vch* and *Ype* is shown in Fig. 4.19. Some regions of the proteins are highly conserved in all four species and overall the proteins seem to be well conserved across the four species.

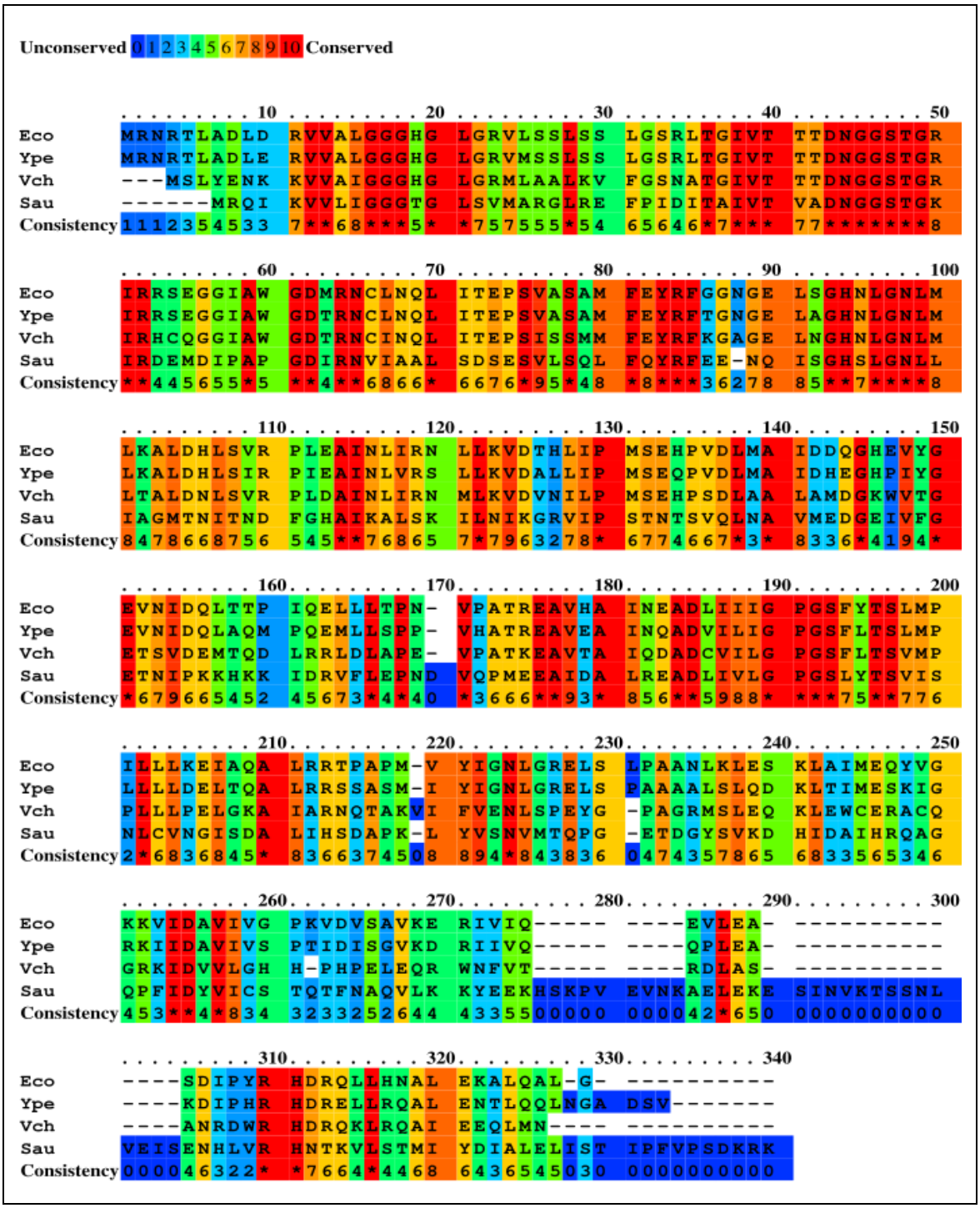
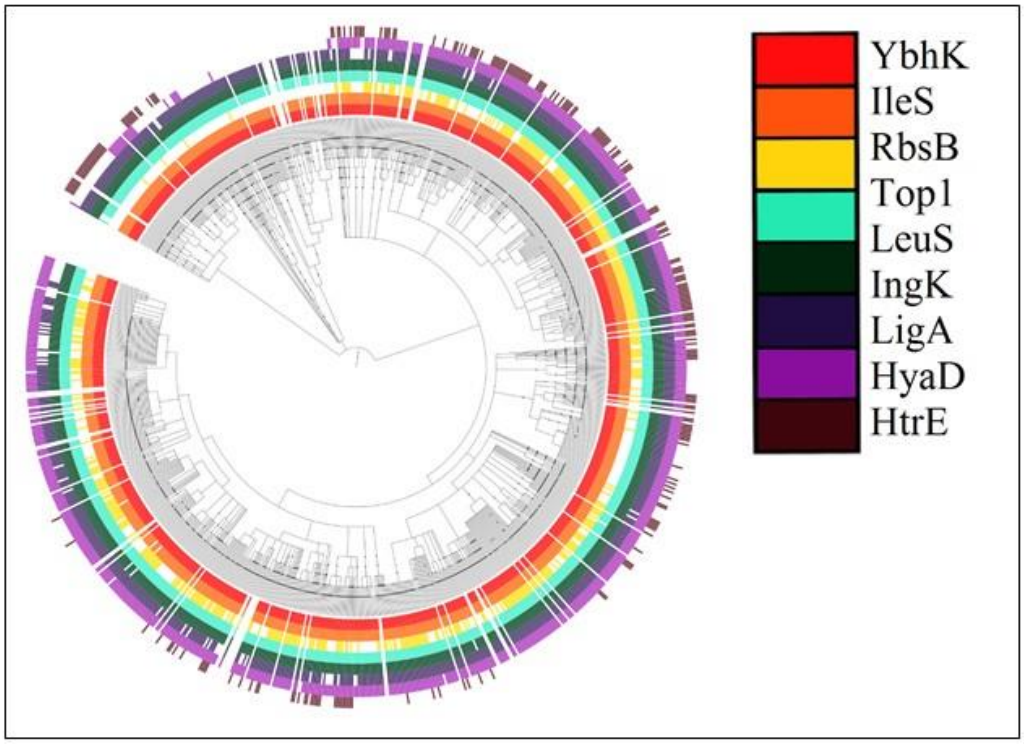


Fig. 4.19 – Protein sequence alignment for YbhK for the four species *Eco, Ype, Vch* and *Sau*

Phylogenetic Analysis. Phylogenetic analysis of YbhK various species revealed that YbhK is present across numerous other bacterial species (Fig. 4.20). Most of the interacting proteins are also widely conserved across the phylogenetic tree. Conservation of protein seems to have a direct relation to conservation across different species in this case. Protein RbsB is not conserved across a few species and hence its interaction does not seem to be conserved across any of the species in this case.

A.



B.

Proteins	Number of species
YbhK	955
IleS	952
RbsB	490
Top1	949
LeuS	948
IngK	878
LigA	817
HyaD	259
HtrE	92

Fig. 4.20 – A. Phylogenetic tree representing the distribution of protein YbhK and its interacting proteins in bacterial species B. Number of species the proteins interacting with the hub protein are present in relative to presence of YbhK represented in the phylogenetic tree

Sequence Identity Analysis for Interacting proteins. The orthologs of proteins interacting with hub protein YbhK in *Sau*, *Vch* and *Ype* were determined using sequence homology (Fig. 4.21). There seems to be a slight correlation between sequence identity and conservation of interaction in this case. All the high identity proteins show an interaction with YbhK in all four species. In this case, HyaD id present only in *Sau* and unknown function protein YjgL is not present in any of the three species.

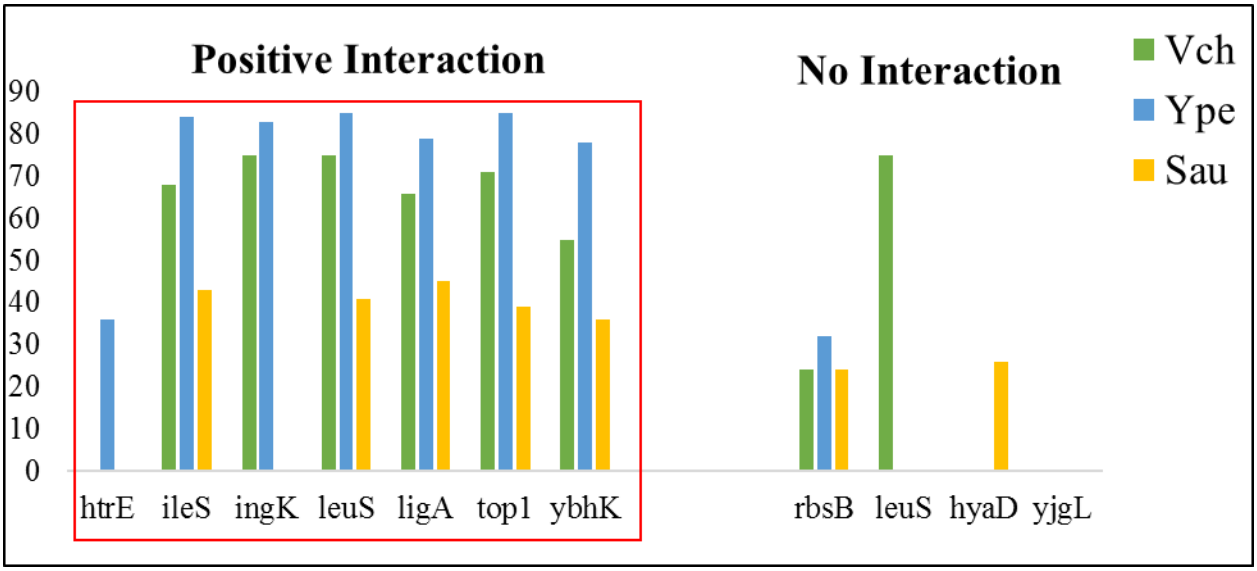


Fig. 4.21 – Sequence identity for proteins tested for interaction with hub protein YbhK in comparison with *Eco*. The proteins in the red box showed positive interactions when tested using yeast two hybrid assay in at least one of the three species used in this study

Network Analysis. Based on literature, YbhK shows eight interactions including self-interaction. However, the self-interaction was not conserved in the retests. Among the seven interactions observed in *Eco*, 4 interactions were conserved in *Ype* and 3 were observed across all four species. The interactions were conserved with proteins involved in DNA binding and ion binding. Based on orthologs studied in *Bacillus subtilis* and *Mycobacterium tuberculosis* as well as UniProt annotation, this protein has functions in glucose metabolism and phosphotransferase system.

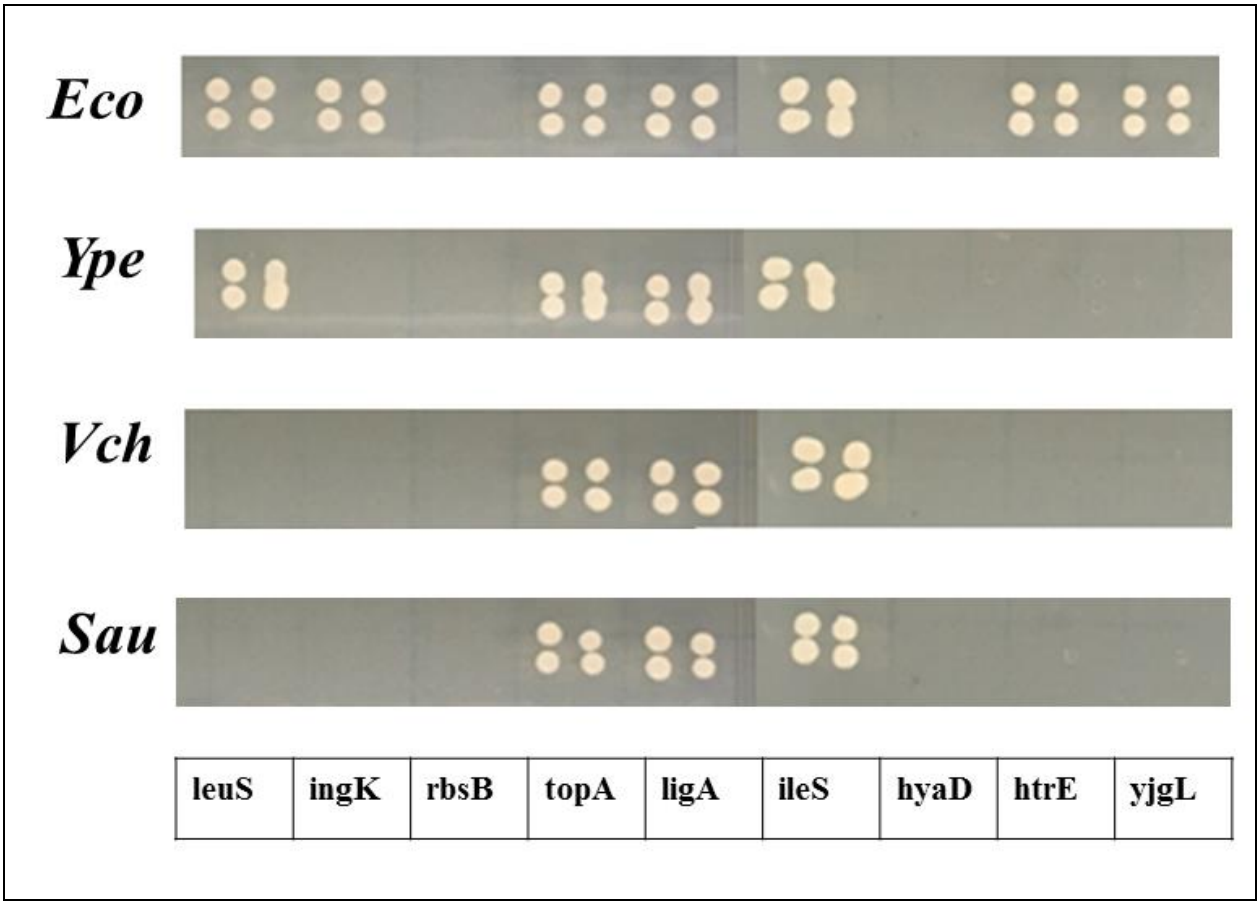


Fig. 4.22 – Yeast two hybrid screen for hub protein YbhK at 10mM 3-AT

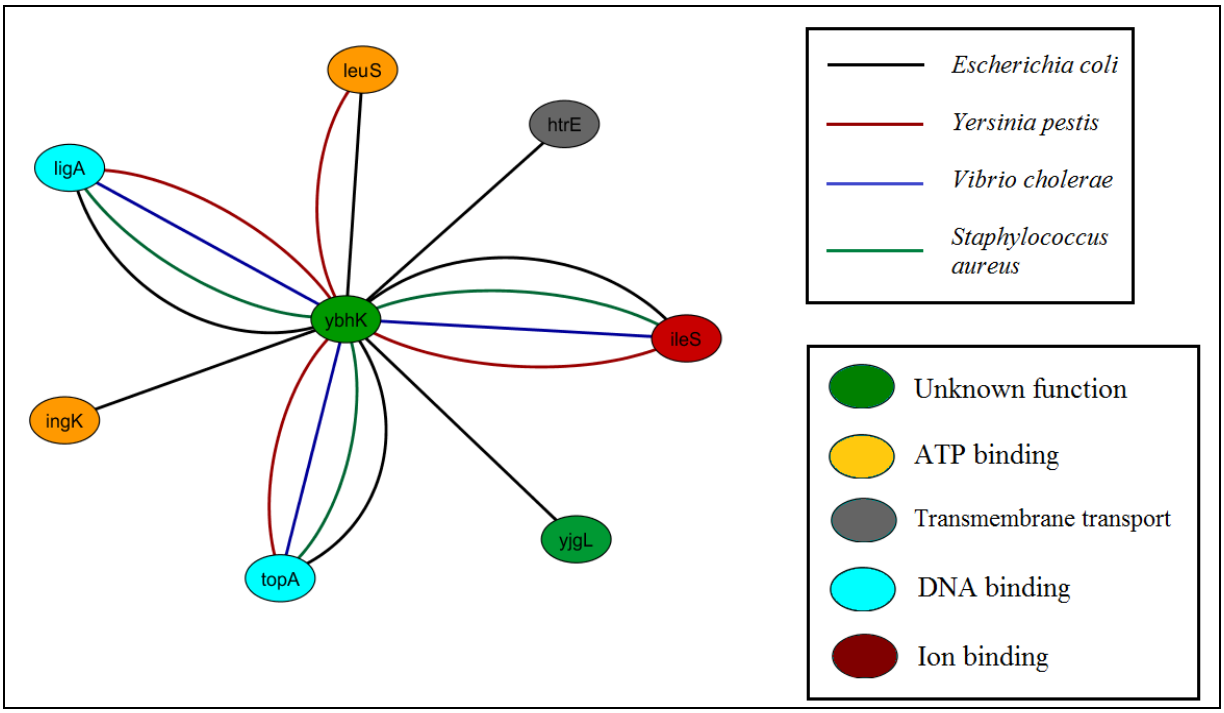


Fig. 4.23 – Protein – protein interaction network for hub protein YbhK

Hub protein – YffB

YffB is an unknown function protein in *Escherichia coli*. It is a 118 amino acid long protein present in the cytosolic region of the cell. YbhK belongs to pFam PF03960, a family for ars operon. The ars operon is present in some bacterial taxa and the major function of this operon is detoxification of arsenate, arsenite, and antimonite. The operon transports arsenite and antimonite outside of the cell. YffB does not have a biological function or molecular function GO annotation. However, based on the COG that YffB has been categorized into, it should have a function in inorganic ion transport and metabolism.

YffB has three other paralogs, ArsC, YfgD, and YfjU. ArsC is the only functionally annotated protein among all the paralogs and has been annotated based on sequence homology as arsenate reductase. ArsC is a 141 amino acid long protein. YfjU protein has been annotated as putative

arsenate reductase protein and is a 104 amino acid long protein. YfgD and YffB have been annotated as unknown function proteins. Sequence alignment of YffB paralogs show that the sequence is not very well conserved across the paralogs (Fig. 4.24). The conservation of amino acids across the four paralogs is consistently very low across the length of the alignment. Proteins structure has been studied for YffB and its paralog ArsC in *Escherichia coli*. Structure alignment of the two proteins reveals ~15% identity and ~30% similarity (Fig. 4.25).

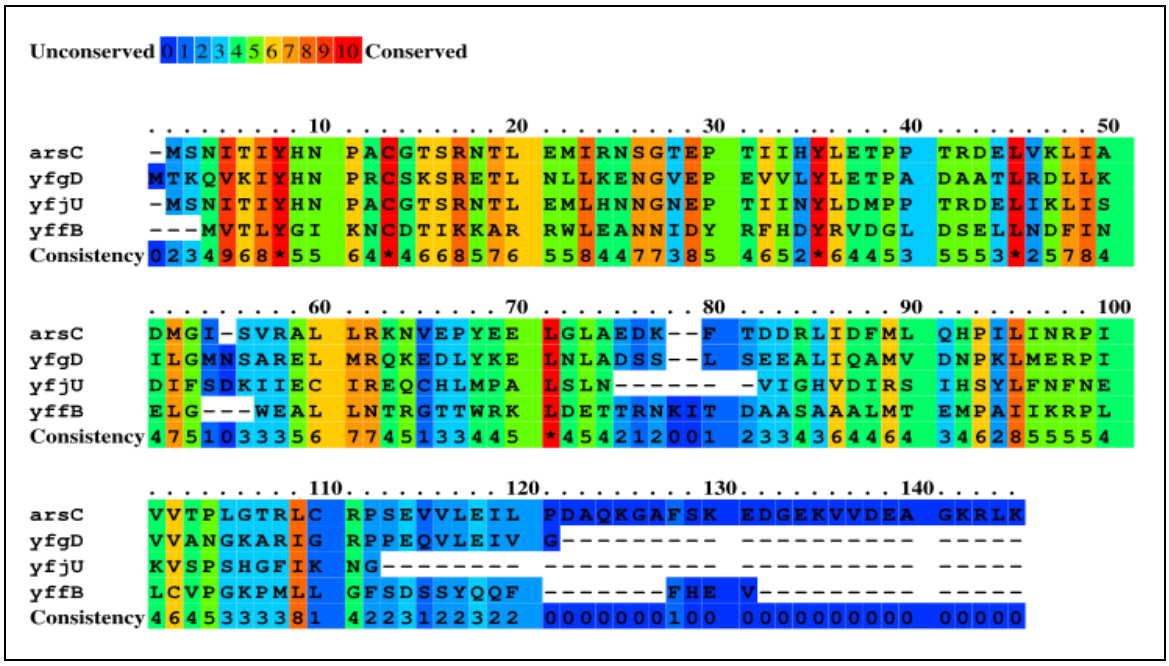


Fig. 4.24 – Protein sequence alignment for paralogs of YffB and its paralogs

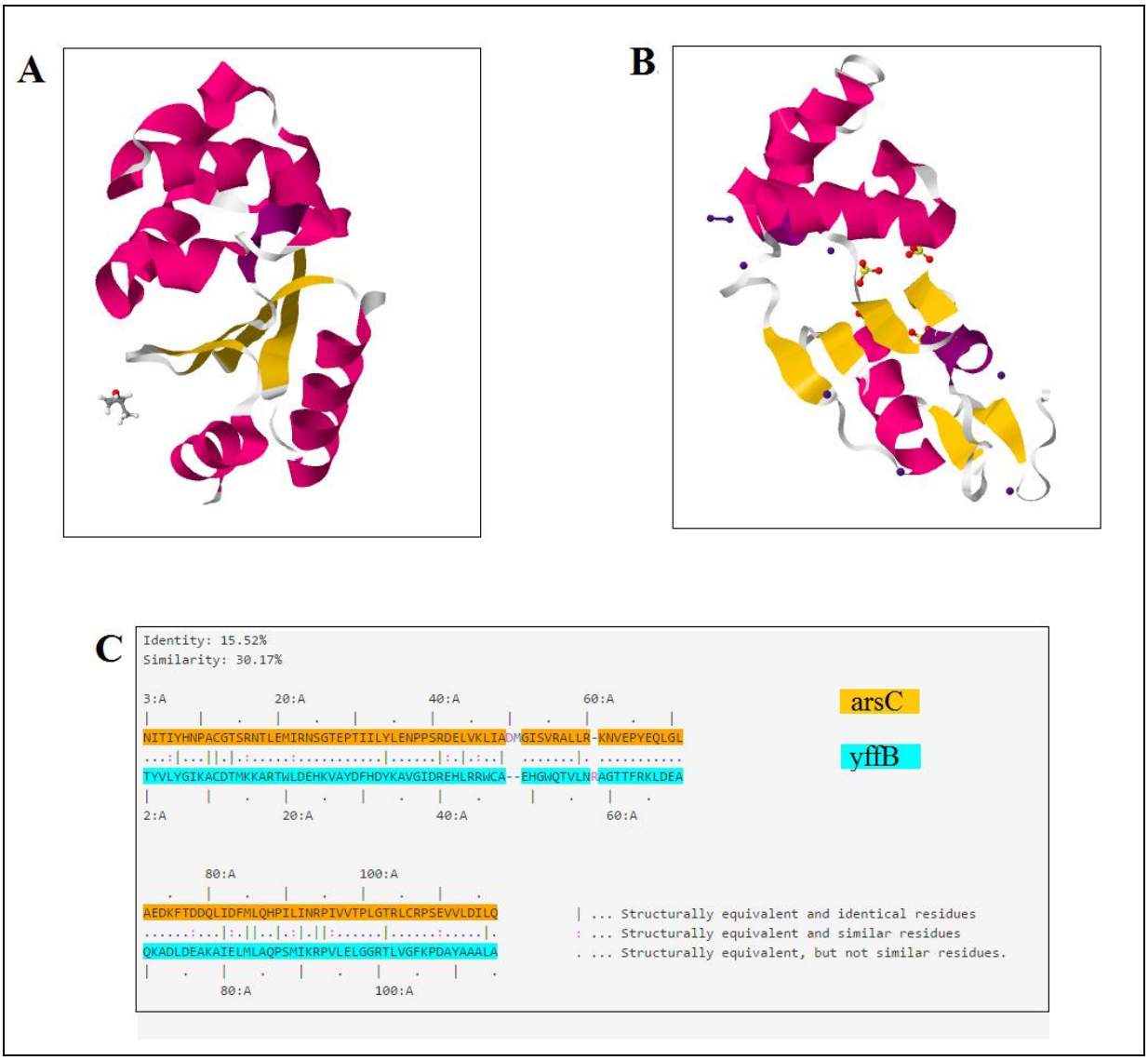


Fig. 4.25 – Structures and structure based alignment for YffB and it paralog ArsC. A. Structure of YffB B. Structure of ArsC C. Structure based alignment for YffB and ArsC

Conservation. YffB is conserved across more than ~2164 species. Protein sequence alignment of YffB homologs in *Eco*, *Sau*, *Vch* and *Ype* is shown in Fig. 4.26. The proteins sequences are overall well conserved across the four species.

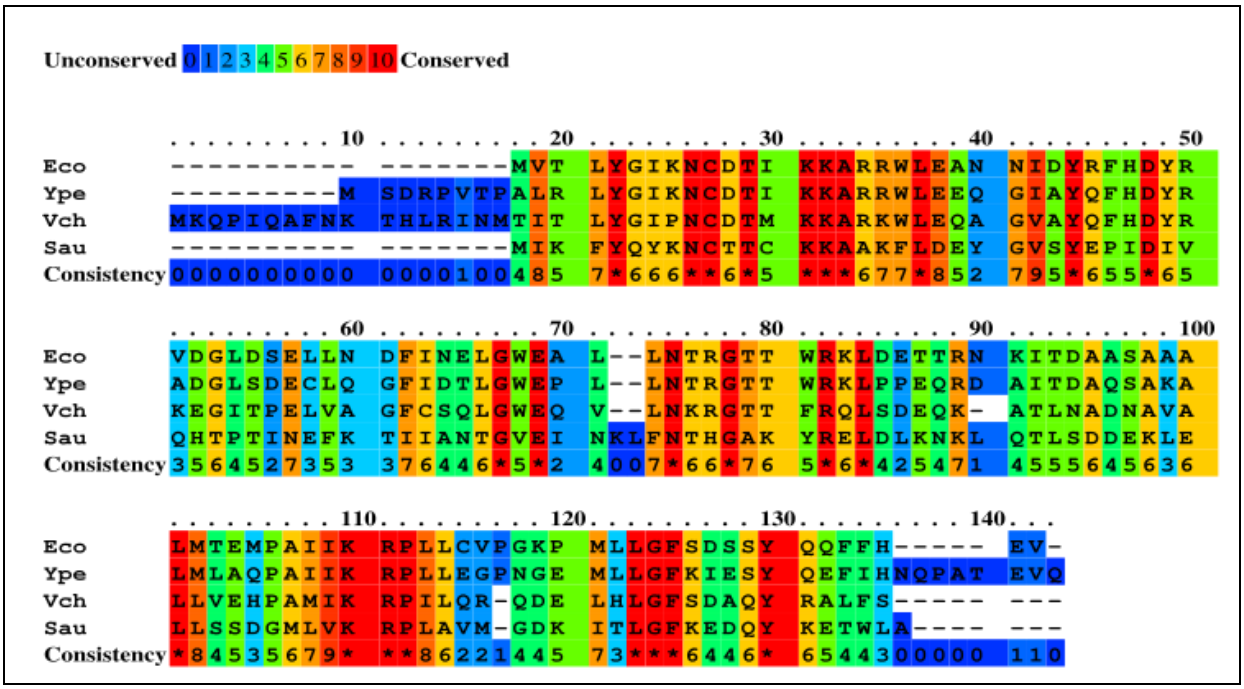
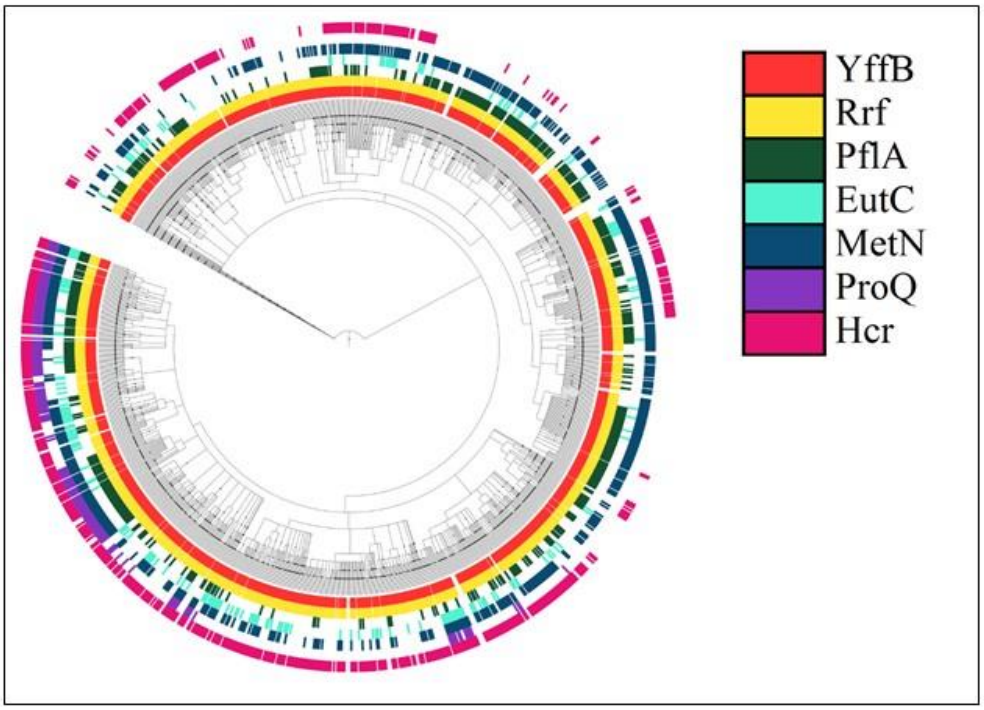


Fig. 4.26 – Protein sequence alignment for YffB for the four species *Eco, Ype, Vch* and *Sau*

Phylogenetic Analysis. Phylogenetic tree for YffB shows high conservation across various bacterial species (Fig. 4.27). However, most of the interacting partners in *Eco* do not show high level of conservation across the bacterial species. Only two interactions (MetN and ProQ) are conserved across all four species. MetN is conserved across most of the bacterial species however, ProQ is present in a few proteobacterial species. There does not seem to be a correlation between conservation of the protein across bacterial species and conservation of interaction.

A.



B.

Proteins	Number of species
YffB	1141
Frr	1135
PflA	574
EutC	242
MetN	753
ProQ	195
Hcr	706
InsN1	498

Fig. 4.27 – A. Phylogenetic tree representing the distribution of protein YffB and its interacting proteins in bacterial species B. Number of species the proteins interacting with the hub protein are present in relative to presence of YffB represented in the phylogenetic tree

Sequence Identity Analysis for Interacting proteins. Two interactions tested using yeast two hybrid system are conserved across all species that the proteins were present in. metN showed 60-75% identity in *Ype, Vch* and *Sau*. proQ on the other hand is absent in most of the bacterial species including *Staphylococcus aureus* and shows less than 50% sequence identity in *Vch* and *Ype*, however, the interactions is conserved in the two species.

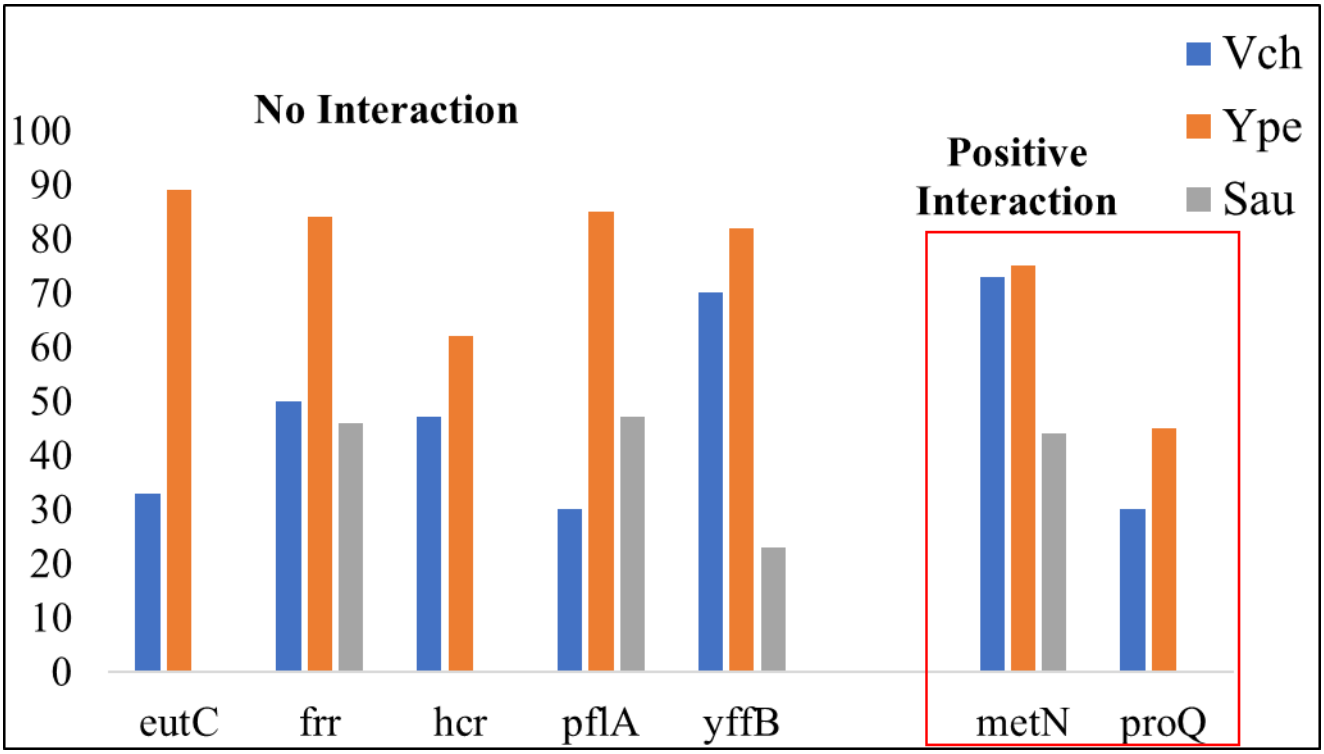


Fig. 4.28 – Sequence identity for proteins tested for interaction with hub protein YffB in comparison with *Eco*. The proteins in the red box showed positive interactions when tested using yeast two hybrid assay in at least one of the three species used in this study

Network Analysis. Based on literature, YffB shows 10 interactions among which only 6 were conserved in *Eco*. Among the 6 interactions observed in *Eco*, only 2 were observed across all the species that the protein was present in. YffB mainly interacts with protein metN which is present

across most bacterial species in the phylogenetic tree. Functional prediction cannot be made based on these interactions. Functionally studied paralog of YffB is ArsC, arsenate reductase. YffB is structurally closely related to ArsC and hence it might have a similar or closely related function.

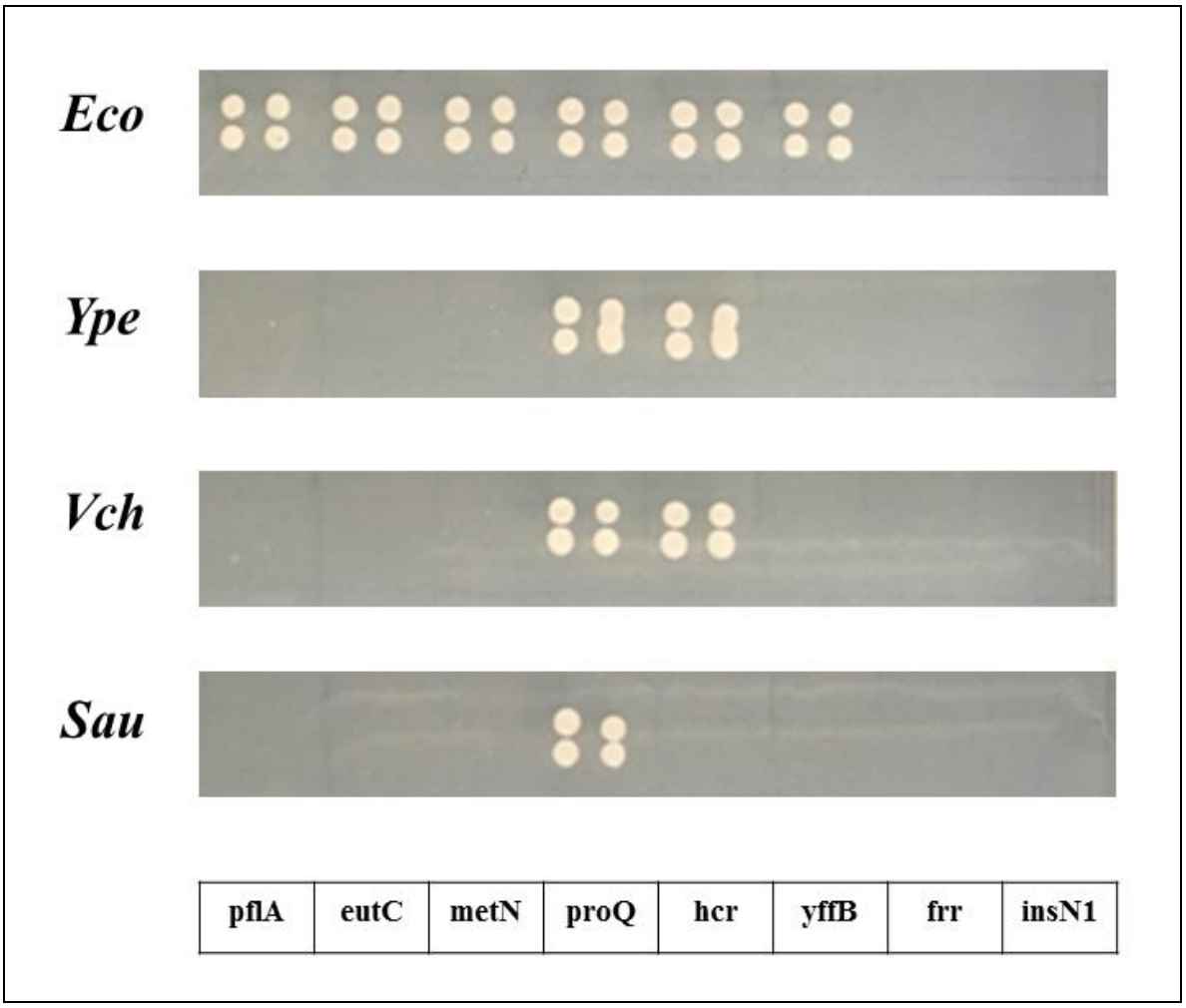


Fig. 4. 29 – Yeast two hybrid screen for hub protein YffB at 10mM 3-AT

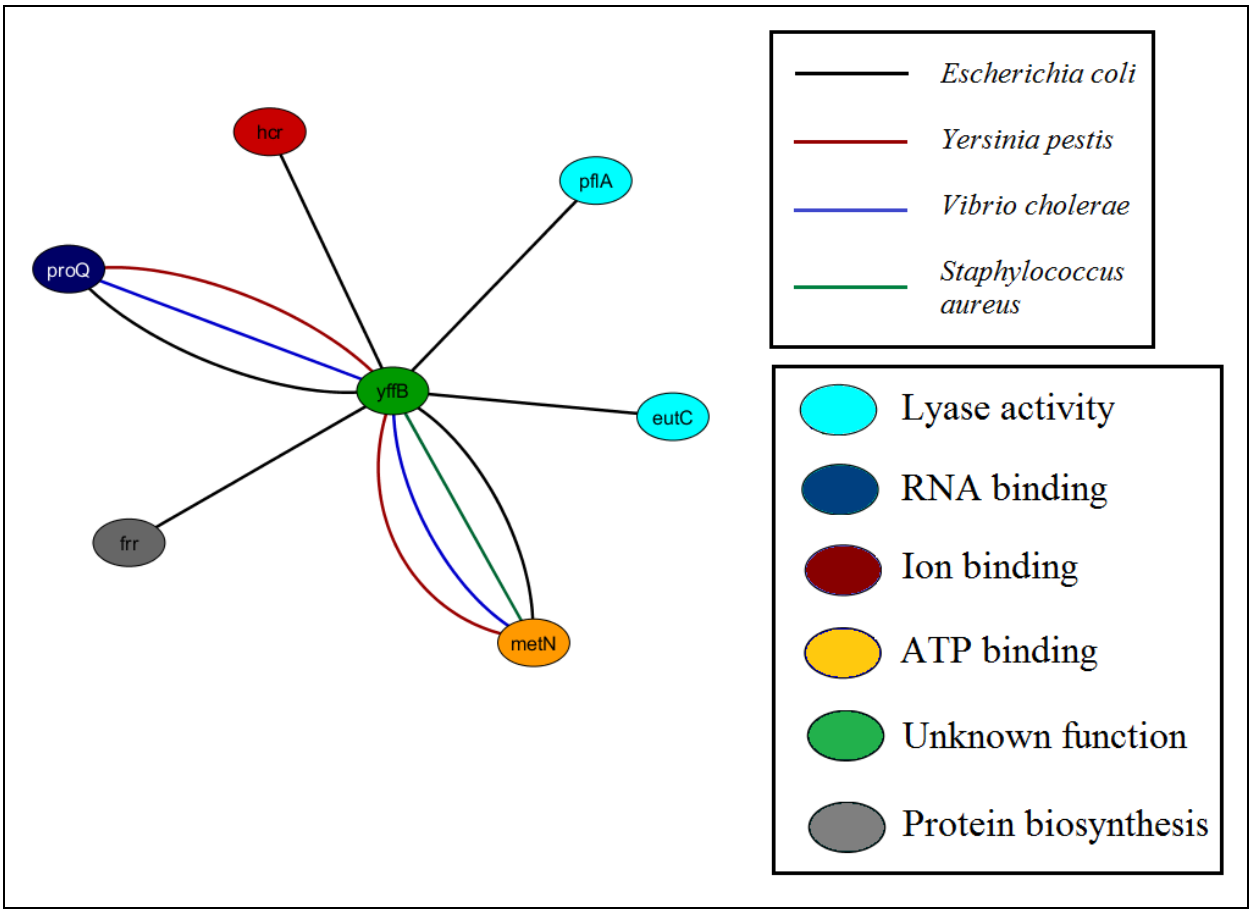


Fig. 4.30 – Protein – protein interaction network for hub protein YffB

Discussion

The four hub proteins YeeD, YjjW, YbhK, and YffB were tested for interactions using yeast two hybrid system. ~85% of the interactions from previous studies for *Escherichia coli* were reproduced during these screens. Interactions for *Vibrio cholerae, Yersinia pestis*, and *Staphylococcus aureus* were tested in parallel with *Escherichia coli.* Sequence analysis, structure analysis, phylogenetic analysis, and network analysis was carried out for each of these proteins as well as their paralogs. The hub proteins are not essential proteins and might be required by the organism only under certain conditions nevertheless these proteins were highly conserved.

For all of the hub proteins close to 50% of the interacting partners were also conserved in all of the species included in this study. Since the interacting proteins are present even is distantly related species, these proteins might either be essential proteins or must be required by cells under different stress conditions. This study yielded a set of high confidence interactions for each of the hub proteins. Though a high percentage of interactions were not conserved, we cannot conclude that these were false positives since various other factors play a role in conservation of interactions.

Sequence alignment for hub proteins. Protein sequence alignments revealed that not all hub proteins were highly conserved across the four species. Among ~60 other hub proteins identified in *Eco*, only four of these hub proteins are present in *Ype*, *Vch* and *Sau* and ~1000 other species.

Hub protein YeeD. YeeD showed low sequence identity in *Ype*, and *Vch* in comparison to *Eco*. However, YeeD showed short sequences along the alignment which were well conserved in all four species. Five PPIs for YeeD were conserved across all four species, TruA, ClpB, DdlB, TreR, and MurA. Analysis of sequence similarity for proteins interacting with YeeD in the four species was carried out. No conclusive trends were observed from sequence similarity analysis for interacting proteins between conserved and not conserved interactions. In case of YeeD, protein structure was available for two of its paralogs, TusA and YedF (unknown function protein). The structures TusA and YedF were fairly similar based on PDB structure similarity analysis inspite of low protein sequence similarity. Network analysis results indicated that YeeD showed conserved interaction with proteins involved in transcription as well as in biosynthetic pathways. Hence network analysis does not lead to conclusive direction for functional annotations for this protein. Targeted studies have not been carried out for YeeD protein or its orthologs YeeD in other species. However, based on interaction data, sequence analysis and

structure analysis of paralogs, YeeD is more likely to be involved in sulfur transferase and hence interacts with proteins from various cellular systems.

Hub protein YjjW. Protein YjjW shows <30% sequence identity in *Ype, Vch* as well as *Sau* in comparison with *Eco*. YjjW is conserved in ~2000 other species. Small stretches of the protein were well conserved across all four species. YjjW is a hub protein interacting with more ~20 proteins. Only two proteins MurI and LdcA showed conserved interactions with YjjW in all four species. Both the proteins are involved in functions related to cell division. YjjW has numerous (more than 20) metal binding sites along the length of the protein and belongs to the S-adenosyl methionine family of proteins. This might be the reason why YjjW has been annotated as a glycyl-radical activating enzyme. However, paralogs of YjjW have been studied experimentally and have been annotated as pyruvate formate-lyase activating enzymes, i.e enzymes involved in anaerobic glucose metabolism pathway. Based on paralogy information as well as interaction data, a hypothesis can be drawn. A relevant functional hypothesis would be that the iron-sulfur cluster activity might be required during certain stages of cell division in presence on non-glucose carbon sources.

Hub protein YbhK. YbhK is a protein present in ~1100 species. The sequence alignment across the four species included in this study indicate a high level of sequence identity. Many stretches of protein sequences are conserved across all four species. These might be the regions involved in conserved PPIs. YbhK does not have any paralogs however its orthologs have been studied in *Bacillus subtilis* and *Mycobacterium smegmatis*. Neither of the studies could determine the cellular function for the protein, but some common observations were made in both studies. Knockout studies for YbhK showed growth defects (slow growing cells, elongation of cells) under certain media conditions, specifically media with non-glucose carbon sources. The

knockout study in *Bacillus subtilis* replaced YvcK with the *Eco* YbhK which helped restore the function of YvcK. Hence, YbhK can be expected to have a function similar to its homologs. Based on these experiments as well as computational analysis it can be hypothesized that this protein is involved in glucose metabolism for krebs cycle intermediates like succinate, oxaloacetic acid, etc. as well as having a functional role in the phosphotransferase system in conjunction will cell shape maintenance. Experiments can be carried out in *Eco* using YbhK knockouts to determine if it shows a similar behavior.

Hub protein YffB. Protein YffB showed very low conservation. It was observed in case of YffB that its interacting partner proQ was present only in a few proteobacteria species, however the interaction was conserved across *Eco, Vch* and *Ype*. The protein ProQ was not present in *Sau*.

Crystal structure of YffB protein is available on protein databank (PDB). Structure similarity analysis of YffB with its paralog ArsC revealed 30% similarity. YffB protein had only two interactions conserved across the four species and rest of the interactions were observed only in *Eco*. Functionally studied paralog of YffB is ArsC annotated as arsenate reductase. YffB is structurally closely related to ArsC and hence YffB might be a part of the process of arsenate reduction. The hypothesis for functional annotations drawn from this study can be used to plan further experiments and study each protein in detail.

CHAPTER V

CONCLUSION

This dissertation explores protein – protein interactions for proteins with unknown function. The dissertation specifically focuses on PPIs for paralogous proteins in *Escherichia coli*, and orthologous proteins across four pathogenic species. The main aim is to study interactions involving unknown function or uncharacterized proteins. For the entire study, the PPIs for Eco were considered as gold standard since the interaction data as well as the clone sets used in this study were obtained from Rajagopala *et al*., 2014 publication.

Paralogous proteins occur in organisms due to gene duplication events. Since the genes share a common ancestry, it is interesting to study these proteins to determine whether they share interacting partners thus implying that they have a similar function. More than half of the *Eco* proteome comprises of paralogs. Among these paralogs quite a few of them do not have a known function. Two paralogous protein groups were tested. The first paralogous protein group included proteins AllD, DlgD and YbiC. Based on literature information, proteins AllD and DlgD work as enzymes in very distinct pathways, however, the nature of their function is very similar. These proteins also share a high percentage of structure similarity. The interaction study did not yield any shared interactions. Hence, the interactions need to be tested more thoroughly using all possible vector combinations. The second paralogous protein pair included proteins SmrA and YfcN. Protein SmrA is a probably endonuclease and has a partial structure. YfcN on

the other hand has no known protein function or structure. The yeast two hybrid screen had only on shared interaction with protein MinC which is a septum ring protein. Reports have shown that endonucleases have an effect on septum ring formation and cytokinesis (Shukla, 2012). Hence, the interaction is relevant. However, in this case also the shared interactions were very low and hence there is a need to test the interactions using all four vector combinations.

Two different approaches were used to compare interactions among pathogenic species. One approach was using COG data to identify orthologs and the other approach involved using sequence homology. The major reason for these approaches was to carry out interactions using the same method in parallel to make the interaction data directly comparable. In both cases the interactions were tested using yeast two hybrid system, however, the approach for determining the proteins being tested was distinct.

In the case of comparative Interactomics, COG data from EggNOG 3.0 was used to determine orthologs between *Eco, Ype, Vch,* and *Sau*. Either or both of the proteins tested for PPI was an uncharacterized protein or protein with a probable/putative function. Based on the conservation of interactions among multiple species, a high confidence interactions dataset was created. Some of the binary interactions were conserved across more than two species. These interactions would be interesting candidates for follow up interaction studies. Two proteins were identified as proteins of interest since they appeared to be hubs for the network and shared multiple interactions conserved across more than one species. It was observed that very few interactions were conserved in *Sau* possibly because it is distantly related to *Eco*. Another reason for low overlap with *Sau* could be the lack of revival of some clones.

For comparative interactomics involving hub proteins, four uncharacterized hub proteins conserved across the four pathogenic species as well as conserved across more than 1,000 other

species were used. Yeast two hybrid assay was used to test the interactions and to enable narrowing down on a particular functional area for these proteins. In most cases, these proteins shared high sequence similarity within species. Also, two of the hub proteins had paralogs which could be useful in narrowing down on the function. A combination of yeast two hybrid data, paralog and structure information along with domain, phylogeny and sequence analysis was used to pin point a functional area for these proteins wherever possible.

The dissertation uses a high throughput array based approach to studying PPIs. Since clone sets are available only for a few organisms, this study tries to use organisms that are closely related to have conserved proteins but having a large enough phylogenetic distance to enable high confidence PPIs. These PPIs can be used for future studies. Each of the uncharacterized protein with multiple high confidence interactions can be further studied to make functional annotations. A literature survey of the proteins along with a summary of useful publications has been included to aid in determining a future direction of work for them. Since a very huge portion of the proteome has no functional annotation at all, this study can aid in having probable functions for multiple proteins based on experimental and data analysis evidence.

BIBLIOGRAPHY

Andersson, S. G., Zomorodipour, a, Andersson, J. O., Sicheritz-Pontén, T., Alsmark, U. C., Podowski, R. M., … Kurland, C. G. (1998). The genome sequence of Rickettsia prowazekii and the origin of mitochondria. *Nature*, *396*(6707), 133–140. http://doi.org/10.1038/24094

Arifuzzaman, M., Maeda, M., Itoh, A., Nishikata, K., Takita, C., Saito, R., … Mori, H. (2006). Large-scale identification of protein – protein interaction of Escherichia coli K-12. *Genome Research*, *16*, 686–691. http://doi.org/10.1101/gr.4527806.8

Batada, N. N., Hurst, L. D., & Tyers, M. (2006). Evolutionary and physiological importance of hub proteins. *PLoS Computational Biology*, *2*(7), 0748–0756. http://doi.org/10.1371/journal.pcbi.0020088

Brückner, A., Polge, C., Lentze, N., Auerbach, D., & Schlattner, U. (2009). Yeast two-hybrid, a powerful tool for systems biology. *International Journal of Molecular Sciences*, *10*(6), 2763–2788. http://doi.org/10.3390/ijms10062763

Chen, L., Xuan, J., Riggins, R. B., Clarke, R., & Wang, Y. (2011). Identifying cancer biomarkers by network-constrained support vector machines. *BMC Systems Biology*, *5*(1), 161. http://doi.org/10.1186/1752-0509-5-161

Cherkasov, A., Hsing, M., Zoraghi, R., Foster, L. J., See, R. H., Stoynov, N., … Reiner, N. E. (2011). Mapping the protein interaction network in methicillin-resistant Staphylococcus aureus. *Journal of Proteome Research*, *10*(3), 1139–1150. http://doi.org/10.1021/pr100918u

Cusa, E., Obradors, N., Baldomà, L., Badía, J., Badi, J., Cusa, E. V. a, & Baldoma, L. (1999). Genetic Analysis of a Chromosomal Region Containing Genes Required for Assimilation of Allantoin Nitrogen and Linked Glyoxylate Metabolism in Escherichia coli Genetic Analysis of a Chromosomal Region Containing Genes Required for Assimilation of Allantoin. *Journal of Bacteriology*, *181*(24), 7479–7484.

David Whitford. (2005). Protein structure and function. In *Protein structure and function* (pp. 1–6). Hoboken N.J.: J. Wiley & Sons.

Deng, W., Burland, V., Iii, G. P., Boutin, A., Mayhew, G. F., Liss, P., … Blattner, F. R. (2002). Genome Sequence of Yersinia pestis KIM Genome Sequence of Yersinia pestis KIM †. *Journal of Bacteriologyteriology*, *184*(16), 4601–4611. http://doi.org/10.1128/JB.184.16.4601

Dey, B., Thukral, S., Krishnan, S., Chakrobarty, M., Gupta, S., Manghani, C., & Rani, V. (2012). DNA-protein interactions: Methods for detection and analysis. *Molecular and Cellular Biochemistry*, *365*(1–2), 279–299. http://doi.org/10.1007/s11010-012-1269-z

Erdin, S., Lisewski, A. M., & Lichtarge, O. (2011). Protein function prediction: Towards integration of similarity metrics. *Current Opinion in Structural Biology*, *21*(2), 180–188. http://doi.org/10.1016/j.sbi.2011.02.001

Feng, S., Zhou, L., Huang, C., Xie, K., & Nice, E. C. (2015). Interactomics: toward protein function and regulation. *Expert Review of Proteomics*, *12*(1), 37–60. http://doi.org/10.1586/14789450.2015.1000870

Finkelstein R A. (1996). *Cholera, Vibrio cholerae O1 and O139, and Other Pathogenic Vibrios*. *Medical Microbiology* (Forth). Galveston TX.

Foulquier, E., Pompeo, F., Freton, C., Cordier, B., Grangeasse, C., & Galinier, A. (2014). PrkC-

mediated phosphorylation of overexpressed YvcK protein regulates PBP1 protein localization in Bacillus subtilis mreB mutant cells. *Journal of Biological Chemistry*, *289*(34), 23662–23669. http://doi.org/10.1074/jbc.M114.562496

Gevers, Dirk; Vandepoele, Klaas; Simillion, Cedric; Van de Peer, Y. (2004). Gene duplication and biased functional retention of paralogs in bacterial genomes. *Trends in Microbiology*, *12*(4), 145–148. http://doi.org/10.1016/j.tim.2004.02.004

Gorke, B., Foulquier, E., & Galinier, A. (2005). YvcK of Bacillus subtilis is required for a normal cell shape and for growth on Krebs cycle intermediates and substrates of the pentose phosphate pathway. *Microbiology*, *151*(11), 3777–3791. http://doi.org/10.1099/mic.0.28172-0

Gui, W. J., Qu, Q. H., Chen, Y. Y., Wang, M., Zhang, X. E., Bi, L. J., & Jiang, T. (2011). Crystal structure of YdaL, a stand-alone small MutS-related protein from Escherichia coli. *Journal of Structural Biology*, *174*(2), 282–289. http://doi.org/10.1016/j.jsb.2011.01.008

Guruharsha, K G; Rual, JF; Zhai, B; Mintseris, J; Vaidya, N; Beekman, C; Wong, C; Rhee, DY; Cenaj, O; McKilli[, E; Shah, S; Spatleton, M; Wan, KH; Yu, CH; Artavanis-Tsakonas, S. (2011). A Protein Complex Network of Drosophila melanogaster. *Cell*, *147*(3), 690–703. http://doi.org/10.1016/j.cell.2011.08.047.A

Häuser, R., Ceol, A., Rajagopala, S. V, Mosca, R., Siszler, G., Wermke, N., … Uetz, P. (2014). A second-generation protein-protein interaction network of Helicobacter pylori. *Molecular & Cellular Proteomics : MCP*, *13*(5), 1318–29. http://doi.org/10.1074/mcp.O113.033571

Hu, P., Janga, S. C., Babu, M., Díaz-Mejía, J. J., Butland, G., Yang, W., … Emili, A. (2009). Global Functional Atlas of Escherichia coli Encompassing Previously Uncharacterized Proteins. *PLoS Biology*, *7*(4), e96. http://doi.org/10.1371/journal.pbio.1000096

Ikeuchi, Y., Shigi, N., Kato, J. I., Nishimura, A., & Suzuki, T. (2006). Mechanistic insights into sulfur relay by multiple sulfur mediators involved in thiouridine biosynthesis at tRNA wobble positions. *Molecular Cell*, *21*(1), 97–108. http://doi.org/10.1016/j.molcel.2005.11.001

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., & Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(8), 4569–74. http://doi.org/10.1073/pnas.061034498

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, *44*(D1), D457–D462. http://doi.org/10.1093/nar/gkv1070

Kaper, J. B., Nataro, J. P., & Mobley, H. L. (2004). Pathogenic Escherichia coli. *Nature Reviews. Microbiology*, *2*(2), 123–140. http://doi.org/10.1038/nrmicro818

Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., … Greenblatt, J. F. (2006). Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature*, *440*(7084), 637–643. http://doi.org/10.1038/nature04670

Kühner, S., van Noort, V., Betts, M. J., Leo-Macias, A., Batisse, C., Rode, M., … Gavin, A.-C. (2009). Proteome organization in a genome-reduced bacterium. *Science (New York, N.Y.)*, *326*(5957), 1235–1240. http://doi.org/10.1126/science.1176343

Li, G., & Young, K. D. (2012). Isolation and identification of new inner membrane-associated proteins that localize to cell poles in Escherichia coli. *Molecular Microbiology*, *84*(2), 276–295. http://doi.org/10.1111/j.1365-2958.2012.08021.x

Li, S., Armstrong, C. M., Bertin, N., Ge, H., Milstein, S., Boxem, M., … Vidal, M. (2004). A

Map of the Interactome Network of the Metazoan C. elegans. *Science (New York, N.Y.)*, *303*(5657), 540–543.

McFedries, A., Schwaid, A., & Saghatelian, A. (2013). Methods for the elucidation of protein-small molecule interactions. *Chemistry and Biology*, *20*(5), 667–673. http://doi.org/10.1016/j.chembiol.2013.04.008

McHugh, C. A., Russell, P., Guttman, M., Chen, M., Manley, J., Licatalosi, D., … Brown, P. (2014). Methods for comprehensive experimental identification of RNA-protein interactions. *Genome Biology*, *15*(1), 203. http://doi.org/10.1186/gb4152

Mehla, J., Caufield, J. H., & Uetz, P. (2015). The Yeast Two-Hybrid System: A Tool for Mapping Protein–Protein Interactions. *Cold Spring Harbor Protocols*, *2015*(5), pdb.top083345. http://doi.org/10.1101/pdb.top083345

Mehla, J., Dedrick, R. M., Harry Caufield, J., Siefring, R., Mair, M., Johnson, A., … Uetz, P. (2015). The protein interactome of mycobacteriophage giles predicts functions for unknown proteins. *Journal of Bacteriology*, *197*(15), 2508–2516. http://doi.org/10.1128/JB.00164-15

Mir, M., Prisic, S., Kang, C. M., Lun, S., Guo, H., Murry, J. P., … Husson, R. N. (2014). Mycobacterial gene cuvA is required for optimal nutrient utilization and virulence. *Infection and Immunity*, *82*(10), 4104–4117. http://doi.org/10.1128/IAI.02207-14

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., & Kanehisa, M. (1999). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *27*(1), 29–34. http://doi.org/10.1093/nar/27.1.29

Peter, K. (2008). Genomics Protocols, *439*, 1–16. http://doi.org/10.1007/978-1-59745-188-8

Powell, S., Szklarczyk, D., Trachana, K., Roth, A., Kuhn, M., Muller, J., … Bork, P. (2012). eggNOG v3.0: Orthologous groups covering 1133 organisms at 41 different taxonomic

ranges. *Nucleic Acids Research*, *40*(D1), 1–6. http://doi.org/10.1093/nar/gkr1060

Rajagopala, S. V, Sikorski, P., Kumar, A., Mosca, R., Vlasblom, J., Arnold, R., … Uetz, P. (2014). The binary protein-protein interaction landscape of Escherichia coli. *Nature Biotechnology*, *32*(3), 285–290. http://doi.org/10.1038/nbt.2831.The

Rajagopala, S. V, Sikorski, P., Kumar, A., Mosca, R., Vlasblom, J., Arnold, R., … Uetz, P. (2014). The binary protein-protein interaction landscape of Escherichia coli. *Nature Biotechnology*, *32*(3), 285–90. http://doi.org/10.1038/nbt.2831

Rolland, T., Ta??an, M., Charloteaux, B., Pevzner, S. J., Zhong, Q., Sahni, N., … Vidal, M. (2014). A proteome-scale map of the human interactome network. *Cell*, *159*(5), 1212–1226. http://doi.org/10.1016/j.cell.2014.10.050

Schwikowski, B., Uetz, P., & Fields, S. (2000). A network of protein-protein interactions in yeast. *Nature Biotechnology*, *18*(12), 1257–1261. http://doi.org/10.1038/82360

Shi, L., Pigeonneau, N., Ventroux, M., Derouiche, A., Bidnenko, V., Mijakovic, I., … Mijakovic, I. (2014). Cross-phosphorylation of bacterial serine/threonine and tyrosine protein kinases on key regulatory residues. *Frontiers in Microbiology*, *5*(SEP), 1–13. http://doi.org/10.3389/fmicb.2014.00538

Shukla, M., Minda, R., Singh, H., Tirumani, S., Chary, K. V. R., & Rao, B. J. (2012). UVI31+ Is a DNA Endonuclease That Dynamically Localizes to Chloroplast Pyrenoids in C. reinhardtii. *PLoS ONE*, *7*(12). http://doi.org/10.1371/journal.pone.0051913

Skwarczynska, M., & Ottmann, C. (2015). Protein–protein interactions as drug targets. *Future Medicinal Chemistry*, *7*, 2195–2219. http://doi.org/10.4155/fmc.15.138

Titz, B., Rajagopala, S. V., Goll, J., Häuser, R., McKevitt, M. T., Palzkill, T., & Uetz, P. (2008). The binary protein interactome of Treponema pallidum - The syphilis spirochete. *PLoS*

ONE, *3*(5). http://doi.org/10.1371/journal.pone.0002292

Uetz, P., Giot, L., Cagney, G., Mansfield, T. a, Judson, R. S., Knight, J. R., … Rothberg, J. M. (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature*, *403*(6770), 623–627. http://doi.org/10.1038/35001009

Vallabhajosyula, R. R., Chakravarti, D., Lutfeali, S., Ray, A., & Raval, A. (2009). Identifying Hubs in protein interaction networks. *PLoS ONE*, *4*(4), 1–10. http://doi.org/10.1371/journal.pone.0005344

Wang, W. U., Chen, C., Lin, K., Fang, Y., & Lieber, C. M. (2005). Label-free detection of small-molecule-protein interactions by using nanowire nanosensors. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(9), 3208–3212. http://doi.org/10.1073/pnas.0406368102

Wang, Y., Cui, T., Zhang, C., Yang, M., Huang, Y., Li, W., … He, Z. G. (2010). Global protein-protein interaction network in the human pathogen mycobacterium tuberculosis H37Rv. *Journal of Proteome Research*, *9*(12), 6665–6677. http://doi.org/10.1021/pr100808n

Wang, Y., Cui, T., Zhang, C., Yang, M., Huang, Y., Li, W., … He, Z. G. (2010). Global protein-protein interaction network in the human pathogen Mycobacterium tuberculosis H37Rv. *J Proteome Res*, *9*(12), 6665–6677. http://doi.org/10.1021/pr100808n

Yew, W. S., Yew, W. S., Gerlt, J. a, & Gerlt, J. a. (2002). Utilization of L-Ascorbate by Escherichia coli K-12: Assignments of Functions to Products of the yjf-sga and yia-sgb Operons. *Society*, *184*(1), 302–306. http://doi.org/10.1128/JB.184.1.302

Yu, H., Braun, P., Yildirim, M. A., Lemmens, I., Venkatesan, K., Sahalie, J., … Vidal, M. (2008). High-Quality Binary Protein Interaction Map of the Yeast Interactome Network. *Science*, *322*(5898), 104–110. http://doi.org/10.1126/science.1158684

Yu, J., Pacifico, S., Liu, G., & Finley, R. L. (2008). DroID: the Drosophila Interactions Database, a comprehensive resource for annotated gene and protein interactions. *BMC Genomics*, *9*, 461. http://doi.org/10.1186/1471-2164-9-461

Zhu, J., & Shimizu, K. (2004). The effect of pfl gene knockout on the metabolism for optically pure D-lactate production by Escherichia coli. *Applied Microbiology and Biotechnology*, *64*(3), 367–375. http://doi.org/10.1007/s00253-003-1499-9

APPENDIX A

1. Construction of bait and prey clones

1) The clone sets for protein coding genes were obtained for all organisms from JCVI and BEI Resources.
2) The clones were present as gateway compatible pDONR vectors in *Escherichia coli* cells. The pDONR plasmids were isolated from *Escherichia coli* cells using Machery Nagel plasmid miniprep kit.
3) The genes were cloned from pDONR vector to expression vector pGADT7g (prey) and pGBGT7g (bait) using gateway cloning procedure. For gateway cloning 80 – 100 ng (~2 ul) of donor (pDONR, ccdB gene) plasmid was mixed with 150 – 200 ng (~1-1.5 ul) of destination (pGADT7g or pGBGT7g, gene of interest) plasmid and added to a PCR tube.
4) ~2 ul of TE buffer and 0.6 ul of LR Clonase was added to the mix and the tube was incubated at 25ºC overnight (18 – 20 hours).

5) Next day 1 ul of Proteinase K was added to the tube and incubated at 37ºC for 10 minutes.

6) 2 ul of the reaction mix was used for transformation into *Escherichia coli* ccdB sensitive chemically competent cells for selection of successfully cloned colonies.

7) Plasmids containing gene of interest were isolated from the transformed cells by plasmid miniprep using Machery Nagel plasmid miniprep kit.

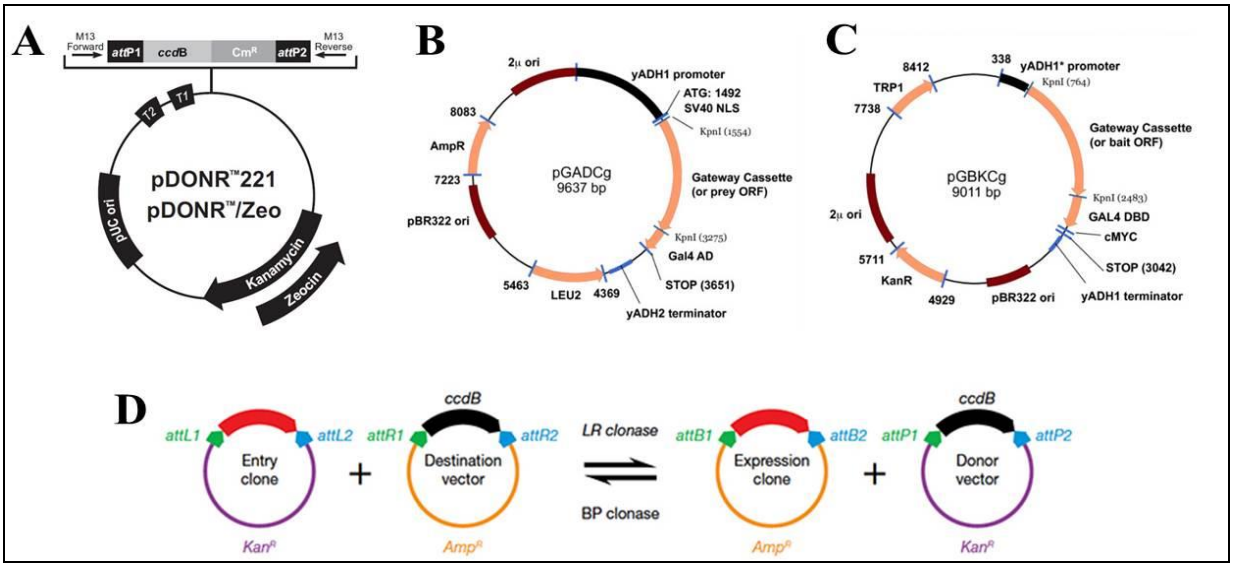


Fig. - A. Plasmid map of pDONR221 B. Plasmid map of prey plasmid (pGADT7g) C. Plasmid map of bait plasmid (pGBKT7g) D. Gateway cloning process

2. Yeast transformation

1) The yeast strains were cultured overnight in 5 ml YEPAD (0.5% yeast extract, 1% peptone, 1% dextrose and 0.01 adenine hemisulphate) liquid medium in a 15 ml tube at 30° C in a shaking incubator.

2) The cells were diluted in 100ml fresh liquid YEPAD medium to an OD600 of 0.2 – 0.3. The cells were allowed to grow for atleast 2 generations (3-4 hours) in a 30° C shaking incubator.
3) The cells were then transferred to two 50 ml tubes. The tubes were centrifuged for 5 minutes at 2,000 rpm to collect the cells. The supernatant was discarded.
4) The cells were washed in 25-50 ml sterile distilled water. The tubes were centrifuged again to precipitate cells. The supernatant was discarded.
5) The cells were then resuspended in 1ml of 0.1 M lithium acetate (LiAc) prepared in Tris EDTA (TE) buffer and incubated on ice for 15 minutes.
6) The cell suspension was then transferred to 1.5 ml centrifuge tubes and briefly centrifuged. The supernatant was discarded and the pellet was resuspended in 1ml LiAc prepared in TE buffer. The tubes were then incubated on ice for 15 – 20 minutes.
7) A mixture was prepared for 20 reactions. This mixture contained 50 µl carrier DNA (10mg/ml) added to a tube containing 2 ml of 40% PEG to which 1 ml of cells prepared from step 6 were added. The mixture was vortexed for 1 minute. 100 µl of this mixture was added to 20 wells of a 96 well plate.
8) 100 ng of the desired plasmid DNA construct was added to the wells in the plate. A negative control containing no plasmid and a positive control containing an empty plasmid were also included.
9) The 96 well plate was sealed with aluminum foil and parafilm. The plate was gently vortexed and placed in a 30° C shaking incubator for 45 minutes.
10) The plate was then transferred to a water bath at 42° C for 30 minutes.

11) The plates was then centrifuged for 10 minutes at 2,000 rpm. The supernatant was discarded and the pellet was resuspended in 50 µl sterile distilled water.
12) The resuspended cells were plated on selective medium plates and incubated for 48 – 72 hours at 30° C.

3. Bait self-activation test

1) Empty prey vector (pGADT7g in Y187) was cultures overnight (16 – 18 hours) in YEPAD medium at 30° C in a shaking incubator.
2) The test baits were mated with the empty prey vector as follows:
    i. Baits were plated on a 96 well plate (omnitray) containing solid YEPAD medium.
    ii. 100 µl of empty prey vector was transferred to a 96 well plate.
    iii. The empty prey was plated using a 96 well pinning assembly onto the plated baits.
    iv. This plate (omni tray) was incubated at 30° C for 48 – 72 hours.
3) The colonies from the YEPAD solid medium were transferred to double-dropout selective medium (-LT) plates. These plates were incubated at 30° C for 48 – 72 hours.
4) These colonies were then transferred to -LTH plates (HIS3 is the reporter gene) supplemented with 3-AT (0mM – 10mM). These plates were incubated at 30° C for ~ 1 week.

4. Yeast two hybrid assay

1) After bait self-activation, the bait and prey were cultured overnight in respective selective media in a 96 well plate at 30° C in a shaking incubator.

2) The baits were pinned on a omni tray containing solid YEPAD medium using the 96 well plate pinning assembly. Preys were pinned on the same plate on top of the baits. The plate was incubated 30° C for 48 – 72 hours.

3) The colonies from this plate were transferred to -LT plate. The plate was incubated 30° C for 48 – 72 hours.

5) The colonies from -LT plate were transferred to -LTH plate supplemented with varying concentrations of 3-AT (0 – 10mM) to detect protein -protein interactions. This plate was incubated at 30° C for ~ 1 week. The results were noted for each 3-AT concentration.

APPENDIX B

Appendix Table 2.1 - List of distinct OGs with paralogs in *Escherichia coli* K12 proteome

Sr. No.	COG ID	COG description	# of proteins in the family
1	COG0006	Xaa-Pro aminopeptidase	2
2	COG0008	Glutamyl- and glutaminyl-tRNA synthetases	2
3	COG0009	Putative translation factor (SUA5)	2
4	COG0021	Transketolase	2
5	COG0028	Thiamine pyrophosphate-requiring enzymes [acetolactate synthase	4
6	COG0031	Cysteine synthase	2
7	COG0036	Pentose-5-phosphate-3-epimerase	2
8	COG0037	Predicted ATPase of the PP-loop superfamily implicated in cell cycle control	2
9	COG0038	Chloride channel protein EriC	2
10	COG0042	tRNA-dihydrouridine synthase	2
11	COG0044	Dihydroorotase and related cyclic amidohydrolases	2
12	COG0050	GTPases - translation elongation factors	2
13	COG0057	Glyceraldehyde-3-phosphate dehydrogenase/erythrose-4-phosphate dehydrogenase	2
14	COG0058	Glucan phosphorylase	2
15	COG0069	Glutamate synthase domain 2	2
16	COG0070	Glutamate synthase domain 3	2
17	COG0071	Molecular chaperone (small heat shock protein)	2

18	COG0073	EMAP domain	2
19	COG0074	Succinyl-CoA synthetase	2
20	COG0076	Glutamate decarboxylase and related PLP-dependent proteins	2
21	COG0078	Ornithine carbamoyltransferase	2
22	COG0084	Mg-dependent DNase	2
23	COG0110	Acetyltransferase (isoleucine patch superfamily)	4
24	COG0111	Phosphoglycerate dehydrogenase and related dehydrogenases	2
25	COG0115	Branched-chain amino acid aminotransferase/4-amino-4-deoxychorismate lyase	2
26	COG0119	Isopropylmalate/homocitrate/citramalate synthases	2
27	COG0129	Dihydroxyacid dehydratase/phosphogluconate dehydratase	4
28	COG0132	Dethiobiotin synthetase	2
29	COG0136	Aspartate-semialdehyde dehydrogenase	2
30	COG0142	Geranylgeranyl pyrophosphate synthase	2
31	COG0144	tRNA and rRNA cytosine-C5-methylases	2
32	COG0147	Anthranilate/para-aminobenzoate synthases component I	2
33	COG0156	7-keto-8-aminopelargonate synthetase and related enzymes	2
34	COG0160	4-aminobutyrate aminotransferase and related aminotransferases	2
35	COG0167	Dihydroorotate dehydrogenase	2
36	COG0168	Trk-type K+ transport systems	2
37	COG0169	Shikimate 5-dehydrogenase	2
38	COG0174	Glutamine synthetase	2
39	COG0175	3'-phosphoadenosine 5'-phosphosulfate sulfotransferase (PAPS reductase)/FAD synthetase and related enzymes	2
40	COG0176	Transaldolase	4
41	COG0183	Acetyl-CoA acetyltransferase	5
42	COG0187	Type IIA topoisomerase (DNA gyrase/topo II	2
43	COG0188	Type IIA topoisomerase (DNA gyrase/topo II	2
44	COG0189	Glutathione synthase/Ribosomal protein S6 modification enzyme (glutaminyl transferase)	2
45	COG0191	Fructose/tagatose bisphosphate aldolase	4

46	COG0204	1-acyl-sn-glycerol-3-phosphate acyltransferase	2
47	COG0208	Ribonucleotide reductase	2
48	COG0209	Ribonucleotide reductase	2
49	COG0210	Superfamily I DNA and RNA helicases	2
50	COG0217	Uncharacterized conserved protein	2
51	COG0223	Methionyl-tRNA formyltransferase	2
52	COG0231	Translation elongation factor P (EF-P)/translation initiation factor 5A (eIF-5A)	2
53	COG0235	Ribulose-5-phosphate 4-epimerase and related epimerases and aldolases	6
54	COG0243	Anaerobic dehydrogenases	10
55	COG0246	Mannitol-1-phosphate/altronate dehydrogenases	5
56	COG0247	Fe-S oxidoreductase	4
57	COG0248	Exopolyphosphatase	2
58	COG0250	Transcription antiterminator	2
59	COG0251	Putative translation initiation inhibitor	5
60	COG0252	L-asparaginase/archaeal Glu-tRNAGln amidotransferase subunit D	2
61	COG0254	Ribosomal protein L31	2
62	COG0257	Ribosomal protein L36	2
63	COG0258	5'-3' exonuclease (including N-terminal domain of PolI)	2
64	COG0260	Leucyl aminopeptidase	2
65	COG0265	Trypsin-like serine proteases	2
66	COG0266	Formamidopyrimidine-DNA glycosylase	2
67	COG0269	3-hexulose-6-phosphate synthase and related proteins	2
68	COG0272	NAD-dependent DNA ligase (contains BRCT domain type II)	2
69	COG0277	FAD/FMN-containing dehydrogenases	4
70	COG0279	Phosphoheptose isomerase	2
71	COG0280	Phosphotransacetylase	2
72	COG0282	Acetate kinase	2
73	COG0288	Carbonic anhydrase	2
74	COG0298	Hydrogenase maturation factor	2
75	COG0304	3-oxoacyl-(acyl-carrier-protein) synthase	2

76	COG0306	Phosphate/sulphate permeases	2
77	COG0312	Predicted Zn-dependent proteases and their inactivated homologs	2
78	COG0316	Uncharacterized conserved protein	4
79	COG0317	Guanosine polyphosphate pyrophosphohydrolases/synthetases	2
80	COG0318	Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II	4
81	COG0322	Nuclease subunit of the excinuclease complex	2
82	COG0329	Dihydrodipicolinate synthase/N-acetylneuraminate lyase	4
83	COG0330	Membrane protease subunits	2
84	COG0339	Zn-dependent oligopeptidases	2
85	COG0346	Lactoylglutathione lyase and related lyases	2
86	COG0347	Nitrogen regulatory protein PII	2
87	COG0348	Polyferredoxin	2
88	COG0350	Methylated DNA-protein cysteine methyltransferase	2
89	COG0363	6-phosphogluconolactonase/Glucosamine-6-phosphate isomerase/deaminase	2
90	COG0365	Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases	2
91	COG0366	Glycosidases	6
92	COG0371	Glycerol dehydrogenase and related enzymes	2
93	COG0372	Citrate synthase	2
94	COG0374	Ni	2
95	COG0375	Zn finger protein HypA/HybF (possibly regulating hydrogenase expression)	2
96	COG0388	Predicted amidohydrolase	2
97	COG0389	Nucleotidyltransferase/DNA polymerase involved in DNA repair	2
98	COG0394	Protein-tyrosine-phosphatase	2
99	COG0395	ABC-type sugar transport system	2
100	COG0398	Uncharacterized conserved protein	2
101	COG0399	Predicted pyridoxal phosphate-dependent enzyme apparently involved in regulation of cell wall biogenesis	2
102	COG0402	Cytosine deaminase and related metal-dependent hydrolases	4
103	COG0406	Fructose-2	2
104	COG0424	Nucleotide-binding protein implicated in inhibition of septum formation	2

105	COG0425	Predicted redox protein	2
106	COG0431	Predicted flavoprotein	2
107	COG0436	Aspartate/tyrosine/aromatic aminotransferase	2
108	COG0437	Fe-S-cluster-containing hydrogenase components 1	8
109	COG0438	Glycosyltransferase	5
110	COG0440	Acetolactate synthase	2
111	COG0443	Molecular chaperone	4
112	COG0444	ABC-type dipeptide/oligopeptide/nickel transport system	4
113	COG0446	Uncharacterized NAD(FAD)-dependent dehydrogenases	2
114	COG0451	Nucleoside-diphosphate-sugar epimerases	4
115	COG0454	Histone acetyltransferase HPA2 and related acetyltransferases	7
116	COG0456	Acetyltransferases	2
117	COG0457	FOG: TPR repeat	4
118	COG0460	Homoserine dehydrogenase	2
119	COG0463	Glycosyltransferases involved in cell wall biogenesis	5
120	COG0469	Pyruvate kinase	2
121	COG0471	Di- and tricarboxylate transporters	5
122	COG0472	UDP-N-acetylmuramyl pentapeptide phosphotransferase/UDP-N-acetylglucosamine-1-phosphate transferase	2
123	COG0473	Isocitrate/isopropylmalate dehydrogenase	2
124	COG0475	Kef-type K+ transport systems	2
125	COG0476	Dinucleotide-utilizing enzymes involved in molybdopterin and thiamine biosynthesis family 2	2
126	COG0477	Permeases of the major facilitator superfamily	4
127	COG0479	Succinate dehydrogenase/fumarate reductase	2
128	COG0488	ATPase components of ABC transporters with duplicated ATPase domains	4
129	COG0489	ATPases involved in chromosome partitioning	2
130	COG0491	Zn-dependent hydrolases	2
131	COG0493	NADPH-dependent glutamate synthase beta chain and related oxidoreductases	5
132	COG0494	NTP pyrophosphohydrolases including oxidative damage repair enzymes	9

133	COG0500	SAM-dependent methyltransferases	9
134	COG0501	Zn-dependent protease with chaperone function	2
135	COG0503	Adenine/guanine phosphoribosyltransferases and related PRPP-binding proteins	2
136	COG0508	Pyruvate/2-oxoglutarate dehydrogenase complex	2
137	COG0512	Anthranilate/para-aminobenzoate synthases component II	2
138	COG0513	Superfamily II DNA and RNA helicases	5
139	COG0516	IMP dehydrogenase/GMP reductase	2
140	COG0517	FOG: CBS domain	2
141	COG0520	Selenocysteine lyase	2
142	COG0521	Molybdopterin biosynthesis enzymes	2
143	COG0523	Putative GTPases (G3E family)	2
144	COG0524	Sugar kinases	8
145	COG0526	Thiol-disulfide isomerase and thioredoxins	4
146	COG0531	Amino acid transporters	12
147	COG0534	Na+-driven multidrug efflux pump	2
148	COG0542	ATPases with chaperone activity	2
149	COG0545	FKBP-type peptidyl-prolyl cis-trans isomerases 1	2
150	COG0548	Acetylglutamate kinase	2
151	COG0549	Carbamate kinase	2
152	COG0550	Topoisomerase IA	2
153	COG0558	Phosphatidylglycerophosphate synthase	2
154	COG0560	Phosphoserine phosphatase	2
155	COG0561	Predicted hydrolases of the HAD superfamily	6
156	COG0564	Pseudouridylate synthases	4
157	COG0565	rRNA methylase	2
158	COG0566	rRNA methylases	2
159	COG0568	DNA-directed RNA polymerase	2
160	COG0578	Glycerol-3-phosphate dehydrogenase	2
161	COG0579	Predicted dehydrogenase	2
162	COG0580	Glycerol uptake facilitator and related permeases (Major Intrinsic Protein Family)	2

163	COG0582	Integrase	9
164	COG0583	Transcriptional regulator	5
165	COG0584	Glycerophosphoryl diester phosphodiesterase	2
166	COG0586	Uncharacterized membrane-associated protein	5
167	COG0589	Universal stress protein UspA and related nucleotide-binding proteins	6
168	COG0593	ATPase involved in DNA replication initiation	2
169	COG0596	Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)	5
170	COG0598	Mg2+ and Co2+ transporters	2
171	COG0600	ABC-type nitrate/sulfonate/bicarbonate transport system	2
172	COG0601	ABC-type dipeptide/oligopeptide/nickel transport systems	5
173	COG0602	Organic radical activating enzymes	2
174	COG0604	NADPH:quinone reductase and related Zn-dependent oxidoreductases	2
175	COG0605	Superoxide dismutase	2
176	COG0607	Rhodanese-related sulfurtransferase	4
177	COG0609	ABC-type Fe3+-siderophore transport system	4
178	COG0612	Predicted Zn-dependent peptidases	2
179	COG0614	ABC-type Fe3+-hydroxamate transport system	2
180	COG0616	Periplasmic serine proteases (ClpP class)	2
181	COG0617	tRNA nucleotidyltransferase/poly(A) polymerase	2
182	COG0621	2-methylthioadenine synthetase	2
183	COG0624	Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase and related deacylases	4
184	COG0625	Glutathione S-transferase	8
185	COG0626	Cystathionine beta-lyases/cystathionine gamma-synthases	2
186	COG0627	Predicted esterase	2
187	COG0628	Predicted permease	4
188	COG0633	Ferredoxin	2
189	COG0635	Coproporphyrinogen III oxidase and related Fe-S oxidoreductases	2
190	COG0637	Predicted phosphatase/phosphohexomutase	5
191	COG0639	Diadenosine tetraphosphatase and related serine/threonine protein phosphatases	2

192	COG0640	Predicted transcriptional regulators	2
193	COG0641	Arylsulfatase regulator (Fe-S oxidoreductase)	2
194	COG0642	Signal transduction histidine kinase	7
195	COG0644	Dehydrogenases (flavoproteins)	4
196	COG0650	Formate hydrogenlyase subunit 4	2
197	COG0651	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter	2
198	COG0652	Peptidyl-prolyl cis-trans isomerase (rotamase) - cyclophilin family	2
199	COG0654	2-polyprenyl-6-methoxyphenol hydroxylase and related FAD-dependent oxidoreductases	4
200	COG0656	Aldo/keto reductases	2
201	COG0663	Carbonic anhydrases/acetyltransferases	2
202	COG0664	cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases	2
203	COG0665	Glycine/D-amino acid oxidases (deaminating)	2
204	COG0666	FOG: Ankyrin repeat	2
205	COG0667	Predicted oxidoreductases (related to aryl-alcohol dehydrogenases)	4
206	COG0668	Small-conductance mechanosensitive channel	4
207	COG0670	Integral membrane protein	2
208	COG0671	Membrane-associated phospholipid phosphatase	4
209	COG0673	Predicted dehydrogenases and related proteins	6
210	COG0680	Ni	2
211	COG0683	ABC-type branched-chain amino acid transport systems	2
212	COG0687	Spermidine/putrescine-binding periplasmic protein	2
213	COG0693	Putative intracellular protease/amidase	2
214	COG0695	Glutaredoxin and related proteins	2
215	COG0697	Permeases of the drug/metabolite transporter (DMT) superfamily	8
216	COG0702	Predicted nucleoside-diphosphate-sugar epimerases	4
217	COG0703	Shikimate kinase	2
218	COG0714	MoxR-like ATPases	2
219	COG0716	Flavodoxins	4

220	COG0719	ABC-type transport system involved in Fe-S cluster assembly	2
221	COG0722	3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthase	2
222	COG0726	Predicted xylanase/chitin deacetylase	2
223	COG0727	Predicted Fe-S-cluster oxidoreductase	2
224	COG0735	Fe2+/Zn2+ uptake regulation proteins	2
225	COG0737	5'-nucleotidase/2'	2
226	COG0739	Membrane proteins related to metalloendopeptidases	2
227	COG0741	Soluble lytic murein transglycosylase and related regulatory proteins (some contain LysM/invasin domains)	4
228	COG0744	Membrane carboxypeptidase (penicillin-binding protein)	2
229	COG0745	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	10
230	COG0747	ABC-type dipeptide transport system	4
231	COG0754	Glutathionylspermidine synthase	2
232	COG0760	Parvulin-like peptidyl-prolyl isomerase	2
233	COG0764	3-hydroxymyristoyl/3-hydroxydecanoyl-(acyl carrier protein) dehydratases	2
234	COG0765	ABC-type amino acid transport system	5
235	COG0768	Cell division protein FtsI/penicillin-binding protein 2	2
236	COG0772	Bacterial cell division membrane protein	2
237	COG0773	UDP-N-acetylmuramate-alanine ligase	2
238	COG0775	Nucleoside phosphorylase	2
239	COG0776	Bacterial nucleoid DNA-binding protein	4
240	COG0778	Nitroreductase	4
241	COG0782	Transcription elongation factor	2
242	COG0784	FOG: CheY-like receiver	5
243	COG0787	Alanine racemase	2
244	COG0789	Predicted transcriptional regulators	5
245	COG0790	FOG: TPR repeat	2
246	COG0791	Cell wall-associated hydrolases (invasion-associated proteins)	4
247	COG0795	Predicted permeases	2

248	COG0800	2-keto-3-deoxy-6-phosphogluconate aldolase	2
249	COG0811	Biopolymer transport proteins	2
250	COG0814	Amino acid permeases	8
251	COG0824	Predicted thioesterase	2
252	COG0826	Collagenase and related proteases	4
253	COG0833	Amino acid transporters	2
254	COG0834	ABC-type amino acid transport/signal transduction systems	7
255	COG0840	Methyl-accepting chemotaxis protein	5
256	COG0841	Cation/multidrug efflux pump	6
257	COG0842	ABC-type multidrug transport system	5
258	COG0847	DNA polymerase III	2
259	COG0848	Biopolymer transport protein	2
260	COG0852	NADH:ubiquinone oxidoreductase 27 kD subunit	2
261	COG0859	ADP-heptose:LPS heptosyltransferase	4
262	COG0860	N-acetylmuramoyl-L-alanine amidase	2
263	COG0861	Membrane protein TerC	4
264	COG1009	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na+/H+ antiporter	2
265	COG1011	Predicted hydrolase (HAD superfamily)	4
266	COG1012	NAD-dependent aldehyde dehydrogenases	11
267	COG1015	Phosphopentomutase	2
268	COG1018	Flavodoxin reductases (ferredoxin-NADPH reductases) family 1	4
269	COG1024	Enoyl-CoA hydratase/carnithine racemase	6
270	COG1028	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	7
271	COG1045	Serine acetyltransferase	2
272	COG1047	FKBP-type peptidyl-prolyl cis-trans isomerases 2	2
273	COG1048	Aconitase A	2
274	COG1052	Lactate dehydrogenase and related dehydrogenases	2
275	COG1053	Succinate dehydrogenase/fumarate reductase	2

276	COG1063	Threonine dehydrogenase and related Zn-dependent dehydrogenases	11
277	COG1064	Zn-dependent alcohol dehydrogenases	2
278	COG1070	Sugar (pentulose and hexulose) kinases	6
279	COG1072	Panthothenate kinase	2
280	COG1073	Hydrolases of the alpha/beta superfamily	4
281	COG1076	DnaJ-domain-containing proteins 1	2
282	COG1080	Phosphoenolpyruvate-protein kinase (PTS system EI component in bacteria)	4
283	COG1082	Sugar phosphate isomerases/epimerases	2
284	COG1088	dTDP-D-glucose 4	2
285	COG1104	Cysteine sulfinate desulfinase/cysteine desulfurase and related enzymes	2
286	COG1105	Fructose-1-phosphate kinase and related fructose-6-phosphate kinase (PfkB)	2
287	COG1109	Phosphomannomutase	2
288	COG1113	Gamma-aminobutyrate permease and related permeases	7
289	COG1120	ABC-type cobalamin/Fe3+-siderophores transport systems	2
290	COG1124	ABC-type dipeptide/oligopeptide/nickel transport system	2
291	COG1126	ABC-type polar amino acid transport system	4
292	COG1129	ABC-type sugar transport system	8
293	COG1131	ABC-type multidrug transport system	2
294	COG1132	ABC-type multidrug transport system	2
295	COG1138	Cytochrome c biogenesis factor	2
296	COG1140	Nitrate reductase beta subunit	2
297	COG1142	Fe-S-cluster-containing hydrogenase components 2	5
298	COG1143	Formate hydrogenlyase subunit 6/NADH:ubiquinone oxidoreductase 23 kD subunit (chain I)	2
299	COG1145	Ferredoxin	2
300	COG1167	Transcriptional regulators containing a DNA-binding HTH domain and an aminotransferase domain (MocR family) and their eukaryotic orthologs	2
301	COG1169	Isochorismate synthase	2
302	COG1171	Threonine dehydratase	2
303	COG1172	Ribose/xylose/arabinose/galactoside ABC-type transport systems	5

304	COG1173	ABC-type dipeptide/oligopeptide/nickel transport systems	5
305	COG1174	ABC-type proline/glycine betaine transport systems	2
306	COG1175	ABC-type sugar transport systems	2
307	COG1176	ABC-type spermidine/putrescine transport system	2
308	COG1177	ABC-type spermidine/putrescine transport system	2
309	COG1178	ABC-type Fe3+ transport system	2
310	COG1180	Pyruvate-formate lyase-activating enzyme	4
311	COG1181	D-alanine-D-alanine ligase and related ATP-grasp enzymes	2
312	COG1186	Protein chain release factor B	2
313	COG1187	16S rRNA uridine-516 pseudouridylate synthase and related pseudouridylate synthases	4
314	COG1190	Lysyl-tRNA synthetase (class II)	2
315	COG1199	Rad3-related DNA helicases	2
316	COG1209	dTDP-glucose pyrophosphorylase	2
317	COG1210	UDP-glucose pyrophosphorylase	2
318	COG1215	Glycosyltransferases	2
319	COG1221	Transcriptional regulators containing an AAA-type ATPase domain and a DNA-binding domain	2
320	COG1226	Kef-type K+ transport systems	2
321	COG1249	Pyruvate/2-oxoglutarate dehydrogenase complex	4
322	COG1251	NAD(P)H-nitrite reductase	2
323	COG1263	Phosphotransferase system IIC components	8
324	COG1264	Phosphotransferase system IIB components	7
325	COG1271	Cytochrome bd-type quinol oxidase	2
326	COG1278	Cold shock proteins	8
327	COG1280	Putative threonine efflux protein	5
328	COG1289	Predicted membrane protein	6
329	COG1292	Choline-glycine betaine transporter	2
330	COG1294	Cytochrome bd-type quinol oxidase	2
331	COG1295	Predicted membrane protein	2

332	COG1299	Phosphotransferase system	5
333	COG1301	Na+/H+-dicarboxylate symporters	2
334	COG1309	Transcriptional regulator	7
335	COG1319	Aerobic-type carbon monoxide dehydrogenase	2
336	COG1335	Amidases related to nicotinamidase	4
337	COG1344	Flagellin and related hook-associated proteins	2
338	COG1349	Transcriptional regulators of sugar metabolism	11
339	COG1357	Uncharacterized low-complexity proteins	2
340	COG1359	Uncharacterized conserved protein	2
341	COG1363	Cellulase M and related proteins	2
342	COG1376	Uncharacterized protein conserved in bacteria	4
343	COG1393	Arsenate reductase and related proteins	4
344	COG1396	Predicted transcriptional regulators	4
345	COG1414	Transcriptional regulator	7
346	COG1434	Uncharacterized conserved protein	2
347	COG1442	Lipopolysaccharide biosynthesis proteins	2
348	COG1445	Phosphotransferase system fructose-specific component IIB	2
349	COG1448	Aspartate/tyrosine/aromatic aminotransferase	2
350	COG1454	Alcohol dehydrogenase	2
351	COG1459	Type II secretory pathway	2
352	COG1464	ABC-type metal ion transport system	2
353	COG1472	Beta-glucosidase-related glycosidases	2
354	COG1473	Metal-dependent amidase/aminoacylase/carboxypeptidase	2
355	COG1475	Predicted transcriptional regulators	2
356	COG1486	Alpha-galactosidases/6-phospho-beta-glucosidases	2
357	COG1494	Fructose-1	2
358	COG1501	Alpha-glucosidases	2
359	COG1502	Phosphatidylserine/phosphatidylglycerophosphate/cardiolipin synthases and related enzymes	4
360	COG1522	Transcriptional regulators	2

361	COG1528	Ferritin-like protein	2
362	COG1529	Aerobic-type carbon monoxide dehydrogenase	2
363	COG1530	Ribonucleases G and E	2
364	COG1538	Outer membrane protein	4
365	COG1539	Dihydroneopterin aldolase	2
366	COG1544	Ribosome-associated protein Y (PSrp-1)	2
367	COG1546	Uncharacterized protein (competence- and mitomycin-induced)	2
368	COG1553	Uncharacterized conserved protein involved in intracellular sulfur reduction	2
369	COG1560	Lauroyl/myristoyl acyltransferase	2
370	COG1566	Multidrug resistance efflux pump	7
371	COG1595	DNA-directed RNA polymerase specialized sigma subunit	2
372	COG1596	Periplasmic protein involved in polysaccharide export	2
373	COG1605	Chorismate mutase	2
374	COG1609	Transcriptional regulators	12
375	COG1620	L-lactate permease	2
376	COG1626	Neutral trehalase	2
377	COG1629	Outer membrane receptor proteins	3
378	COG1643	HrpA-like helicases	2
379	COG1651	Protein-disulfide isomerase	2
380	COG1653	ABC-type sugar transport system	2
381	COG1662	Transposase and inactivated derivatives	6
382	COG1670	Acetyltransferases	3
383	COG1680	Beta-lactamase class C and other penicillin binding proteins	3
384	COG1686	D-alanyl-D-alanine carboxypeptidase	4
385	COG1690	Uncharacterized conserved protein	2
386	COG1702	Phosphate starvation-inducible protein PhoH	2
387	COG1734	DnaK suppressor protein	2
388	COG1737	Transcriptional regulators	4
389	COG1740	Ni	2
390	COG1741	Pirin-related protein	2

391	COG1760	L-serine deaminase	3
392	COG1762	Phosphotransferase system mannitol/fructose-specific IIA domain (Ntr-type)	6
393	COG1802	Transcriptional regulators	4
394	COG1804	Predicted acyl-CoA transferases/carnitine dehydratase	3
395	COG1820	N-acetylglucosamine-6-phosphate deacetylase	2
396	COG1826	Sec-independent protein secretion pathway components	3
397	COG1830	DhnA-type fructose-1	2
398	COG1838	Tartrate dehydratase beta subunit/Fumarate hydratase class I	3
399	COG1842	Phage shock protein A (IM30)	2
400	COG1846	Transcriptional regulators	3
401	COG1853	Conserved protein/domain typically associated with flavoprotein oxygenases	2
402	COG1879	ABC-type sugar transport system	6
403	COG1881	Phospholipid-binding protein	2
404	COG1882	Pyruvate-formate lyase	4
405	COG1896	Predicted hydrolases of HD superfamily	2
406	COG1925	Phosphotransferase system	3
407	COG1929	Glycerate kinase	2
408	COG1937	Uncharacterized protein conserved in bacteria	2
409	COG1940	Transcriptional regulator/sugar kinase	7
410	COG1957	Inosine-uridine nucleoside N-ribohydrolase	3
411	COG1959	Predicted transcriptional regulator	2
412	COG1960	Acyl-CoA dehydrogenases	4
413	COG1961	Site-specific recombinases	5
414	COG1966	Carbon starvation protein	2
415	COG1972	Nucleoside permease	3
416	COG1974	SOS-response transcriptional repressors (RecA-mediated autopeptidases)	3
417	COG1975	Xanthine and CO dehydrogenases maturation factor	2
418	COG1982	Arginine/lysine/ornithine decarboxylases	5
419	COG1985	Pyrimidine reductase	2
420	COG1988	Predicted membrane-bound metal-dependent hydrolases	2

421	COG1989	Type II secretory pathway	2
422	COG2003	DNA repair proteins	4
423	COG2017	Galactose mutarotase and related enzymes	3
424	COG2025	Electron transfer flavoprotein	3
425	COG2050	Uncharacterized protein	4
426	COG2055	Malate/L-lactate dehydrogenases	3
427	COG2066	Glutaminase	2
428	COG2076	Membrane transporters of cations and cationic drugs	4
429	COG2080	Aerobic-type carbon monoxide dehydrogenase	2
430	COG2084	3-hydroxyisobutyrate dehydrogenase and related beta-hydroxyacid dehydrogenases	4
431	COG2086	Electron transfer flavoprotein	3
432	COG2091	Phosphopantetheinyl transferase	2
433	COG2095	Multiple antibiotic transporter	3
434	COG2116	Formate/nitrite family of transporters	4
435	COG2128	Uncharacterized conserved protein	2
436	COG2132	Putative multicopper oxidases	2
437	COG2141	Coenzyme F420-dependent N5	3
438	COG2146	Ferredoxin subunits of nitrite reductase and ring-hydroxylating dioxygenases	2
439	COG2161	Antitoxin of toxin-antitoxin stability system	2
440	COG2165	Type II secretory pathway	4
441	COG2166	SufE protein probably involved in Fe-S center assembly	2
442	COG2180	Nitrate reductase delta subunit	2
443	COG2181	Nitrate reductase gamma subunit	2
444	COG2186	Transcriptional regulators	9
445	COG2188	Transcriptional regulators	6
446	COG2190	Phosphotransferase system IIA components	2
447	COG2194	Predicted membrane-associated	5
448	COG2195	Di- and tripeptidases	2
449	COG2197	Response regulator containing a CheY-like receiver domain and an HTH DNA-	7

		binding domain	
450	COG2199	FOG: GGDEF domain	8
451	COG2200	FOG: EAL domain	4
452	COG2204	Response regulator containing CheY-like receiver	4
453	COG2207	AraC-type DNA-binding domain-containing proteins	6
454	COG2211	Na+/melibiose symporter and related transporters	6
455	COG2213	Phosphotransferase system	2
456	COG2217	Cation transport ATPase	2
457	COG2222	Predicted phosphosugar isomerases	2
458	COG2223	Nitrate/nitrite transporter	2
459	COG2225	Malate synthase	2
460	COG2233	Xanthine/uracil permeases	6
461	COG2234	Predicted aminopeptidases	2
462	COG2240	Pyridoxal/pyridoxine/pyridoxamine kinase	2
463	COG2244	Membrane protein involved in the export of O-antigen and teichoic acid	4
464	COG2249	Putative NADPH-quinone reductase (modulator of drug activity B)	4
465	COG2252	Permeases	4
466	COG2259	Predicted membrane protein	2
467	COG2261	Predicted membrane protein	2
468	COG2265	SAM-dependent methyltransferases related to tRNA (uracil-5-)-methyltransferase	3
469	COG2267	Lysophospholipase	2
470	COG2271	Sugar phosphate permease	3
471	COG2315	Uncharacterized protein conserved in bacteria	2
472	COG2337	Growth inhibitor	2
473	COG2376	Dihydroxyacetone kinase	2
474	COG2378	Predicted transcriptional regulator	2
475	COG2382	Enterochelin esterase and related enzymes	2
476	COG2390	Transcriptional regulator	2
477	COG2391	Predicted transporter component	2
478	COG2440	Ferredoxin-like protein	3

479	COG2606	Uncharacterized conserved protein	2
480	COG2610	H+/gluconate symporter and related permeases	7
481	COG2704	Anaerobic C4-dicarboxylate transporter	2
482	COG2721	Altronate dehydratase	2
483	COG2723	Beta-glucosidase/6-phospho-beta-glucosidase/beta-galactosidase	3
484	COG2731	Beta-galactosidase	4
485	COG2771	DNA-binding HTH domain-containing proteins	4
486	COG2801	Transposase and inactivated derivatives	7
487	COG2804	Type II secretory pathway	2
488	COG2807	Cyanate permease	2
489	COG2813	16S RNA G1207 methylase RsmC	2
490	COG2814	Arabinose efflux permease	4
491	COG2823	Predicted periplasmic or secreted lipoprotein	2
492	COG2826	Transposase and inactivated derivatives	3
493	COG2837	Predicted iron-dependent peroxidase	2
494	COG2840	Uncharacterized protein conserved in bacteria	2
495	COG2860	Predicted membrane protein	2
496	COG2864	Cytochrome b subunit of formate dehydrogenase	2
497	COG2879	Uncharacterized small protein	2
498	COG2885	Outer membrane protein and related peptidoglycan-associated (lipo)proteins	4
499	COG2890	Methylase of polypeptide chain release factors	2
500	COG2893	Phosphotransferase system	2
501	COG2897	Rhodanese-related sulfurtransferase	2
502	COG2916	DNA-binding protein H-NS	2
503	COG2949	Uncharacterized membrane protein	2
504	COG2963	Transposase and inactivated derivatives	17
505	COG2982	Uncharacterized protein involved in outer membrane biogenesis	2
506	COG2985	Predicted permease	2
507	COG2995	Uncharacterized paraquat-inducible protein A	2
508	COG3005	Nitrate/TMAO reductases	3

509	COG3008	Paraquat-inducible protein B	2
510	COG3023	Negative regulator of beta-lactamase expression	2
511	COG3031	Type II secretory pathway	2
512	COG3038	Cytochrome B561	3
513	COG3039	Transposase and inactivated derivatives	18
514	COG3042	Putative hemolysin	2
515	COG3065	Starvation-inducible outer membrane lipoprotein	2
516	COG3069	C4-dicarboxylate transporter	2
517	COG3077	DNA-damage-inducible protein J	2
518	COG3088	Uncharacterized protein involved in biosynthesis of c-type cytochromes	2
519	COG3093	Plasmid maintenance system antidote protein	2
520	COG3104	Dipeptide/tripeptide permease	4
521	COG3111	Uncharacterized conserved protein	2
522	COG3119	Arylsulfatase A and related enzymes	3
523	COG3121	P pilus assembly protein	4
524	COG3131	Periplasmic glucans biosynthesis protein	2
525	COG3150	Predicted esterase	2
526	COG3152	Predicted membrane protein	3
527	COG3153	Predicted acetyltransferase	2
528	COG3188	P pilus assembly protein	9
529	COG3203	Outer membrane protein (porin)	4
530	COG3206	Uncharacterized protein involved in exopolysaccharide biosynthesis	2
531	COG3209	Rhs family protein	7
532	COG3250	Beta-galactosidase/beta-glucuronidase	3
533	COG3260	Ni	2
534	COG3261	Ni	2
535	COG3262	Ni, Fe hydrogenase component III	2
536	COG3264	Small-conductance mechanosensitive channel	2
537	COG3265	Gluconate kinase	2
538	COG3275	Putative regulator of cell autolysis	2

539	COG3279	Response regulator of the LytR/AlgR family	2
540	COG3290	Signal transduction histidine kinase regulating citrate/malate metabolism	2
541	COG3302	DMSO reductase anchor subunit	2
542	COG3381	Uncharacterized component of anaerobic dehydrogenases	3
543	COG3385	FOG: Transposase and inactivated derivatives	4
544	COG3396	Uncharacterized conserved protein	2
545	COG3414	Phosphotransferase system	3
546	COG3444	Phosphotransferase system	2
547	COG3449	DNA gyrase inhibitor	2
548	COG3468	Type V secretory pathway	4
549	COG3539	P pilus assembly protein	4
550	COG3596	Predicted GTPase	2
551	COG3604	Transcriptional regulator containing GAF	3
552	COG3622	Hydroxypyruvate isomerase	2
553	COG3623	Putative L-xylulose-5-phosphate 3-epimerase	2
554	COG3677	Transposase and inactivated derivatives	7
555	COG3678	P pilus assembly/Cpx signaling pathway	3
556	COG3685	Uncharacterized protein conserved in bacteria	2
557	COG3691	Uncharacterized protein conserved in bacteria	2
558	COG3710	DNA-binding winged-HTH domains	2
559	COG3711	Transcriptional antiterminator	2
560	COG3713	Outer membrane protein V	2
561	COG3715	Phosphotransferase system	2
562	COG3716	Phosphotransferase system	2
563	COG3722	Transcriptional regulator	2
564	COG3729	General stress protein	2
565	COG3765	Chain length determinant protein	3
566	COG3772	Phage-related lysozyme (muraminidase)	2
567	COG3775	Phosphotransferase system	2
568	COG3836	4-dihydroxyhept-2-ene-1	2

569	COG3839	ABC-type sugar transport systems	3
570	COG3842	ABC-type spermidine/putrescine transport systems	4
571	COG4160	ABC-type arginine/histidine transport system	2
572	COG4166	ABC-type oligopeptide transport system	5
573	COG4215	ABC-type arginine transport system	2
574	COG4220	Phage DNA packaging protein	2
575	COG4238	Murein lipoprotein	2
576	COG4533	ABC-type uncharacterized transport system	2
577	COG4565	Response regulator of citrate/malate metabolism	2
578	COG4573	Predicted tagatose 6-phosphate kinase	2
579	COG4575	Uncharacterized conserved protein	3
580	COG4577	Carbon dioxide concentrating mechanism/carboxysome shell protein	2
581	COG4580	Maltoporin (phage lambda and maltose receptor)	2
582	COG4591	ABC-type transport system	2
583	COG4608	ABC-type oligopeptide transport system	2
584	COG4638	Phenylpropionate dioxygenase and related ring-hydroxylating dioxygenases	2
585	COG4668	Mannitol/fructose-specific phosphotransferase system	2
586	COG4682	Predicted membrane protein	2
587	COG4771	Outer membrane receptor for ferrienterochelin and colicins	2
588	COG4795	Type II secretory pathway	2
589	COG4886	Leucine-rich repeat (LRR) protein	2
590	COG4917	Ethanolamine utilization protein	2
591	COG4943	Predicted signal transduction protein containing sensor and EAL domains	2
592	COG4948	L-alanine-DL-glutamate epimerase and related enzymes of enolase superfamily	6
593	COG4992	Ornithine/acetylornithine aminotransferase	3
594	COG5013	Nitrate reductase alpha subunit	2
595	COG5339	Uncharacterized protein conserved in bacteria	2
596	COG5433	Transposase	6
597	COG5463	Predicted integral membrane protein	2
598	COG5464	Uncharacterized conserved protein	5

599	COG5562	Phage envelope protein	3
600	COG5645	Predicted periplasmic lipoprotein	2

Appendix Table 2.2 – *Escherichia coli* interactions from literature involving paralogs annotated as 'uncharacterized' or 'predicted'

Bait UniProt	Prey UniProt	Bait COG	Prey COG	Bait Name	Prey Name
P0A8N0	P24242	COG3120	COG1609	MATP	ASCG
P23862	P24242	COG3923	COG1609	PRIC	ASCG
P32712	P00805	COG4235	COG0252	NRFG	ASPG2
P0A805	P00805	COG0233	COG0252	RRF	ASPG2
P76502	P00805	COG2062	COG0252	SIXA	ASPG2
P0A9Q5	P0AC02	COG0777	COG4105	ACCD	BAMD
P27254	P0A6E9	COG1703	COG0132	ARGK	BIOD2
P19624	P0A6E9	COG1995	COG0132	PDXA	BIOD2
P76407	P0ABH7	COG1597	COG0372	YEGS	CISY
P00861	P0AE88	COG0019	COG0745	DCDA	CPXR
P0AGI4	P08368	COG4214	COG0745	XYLH	CREB
P19925	P0ACJ8	COG2977	COG0664	ENTD	CRP
P0ACG8	P0ACJ8	COG1188	COG0664	HSLR	CRP
P06993	P0ACJ8	COG2909	COG0664	MALT	CRP
P0A8R0	P0ACJ8	COG0684	COG0664	RRAA	CRP
P76016	P76015	COG3284	COG2376	DHAR	DHAK
Q47149	Q47150	COG3041	COG3077	YAFQ	DINJ
P0AEF0	P0ACB0	COG1484	COG0305	DNAC	DNAB
P10443	P03007	COG0587	COG0847	DPO3A	DPO3E
P06710	P03007	COG2812	COG0847	DPO3X	DPO3E
P28630	P03007	COG1466	COG0847	HOLA	DPO3E
P06710	P06710	COG2812	COG2812	DPO3X	DPO3X
P0A8R0	P06710	COG0684	COG2812	RRAA	DPO3X
P62620	P77188	COG0821	0	ISPG	ECPB
P64564	P76938	COG0762	COG3101	YGGT	EPMC

P19636	P76551	COG4302	COG4819	EUTC	EUTA
P0ABA4	P19636	COG0712	COG4302	ATPD	EUTC
P12996	P19636	COG0502	COG4302	BIOB	EUTC
P0ABU5	P19636	COG3155	COG4302	ELBB	EUTC
P76551	P19636	COG4819	COG4302	EUTA	EUTC
P75825	P19636	COG1151	COG4302	HCP	EUTC
P0ACG8	P19636	COG1188	COG4302	HSLR	EUTC
P06993	P19636	COG2909	COG4302	MALT	EUTC
P0A6Z6	P19636	COG0864	COG4302	NIKR	EUTC
P0C037	P19636	COG3123	COG4302	YAIE	EUTC
P76555	P77218	COG4766	COG0280	EUTQ	EUTD
P52060	P0AEJ8	COG1872	COG4576	YGGU	EUTN
P76555	P76555	COG4766	COG4766	EUTQ	EUTQ
P61889	P76555	COG0039	COG4766	MDH	EUTQ
P0A6W9	P37009	COG2918	COG3842	GSH1	FBPC
P0A8P1	P0A8P3	COG2360	COG2924	LFTR	FETP
P0A9Q5	P39405	COG0777	COG4114	ACCD	FHUF
P27254	P39405	COG1703	COG4114	ARGK	FHUF
P0ABA6	P39405	COG0224	COG4114	ATPG	FHUF
P0A9I1	P39405	COG2301	COG4114	CITE	FHUF
P60664	P39405	COG0107	COG4114	HIS6	FHUF
P0A6Z6	P39405	COG0864	COG4114	NIKR	FHUF
P0ACM9	P39405	COG0096	COG4114	YIHL	FHUF
P77580	P68646	COG4569	COG2440	ACDH	FIXX
P19624	P68646	COG1995	COG2440	PDXA	FIXX
P0AGI4	P68646	COG4214	COG2440	XYLH	FIXX
P0ABA4	P43533	COG0712	COG3418	ATPD	FLGN
P24216	P0ABY2	COG1345	0	FLID	FLIT

P0AAF6	P0AC47	COG4161	COG0479	ARTP	FRDB
P15028	P0ACK8	COG4594	COG1349	FECB	FUCR
P33235	P75913	COG1256	COG0111	FLGK	GHRA
P77580	P0ACL5	COG4569	COG2186	ACDH	GLCC
P0ABA6	P0ACL5	COG0224	COG2186	ATPG	GLCC
P0A6G5	P0ACL5	COG3697	COG2186	CITX	GLCC
P0ADI4	P0ACL5	COG1535	COG2186	ENTB	GLCC
P52613	P0ACL5	COG2882	COG2186	FLIJ	GLCC
P28630	P0ACL5	COG1466	COG2186	HOLA	GLCC
P0ACG8	P0ACL5	COG1188	COG2186	HSLR	GLCC
P0A8N0	P0ACL5	COG3120	COG2186	MATP	GLCC
P30749	P0ACL5	COG0314	COG2186	MOAE	GLCC
P22524	P0ACL5	COG3095	COG2186	MUKE	GLCC
P22634	P0ACL5	COG0796	COG2186	MURI	GLCC
P23862	P0ACL5	COG3923	COG2186	PRIC	GLCC
P37177	P0ACL5	COG3605	COG2186	PT1P	GLCC
P09184	P0ACL5	COG3727	COG2186	VSR	GLCC
P0A8P6	P0ACL5	COG4973	COG2186	XERC	GLCC
P45955	P0ACL5	COG1729	COG2186	YBGF	GLCC
P0A8C4	P0ACL5	COG3079	COG2186	YGFB	GLCC
P0AFB5	P27249	COG3852	COG2844	NTRB	GLND
P27249	P0AC55	COG2844	COG0347	GLND	GLNK
P05847	P0ACL0	COG1951	COG1349	TTDA	GLPR
P30749	P75824	COG0314	COG1018	MOAE	HCR
P0A988	P69931	COG0592	COG0593	DPO3B	HDA
P67701	P64578	COG5499	COG4680	HIGA	HIGB
Q46871	P28630	COG2375	COG1466	YQJH	HOLA
P06710	P28632	COG2812	COG3050	DPO3X	HOLD

P28905	P28632	COG2927	COG3050	HOLC	HOLD
P0ACD4	P0A6L9	COG0822	COG1076	ISCU	HSCB
P21170	P0A6L9	COG1166	COG1076	SPEA	HSCB
P28224	P0ACG8	COG3895	COG1188	MLIC	HSLR
P21170	P0AAK4	COG1166	COG1142	SPEA	HYDN
P11071	P08200	COG4579	COG0538	ACEK	IDH
P0A9I1	P05827	COG2301	COG0583	CITE	ILVY
P00936	P05827	COG3072	COG0583	CYAA	ILVY
P23862	P05827	COG3923	COG0583	PRIC	ILVY
P0A853	P0AEW6	COG3033	COG0524	TNAA	INGK
P17109	P37647	COG1165	COG0524	MEND	KDGK
P21865	P21866	COG2205	COG0745	KDPD	KDPE
P37177	P0A756	COG3605	COG2249	PT1P	KEFG
P23862	P75957	COG3923	COG1136	PRIC	LOLD
P0A6S5	P0A725	COG2919	COG0774	FTSB	LPXC
P0A9I1	P0ACJ0	COG2301	COG1522	CITE	LRP
P14175	P75826	COG4175	COG2431	PROV	LYSO
P06993	P23256	COG2909	COG1168	MALT	MALY
P0A9Q5	P0A8N0	COG0777	COG3120	ACCD	MATP
P15028	P0A8N0	COG4594	COG3120	FECB	MATP
P77390	P69741	COG3053	COG1740	CITC	MBHT
P30131	P30750	COG0068	COG1135	HYPF	METN
P0A9I1	P18196	COG2301	COG0850	CITE	MINC
P28630	P18196	COG1466	COG0850	HOLA	MINC
P0ACG8	P18196	COG1188	COG0850	HSLR	MINC
P30131	P18196	COG0068	COG0850	HYPF	MINC
P0A9I3	P41052	COG2716	COG2951	GCVR	MLTB
P05050	Q46864	COG3145	COG1396	ALKB	MQSA

P0A9I1	Q46864	COG2301	COG1396	CITE	MQSA
P41052	Q46864	COG2951	COG1396	MLTB	MQSA
P15067	P0A744	COG1523	COG0225	GLGX	MSRA
P23305	P22634	COG3159	COG0796	YIGA	MURI
P04951	P67430	COG1212	0	KDSB	NEMR
P28630	P17117	COG1466	COG0778	HOLA	NFSA
P07109	P0A8D0	COG4598	COG1327	HISP	NRDR
P0AFB5	P0AFB8	COG3852	COG2204	NTRB	NTRC
P28630	P0AFD1	COG1466	COG1905	HOLA	NUOE
P33916	P76027	COG4172	COG0444	YEJF	OPPD
P0A9I1	P77467	COG2301	COG1024	CITE	PAAG
P06993	P77467	COG2909	COG1024	MALT	PAAG
P0ABA6	P76086	COG0224	COG3327	ATPG	PAAX
P29131	P76086	COG3087	COG3327	FTSN	PAAX
P0AE08	P28305	COG0450	COG0115	AHPC	PABC
P07001	P07001	COG3288	COG3288	PNTA	PNTA
P10378	P07003	COG1021	COG0028	ENTE	POXB
P12996	P45577	COG0502	COG3109	BIOB	PROQ
P0ACG8	P45577	COG1188	COG3109	HSLR	PROQ
P0A7X6	P45577	COG0806	COG3109	RIMM	PROQ
P46144	P45577	COG1418	COG3109	YEDJ	PROQ
P14175	P14175	COG4175	COG4175	PROV	PROV
P04994	P23857	COG1570	COG0607	EX7L	PSPE
P22634	P0A8A4	COG0796	COG1806	MURI	PSRP
P21865	P69829	COG2205	COG1762	KDPD	PTSN
P16685	P04983	COG3624	COG1129	PHNG	RBSA
P76502	P64534	COG2062	COG5455	SIXA	RCNB
P0ACG8	P64530	COG1188	COG1937	HSLR	RCNR

P06993	P0A7G6	COG2909	COG0468	MALT	RECA
P0A7G6	P33596	COG0468	COG2137	RECA	RECX
P22731	P0A7X6	COG0410	COG0806	LIVF	RIMM
P0ABU5	P0A7N1	COG3155	COG0254	ELBB	RL31B
P15005	P63177	COG1401	COG0566	MCRB	RLMB
P31806	P63177	COG0062	COG0566	NNR	RLMB
P0A8P1	P0AA37	COG2360	COG0564	LFTR	RLUA
P22634	P0AFW4	COG0796	COG0782	MURI	RNK
P24554	P0AFW4	COG1066	COG0782	RADA	RNK
P0A988	P0AG07	COG0592	COG0036	DPO3B	RPE
P76344	P0A776	COG3443	COG0494	ZINT	RPPH
P0A9I3	P0A805	COG2716	COG0233	GCVR	RRF
P0C0L2	P0A805	COG1764	COG0233	OSMC	RRF
P52612	P0A7W5	COG1157	COG0096	FLII	RS8
P04951	P0ACU2	COG1212	COG1309	KDSB	RUTR
P0ACG8	P07014	COG1188	COG0479	HSLR	SDHB
P15067	P0ABW5	COG1523	0	GLGX	SFMA
P05055	P21507	COG1185	COG0513	PNP	SRMB
P36979	P21507	COG0820	COG0513	RLMN	SRMB
P15067	P0AGE0	COG1523	COG0629	GLGX	SSB
P28903	P30138	COG1328	COG0476	NRDD	THIF
P28630	P0AGJ2	COG1466	COG0566	HOLA	TRMH
P05020	P0AGJ2	COG0418	COG0566	PYRC	TRMH
P12996	P00895	COG0502	COG0147	BIOB	TRPE
P51020	P00895	COG0091	COG0147	HOA	TRPE
P26602	P26602	COG3161	COG3161	UBIC	UBIC
P61175	P17993	COG0091	COG2227	RL22	UBIG
P77649	P17993	COG0397	COG2227	YDIU	UBIG

P77390	P25534	COG3053	COG0654	CITC	UBIH
P0A7G6	P0AG11	COG0468	COG1974	RECA	UMUD
P0A887	P0AG11	COG2226	COG1974	UBIE	UMUD
P0ADW6	P08390	COG1242	COG0136	YHCC	USG
P06136	P09184	COG1589	COG3727	FTSQ	VSR
P22634	Q47147	COG0796	COG0121	MURI	YAFJ
P69441	P77806	COG0563	COG0436	KAD	YBDL
P0AC75	P37909	COG1519	0	KDTA	YBGD
P19624	P37909	COG1995	0	PDXA	YBGD
P21170	P76162	COG1166	0	SPEA	YDFU
Q46896	P76272	COG1518	COG3008	CAS1	YEBT
P28630	P64526	COG1466	0	HOLA	YEEW
P0ADI4	P33030	COG1535	COG0523	ENTB	YEIR
P0ACG8	P33030	COG1188	COG0523	HSLR	YEIR
P15067	P76500	COG1523	0	GLGX	YFCQ
P0ABE2	P52102	COG0271	COG1145	BOLA	YFHL
P60664	P52102	COG0107	COG1145	HIS6	YFHL
P0AEZ3	P52102	COG2894	COG1145	MIND	YFHL
Q47690	P39834	COG2040	0	MMUM	YGIL
P76215	P28638	COG2988	COG0863	ASTE	YHDJ
P28630	P0AFU6	COG1466	0	HOLA	YIIF
P0A9I1	Q79E92	COG2301	COG2963	CITE	YKGN
P28630	Q79E92	COG1466	COG2963	HOLA	YKGN
P0ACG8	Q79E92	COG1188	COG2963	HSLR	YKGN
P04951	Q79E92	COG1212	COG2963	KDSB	YKGN
P46144	Q79E92	COG1418	COG2963	YEDJ	YKGN
P15067	P67553	COG1523	0	GLGX	YNFC
P0A6F5	P76524	COG0459	COG0006	CH60	YPDF

P0ADI4	P75862	COG1535	0	ENTB	ZAPC
P26616	P76344	COG0281	COG3443	MAO1	ZINT
P76502	P76344	COG2062	COG3443	SIXA	ZINT
P60785	P64423	COG0481	COG0598	LEPA	ZNTB
Q46896	P76339	COG1518	COG0642	CAS1	YEDV
P0A9D8	P77609	COG2171	0	DAPD	FLXA
P04994	P76168	COG1570	COG0582	EX7L	INTQ
P06136	P76080	COG1589	COG2151	FTSQ	PAAD
P15067	P37909	COG1523	0	GLGX	YBGD
P28630	P75883	COG1466	0	HOLA	GFCC
P28630	P56256	COG1466	COG1142	HOLA	YSAA
P0ACG8	P77174	COG1188	COG1475	HSLR	YBDM
P23930	P77551	COG0815	0	LNT	RZPR
P22634	P39409	COG0796	COG1180	MURI	YJJW
P19624	P39834	COG1995	0	PDXA	YGIL
P0AFM2	P77165	COG2113	COG2080	PROX	YAGT
P0A8R0	P77174	COG0684	COG1475	RRAA	YBDM
P05523	P24242	COG0266	COG1609	FPG	ASCG
P23878	P17444	COG1120	COG2303	FEPC	BETA
P0A7M9	P12996	COG0254	COG0502	RL31	BIOB
P77444	Q46896	COG0520	COG1518	SUFS	CAS1
P0ABH7	P0ABH7	COG0372	COG0372	CISY	CISY
Q46925	Q46925	COG0520	COG0520	CSDA	CSDA
Q46925	P0AGF2	COG0520	COG2166	CSDA	CSDE
P06959	P0A9P0	COG0508	COG1249	ODP2	DLDH
P77806	P10443	COG0436	COG0587	YBDL	DPO3A
P03007	P03007	COG0847	COG0847	DPO3E	DPO3E
P0ADF8	P76938	COG0440	COG3101	ILVN	EPMC

P06134	P0AEJ8	COG0350	COG4576	ADA	EUTN
P0A7M9	P0AEJ8	COG0254	COG4576	RL31	EUTN
P0ABU7	P0ABU7	COG0811	COG0811	EXBB	EXBB
P23878	P23878	COG1120	COG1120	FEPC	FEPC
P77398	P68646	COG0223	COG2440	ARNA	FIXX
P77806	P68646	COG0436	COG2440	YBDL	FIXX
P0A9L3	P0A9L3	COG0545	COG0545	FKBB	FKBB
P0A6J8	P0A9A6	COG1181	COG0206	DDLA	FTSZ
P76213	P0ACL5	COG0322	COG2186	CHO	GLCC
P0AC55	P0AC55	COG0347	COG0347	GLNK	GLNK
P68688	P0AC62	COG0695	COG0695	GLRX1	GLRX3
P06134	P69931	COG0350	COG0593	ADA	HDA
P69931	P69931	COG0593	COG0593	HDA	HDA
P0A9L3	P0AAJ8	COG0545	COG0437	FKBB	HYBA
P0AAM3	P0AAM7	COG0298	COG0298	HYPC	HYBG
P0AAM7	P16431	COG0298	COG3261	HYBG	HYCE
P0A9L3	P0AAK4	COG0545	COG1142	FKBB	HYDN
P0AAM7	P24192	COG0298	COG0409	HYBG	HYPD
P03018	P25665	COG0210	COG0620	UVRD	METE
P16703	P0AAG8	COG0031	COG1129	CYSM	MGLA
Q46864	Q46896	COG1396	COG1518	MQSA	CAS1
P0A6Y8	P36659	COG0443	COG2214	DNAK	CBPA
P25524	P25524	COG0402	COG0402	CODA	CODA
P52106	P52106	COG2771	COG2771	CSGD	CSGD
P0A6L2	P0A6L2	COG0329	COG0329	DAPA	DAPA
P0AFI5	P0A6P5	COG1686	COG1160	PBP7	DER
P76114	P0AC73	COG1802	COG2731	MCBR	EBGC
P0A8Y8	P0ADI4	COG2050	COG1535	ENTH	ENTB

P75728	P76938	COG0654	COG3101	UBIF	EPMC
P24178	P19636	COG1393	COG4302	YFFB	EUTC
P63389	P0AEL3	COG0488	COG1918	YHES	FEOA
P0ACF4	P39405	COG0776	COG4114	DBHB	FHUF
P0AB74	P0AEM6	COG0191	COG1191	KBAY	FLIA
P61949	P0ACL5	COG0716	COG2186	FLAV	GLCC
P0A6Y8	P09372	COG0443	COG0576	DNAK	GRPE
P0A6Y8	P0A6Z3	COG0443	COG0326	DNAK	HTPG
P08142	P08142	COG0028	COG0028	ILVB	ILVB
P00893	P00893	COG0028	COG0028	ILVI	ILVI
P08142	P0ADG1	COG0028	COG3978	ILVB	ILVM
P08142	P0ADF8	COG0028	COG0440	ILVB	ILVN
P00893	Q46864	COG0028	COG1396	ILVI	MQSA
P0A9P6	P0A9P6	COG0513	COG0513	DEAD	DEAD
P0ACS2	P18776	COG0789	COG0437	SOXR	DMSB
P02925	P15042	COG1879	COG0272	RBSB	DNLJ
P39265	P19636	COG1879	COG4302	ALSB	EUTC
P37680	P68646	COG0235	COG2440	SGBE	FIXX
P11553	P11553	COG1070	COG1070	FUCK	FUCK
P37680	P0AAK4	COG0235	COG1142	SGBE	HYDN
P28904	P24218	COG0366	COG0582	TREC	INTD
P37680	P0AAG8	COG0235	COG1129	SGBE	MGLA
P08401	P08368	COG0642	COG0745	CREC	CREB
P0ACT6	P68646	COG1309	COG2440	UIDR	FIXX
P0A968	P0AAG0	COG1278	COG0444	CSPD	DPPD
P45543	P45543	COG0524	COG0524	FRLD	FRLD
P45543	Q46839	COG0524	COG1620	FRLD	GLCA
P30235	P0ACL5	COG0524	COG2186	PSUK	GLCC

P0ACA7	P0ACA7	COG0625	COG0625	GSTB	GSTB
P0ACA7	P06987	COG0625	COG0131	GSTB	HIS7
P0ACA3	P54745	COG0625	COG1299	SSPA	HRSA
P0AEW6	P0AEW6	COG0524	COG0524	INGK	INGK
P31460	P31460	COG2186	COG2186	DGOR	DGOR
P08337	P0AC47	COG0494	COG0479	MUTT	FRDB
P69228	P69228	COG0745	COG0745	BAER	BAER
P08368	P08368	COG0745	COG0745	CREB	CREB
P0ACZ8	P0ACZ8	COG0745	COG0745	CUSR	CUSR
P18775	P0ACL5	COG0243	COG2186	DMSA	GLCC
P07658	P75825	COG0243	COG1151	FDHF	HCP
P24242	P24242	COG1609	COG1609	ASCG	ASCG
P0ACP1	P0ACP1	COG1609	COG1609	CRA	CRA
P0ACN7	P0AEJ8	COG1609	COG4576	CYTR	EUTN
P26266	P76015	COG3765	COG2376	FEPE	DHAK
Q2M7R5	P18776	0	COG0437	YIBT	DMSB
P33129	P15042	COG3188	COG0272	HTRE	DNLJ
P11989	P06710	COG3711	COG2812	BGLG	DPO3X
P0AB33	P19636	0	COG4302	BSSS	EUTC
P0AEF4	P19636	COG4565	COG4302	DPIA	EUTC
P0AAY6	P19636	0	COG4302	YBJN	EUTC
P76275	P19636	0	COG4302	YEBW	EUTC
P32162	P0AEJ8	COG3691	COG4576	YIIS	EUTN
P32700	P0AEJ8	0	COG4576	YJCB	EUTN
P39325	P0AEJ8	0	COG4576	YTFQ	EUTN
P76334	P63746	0	COG4810	YEDR	EUTS
P03030	P39405	0	COG4114	LYSR	FHUF
P0ACL9	P39405	COG2186	COG4114	PDHR	FHUF

Q47274	P39405	0	COG4114	REQ1	FHUF
P31063	P39405	0	COG4114	YEDD	FHUF
P03014	P68646	COG1961	COG2440	PINE	FIXX
P06864	P0A6S3	COG3250	COG1706	BGA2	FLGI
P0AGB6	P52612	COG1595	COG1157	RPOE	FLII
Q46864	P0A9E5	COG1396	COG0664	MQSA	FNR
P0AC33	P0AC33	COG1838	COG1838	FUMA	FUMA
P03024	P03024	COG1609	COG1609	GALR	GALR
P03024	P25748	COG1609	COG1609	GALR	GALS
P0ACN7	P0ACL5	COG1609	COG2186	CYTR	GLCC
Q46864	P0ACL5	COG1396	COG2186	MQSA	GLCC
P69816	P0ACL5	COG1445	COG2186	PTFB2	GLCC
P69829	P0ACL5	COG1762	COG2186	PTSN	GLCC
P0AGB6	P0ACL5	COG1595	COG2186	RPOE	GLCC
P0ACU2	P0ACL5	COG1309	COG2186	RUTR	GLCC
P36673	P0ACL5	COG1609	COG2186	TRER	GLCC
P37671	P0ACL5	COG1414	COG2186	YIAJ	GLCC
P37902	P37902	COG0834	COG0834	GLTI	GLTI
P37902	P0AAG3	COG0834	COG1126	GLTI	GLTL
P77399	P77293	COG1024	COG0463	FADJ	GTRB
P0AAK4	P75824	COG1142	COG1018	HYDN	HCR
P0AFI5	P69931	COG1686	COG0593	PBP7	HDA
P32677	P77319	0	COG0443	YIJO	HSCC
P19930	P19930	COG0680	COG0680	HYAD	HYAD
P31666	P05827	COG0726	COG0583	YADE	ILVY
P06864	P32053	COG3250	COG0582	BGA2	INTA
P0AGB6	P18811	COG1595	COG1609	RPOE	MALI
P04983	P06993	COG1129	COG2909	RBSA	MALT

P19930	P0ACD8	COG0680	COG0374	HYAD	MBHL
P19930	P69739	COG0680	COG1740	HYAD	MBHS
P45543	P05704	COG0524	COG0840	FRLD	MCP3
P68688	P30750	COG0695	COG1135	GLRX1	METN
P30147	P30750	COG3622	COG1135	HYI	METN
P0AEI6	P30750	COG0494	COG1135	NUDJ	METN
P37759	P30750	COG1088	COG1135	RMLB1	METN
P77689	P30750	COG0719	COG1135	SUFD	METN
P77444	P30750	COG0520	COG1135	SUFS	METN
P77585	P30750	COG1363	COG1135	YPDE	METN
P76540	P0AAG8	COG4577	COG1129	EUTK	MGLA
P71229	P0AAG8	COG3604	COG1129	HYFR	MGLA
P52062	P0AAG8	COG0635	COG1129	YGGW	MGLA
P11989	P41052	COG3711	COG2951	BGLG	MLTB
P75883	P41052	0	COG2951	GFCC	MLTB
P39363	P41052	0	COG2951	SGCA	MLTB
P0ABH9	Q46864	COG0542	COG1396	CLPA	MQSA
P0ACN7	Q46864	COG1609	COG1396	CYTR	MQSA
P69931	Q46864	COG0593	COG1396	HDA	MQSA
P0AD57	Q46864	COG0142	COG1396	ISPB	MQSA
Q46864	Q46864	COG1396	COG1396	MQSA	MQSA
P0AFI5	Q46864	COG1686	COG1396	PBP7	MQSA
P29745	Q46864	COG2195	COG1396	PEPT	MQSA
P04983	Q46864	COG1129	COG1396	RBSA	MQSA
P0ACQ0	Q46864	COG1609	COG1396	RBSR	MQSA
P0DMC7	Q46864	COG2197	COG1396	RCSB	MQSA
P0AGB6	Q46864	COG1595	COG1396	RPOE	MQSA
P08337	P08337	COG0494	COG0494	MUTT	MUTT

P38489	P38489	COG0778	COG0778	NFSB	NFSB
P76097	P33593	COG5383	COG0444	YDCJ	NIKD
P68688	P0A6Z6	COG0695	COG0864	GLRX1	NIKR
P77585	P0A6Z6	COG1363	COG0864	YPDE	NIKR
P0AC55	P0AFB5	COG0347	COG3852	GLNK	NTRB
P05523	P0AFD1	COG0266	COG1905	FPG	NUOE
P75824	P0AFD1	COG1018	COG1905	HCR	NUOE
P0AFI5	P0AFD1	COG1686	COG1905	PBP7	NUOE
P0AFI0	P0AFI0	COG0028	COG0028	OXC	OXC
P28305	P28305	COG0115	COG0115	PABC	PABC
P32131	P16687	COG0635	COG3626	HEMN	PHNI
P0AFJ5	P0AFJ5	COG0745	COG0745	PHOB	PHOB
P0A7M9	P0AFJ5	COG0254	COG0745	RL31	PHOB
P18775	P07000	COG0243	COG2267	DMSA	PLDB
P0A9L5	P0A9L5	COG0760	COG0760	PPIC	PPIC
P77580	P25889	COG4569	COG0167	ACDH	PREA
P05523	P15373	COG0266	COG2002	FPG	PRLF
P25536	P15373	COG0424	COG2002	YHDE	PRLF
P61949	P45577	COG0716	COG3109	FLAV	PROQ
P23882	P45577	COG0223	COG3109	FMT	PROQ
P0AGA6	P45577	COG2197	COG3109	UHPA	PROQ
P37759	P09546	COG1088	COG0506	RMLB1	PUTA
P28304	P28304	COG0604	COG0604	QOR1	QOR1
P0AEE5	P04983	COG1879	COG1129	DGAL	RBSA
P04983	P04983	COG1129	COG1129	RBSA	RBSA
P0DMC7	P39838	COG2197	COG2198	RCSB	RCSD
P61949	P33596	COG0716	COG2137	FLAV	RECX
P09980	P09980	COG0210	COG0210	REP	REP

P03018	P09980	COG0210	COG0210	UVRD	REP
P37182	P0A7I4	COG0680	COG4108	HYBD	RF3
P0A8J8	P0A8J8	COG0513	COG0513	RHLB	RHLB
P61949	P0A7L8	COG0716	COG0211	FLAV	RL27
P37387	P0A7K2	COG4213	COG0222	XYLF	RL7
P61949	P0A776	COG0716	COG0494	FLAV	RPPH
P61949	P0A7V8	COG0716	COG0522	FLAV	RS4
P37653	P77285	COG1215	COG4659	BCSA	RSXG
P77285	P77285	COG4659	COG4659	RSXG	RSXG
P0AF12	P07026	COG0775	COG2771	MTNN	SDIA
P38506	P07026	COG0258	COG2771	XNI	SDIA
P0A910	P0AG86	COG2885	COG1952	OMPA	SECB
P21513	P21507	COG1530	COG0513	RNE	SRMB
P0ACA3	P0ACA3	COG0625	COG0625	SSPA	SSPA
P02925	P06612	COG1879	COG0550	RBSB	TOP1
P38684	P39453	COG0745	COG0784	TORR	TORS
P04846	P0AGJ2	COG1464	COG0566	NLPA	TRMH
P0AGJ2	P0AGJ2	COG0566	COG0566	TRMH	TRMH
P75728	P75728	COG0654	COG0654	UBIF	UBIF
P77580	P25534	COG4569	COG0654	ACDH	UBIH
P0ACT6	P0ACT6	COG1309	COG1309	UIDR	UIDR
P10908	P0AG11	COG0584	COG1974	UGPQ	UMUD
P0ADK6	P08390	0	COG0136	YIBA	USG
P23843	P03018	COG4166	COG0210	OPPA	UVRD
P0AAF3	P75691	COG1129	COG1064	ARAG	YAHK
P61949	P76135	COG0716	COG2207	FLAV	YDEO
P50457	P33218	COG0160	COG2979	PUUE	YEBE
P77788	P03813	COG0494	COG1794	NUDG	YGEA

P0A8C4	P0AGK4	COG3079	COG1534	YGFB	YHBY
P77580	P28638	COG4569	COG0863	ACDH	YHDJ
P0ABU5	P28638	COG3155	COG0863	ELBB	YHDJ
P37680	P28638	COG0235	COG0863	SGBE	YHDJ
P76010	P28638	COG5581	COG0863	YCGR	YHDJ
P0A8A0	P28638	COG0217	COG0863	YEBC	YHDJ
P76114	P37631	COG1802	COG2081	MCBR	YHIN
P39220	P0AF76	0	COG3789	YABP	YJFI
P05050	Q79E92	COG3145	COG2963	ALKB	YKGN
P77390	Q79E92	COG3053	COG2963	CITC	YKGN
P64423	P64423	COG0598	COG0598	ZNTB	ZNTB
P29131	P39409	COG3087	COG1180	FTSN	YJJW
P06993	P45549	COG2909	COG1015	MALT	YHFW
P06993	P77174	COG2909	COG1475	MALT	YBDM
P06993	P31574	COG2909	COG2025	MALT	FIXB
P39358	P39358	COG0129	COG0129	YJHG	YJHG
P0AFI5	P11071	COG1686	COG4579	PBP7	ACEK
P0ACS9	P0ACS9	COG1309	COG1309	ACRR	ACRR
P33931	P0ACS9	COG4133	COG1309	CCMA	ACRR
Q46864	P0ACS9	COG1396	COG1309	MQSA	ACRR
P0A7W5	P26646	COG0096	COG0604	RS8	ACUI
P0AB71	P0AB71	COG0191	COG0191	ALF	ALF
P0A955	P0A955	COG0800	COG0800	ALKH	ALKH
P06864	P0A955	COG3250	COG0800	BGA2	ALKH
P32718	P32718	COG1940	COG1940	ALSK	ALSK
P32131	P0AD70	COG0635	COG1680	HEMN	AMPH
P42593	P04825	COG0446	COG0308	FADH	AMPN
P0AAF3	P0AAF3	COG1129	COG1129	ARAG	ARAG

P28630	P75993	COG1466	0	HOLA	ARIR

Appendix Table 3.1 - List of OG IDs and description (395) for OGs present in all species being studied (*Eco, Ype, Vch, Sau, Rpr*)

Sr No.	OG ID	OG Desription
1	COG0006	Xaa-Pro aminopeptidase
2	COG0008	Glutamyl- and glutaminyl-tRNA synthetases
3	COG0009	Putative translation factor (SUA5)
4	COG0012	Predicted GTPase
5	COG0013	Alanyl-tRNA synthetase
6	COG0016	Phenylalanyl-tRNA synthetase alpha subunit
7	COG0018	Arginyl-tRNA synthetase
8	COG0020	Undecaprenyl pyrophosphate synthase
9	COG0024	Methionine aminopeptidase
10	COG0030	Dimethyladenosine transferase (rRNA methylation)
11	COG0037	Predicted ATPase of the PP-loop superfamily implicated in cell cycle control
12	COG0039	Malate/lactate dehydrogenases
13	COG0042	tRNA-dihydrouridine synthase
14	COG0045	Succinyl-CoA synthetase
15	COG0048	Ribosomal protein S12
16	COG0049	Ribosomal protein S7
17	COG0050	GTPases - translation elongation factors
18	COG0051	Ribosomal protein S10
19	COG0052	Ribosomal protein S2
20	COG0053	Predicted Co/Zn/Cd cation transporters
21	COG0055	F0F1-type ATP synthase
22	COG0056	F0F1-type ATP synthase
23	COG0060	Isoleucyl-tRNA synthetase
24	COG0061	Predicted sugar kinase
25	COG0071	Molecular chaperone (small heat shock protein)

26	COG0072	Phenylalanyl-tRNA synthetase beta subunit
27	COG0073	EMAP domain
28	COG0074	Succinyl-CoA synthetase
29	COG0080	Ribosomal protein L11
30	COG0081	Ribosomal protein L1
31	COG0084	Mg-dependent DNase
32	COG0085	DNA-directed RNA polymerase
33	COG0086	DNA-directed RNA polymerase
34	COG0087	Ribosomal protein L3
35	COG0088	Ribosomal protein L4
36	COG0089	Ribosomal protein L23
37	COG0090	Ribosomal protein L2
38	COG0091	Ribosomal protein L22
39	COG0092	Ribosomal protein S3
40	COG0093	Ribosomal protein L14
41	COG0094	Ribosomal protein L5
42	COG0096	Ribosomal protein S8
43	COG0097	Ribosomal protein L6P/L9E
44	COG0098	Ribosomal protein S5
45	COG0099	Ribosomal protein S13
46	COG0100	Ribosomal protein S11
47	COG0101	Pseudouridylate synthase
48	COG0102	Ribosomal protein L13
49	COG0103	Ribosomal protein S9
50	COG0105	Nucleoside diphosphate kinase
51	COG0112	Glycine/serine hydroxymethyltransferase
52	COG0113	Delta-aminolevulinic acid dehydratase

53	COG0114	Fumarase
54	COG0115	Branched-chain amino acid aminotransferase/4-amino-4-deoxychorismate lyase
55	COG0124	Histidyl-tRNA synthetase
56	COG0125	Thymidylate kinase
57	COG0130	Pseudouridine synthase
58	COG0136	Aspartate-semialdehyde dehydrogenase
59	COG0142	Geranylgeranyl pyrophosphate synthase
60	COG0143	Methionyl-tRNA synthetase
61	COG0152	Phosphoribosylaminoimidazolesuccinocarboxamide (SAICAR) synthase
62	COG0156	7-keto-8-aminopelargonate synthetase and related enzymes
63	COG0162	Tyrosyl-tRNA synthetase
64	COG0164	Ribonuclease HII
65	COG0172	Seryl-tRNA synthetase
66	COG0173	Aspartyl-tRNA synthetase
67	COG0174	Glutamine synthetase
68	COG0177	Predicted EndoIII-related endonuclease
69	COG0178	Excinuclease ATPase subunit
70	COG0180	Tryptophanyl-tRNA synthetase
71	COG0181	Porphobilinogen deaminase
72	COG0183	Acetyl-CoA acetyltransferase
73	COG0184	Ribosomal protein S15P/S13E
74	COG0187	Type IIA topoisomerase (DNA gyrase/topo II
75	COG0188	Type IIA topoisomerase (DNA gyrase/topo II
76	COG0190	10-methylene-tetrahydrofolate dehydrogenase/Methenyl tetrahydrofolate cyclohydrolase
77	COG0193	Peptidyl-tRNA hydrolase
78	COG0194	Guanylate kinase
79	COG0195	Transcription elongation factor

80	COG0197	Ribosomal protein L16/L10E
81	COG0198	Ribosomal protein L24
82	COG0199	Ribosomal protein S14
83	COG0200	Ribosomal protein L15
84	COG0201	Preprotein translocase subunit SecY
85	COG0202	DNA-directed RNA polymerase
86	COG0203	Ribosomal protein L17
87	COG0204	1-acyl-sn-glycerol-3-phosphate acyltransferase
88	COG0206	Cell division GTPase
89	COG0208	Ribonucleotide reductase
90	COG0209	Ribonucleotide reductase
91	COG0210	Superfamily I DNA and RNA helicases
92	COG0211	Ribosomal protein L27
93	COG0212	5-formyltetrahydrofolate cyclo-ligase
94	COG0215	Cysteinyl-tRNA synthetase
95	COG0216	Protein chain release factor A
96	COG0217	Uncharacterized conserved protein
97	COG0218	Predicted GTPase
98	COG0220	Predicted S-adenosylmethionine-dependent methyltransferase
99	COG0222	Ribosomal protein L7/L12
100	COG0223	Methionyl-tRNA formyltransferase
101	COG0224	F0F1-type ATP synthase
102	COG0228	Ribosomal protein S16
103	COG0231	Translation elongation factor P (EF-P)/translation initiation factor 5A (eIF-5A)
104	COG0233	Ribosome recycling factor
105	COG0234	Co-chaperonin GroES (HSP10)
106	COG0236	Acyl carrier protein

107	COG0237	Dephospho-CoA kinase
108	COG0238	Ribosomal protein S18
109	COG0240	Glycerol-3-phosphate dehydrogenase
110	COG0242	N-formylmethionyl-tRNA deformylase
111	COG0244	Ribosomal protein L10
112	COG0249	Mismatch repair ATPase (MutS family)
113	COG0250	Transcription antiterminator
114	COG0253	Diaminopimelate epimerase
115	COG0254	Ribosomal protein L31
116	COG0255	Ribosomal protein L29
117	COG0256	Ribosomal protein L18
118	COG0258	5'-3' exonuclease (including N-terminal domain of PolI)
119	COG0260	Leucyl aminopeptidase
120	COG0261	Ribosomal protein L21
121	COG0264	Translation elongation factor Ts
122	COG0265	Trypsin-like serine proteases
123	COG0267	Ribosomal protein L33
124	COG0268	Ribosomal protein S20
125	COG0272	NAD-dependent DNA ligase (contains BRCT domain type II)
126	COG0275	Predicted S-adenosylmethionine-dependent methyltransferase involved in cell envelope biogenesis
127	COG0276	Protoheme ferro-lyase (ferrochelatase)
128	COG0280	Phosphotransacetylase
129	COG0281	Malic enzyme
130	COG0282	Acetate kinase
131	COG0283	Cytidylate kinase
132	COG0285	Folylpolyglutamate synthase
133	COG0289	Dihydrodipicolinate reductase

134	COG0290	Translation initiation factor 3 (IF-3)
135	COG0292	Ribosomal protein L20
136	COG0304	3-oxoacyl-(acyl-carrier-protein) synthase
137	COG0305	Replicative DNA helicase
138	COG0313	Predicted methyltransferases
139	COG0316	Uncharacterized conserved protein
140	COG0318	Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II
141	COG0319	Predicted metal-dependent hydrolase
142	COG0320	Lipoate synthase
143	COG0322	Nuclease subunit of the excinuclease complex
144	COG0323	DNA mismatch repair enzyme (predicted ATPase)
145	COG0324	tRNA delta(2)-isopentenylpyrophosphate transferase
146	COG0328	Ribonuclease HI
147	COG0329	Dihydrodipicolinate synthase/N-acetylneuraminate lyase
148	COG0331	(acyl-carrier-protein) S-malonyltransferase
149	COG0332	3-oxoacyl-[acyl-carrier-protein] synthase III
150	COG0333	Ribosomal protein L32
151	COG0335	Ribosomal protein L19
152	COG0336	tRNA-(guanine-N1)-methyltransferase
153	COG0340	Biotin-(acetyl-CoA carboxylase) ligase
154	COG0341	Preprotein translocase subunit SecF
155	COG0342	Preprotein translocase subunit SecD
156	COG0343	Queuine/archaeosine tRNA-ribosyltransferase
157	COG0353	Recombinational DNA repair protein (RecF pathway)
158	COG0355	F0F1-type ATP synthase
159	COG0356	F0F1-type ATP synthase
160	COG0357	Predicted S-adenosylmethionine-dependent methyltransferase involved in bacterial cell division

161	COG0358	DNA primase (bacterial type)
162	COG0359	Ribosomal protein L9
163	COG0360	Ribosomal protein S6
164	COG0361	Translation initiation factor 1 (IF-1)
165	COG0372	Citrate synthase
166	COG0381	UDP-N-acetylglucosamine 2-epimerase
167	COG0407	Uroporphyrinogen-III decarboxylase
168	COG0436	Aspartate/tyrosine/aromatic aminotransferase
169	COG0438	Glycosyltransferase
170	COG0441	Threonyl-tRNA synthetase
171	COG0442	Prolyl-tRNA synthetase
172	COG0443	Molecular chaperone
173	COG0445	NAD/FAD-utilizing enzyme apparently involved in cell division
174	COG0450	Peroxiredoxin
175	COG0457	FOG: TPR repeat
176	COG0459	Chaperonin GroEL (HSP60 family)
177	COG0463	Glycosyltransferases involved in cell wall biogenesis
178	COG0465	ATP-dependent Zn proteases
179	COG0468	RecA/RadA recombinase
180	COG0470	ATPase involved in DNA replication
181	COG0472	UDP-N-acetylmuramyl pentapeptide phosphotransferase/UDP-N-acetylglucosamine-1-phosphate transferase
182	COG0473	Isocitrate/isopropylmalate dehydrogenase
183	COG0477	Permeases of the major facilitator superfamily
184	COG0479	Succinate dehydrogenase/fumarate reductase
185	COG0480	Translation elongation factors (GTPases)
186	COG0481	Membrane GTPase LepA
187	COG0482	Predicted tRNA(5-methylaminomethyl-2-thiouridylate) methyltransferase

188	COG0483	Archaeal fructose-1
189	COG0484	DnaJ-class molecular chaperone with C-terminal Zn finger domain
190	COG0486	Predicted GTPase
191	COG0488	ATPase components of ABC transporters with duplicated ATPase domains
192	COG0489	ATPases involved in chromosome partitioning
193	COG0492	Thioredoxin reductase
194	COG0493	NADPH-dependent glutamate synthase beta chain and related oxidoreductases
195	COG0494	NTP pyrophosphohydrolases including oxidative damage repair enzymes
196	COG0495	Leucyl-tRNA synthetase
197	COG0497	ATPase involved in DNA repair
198	COG0500	SAM-dependent methyltransferases
199	COG0504	CTP synthase (UTP-ammonia lyase)
200	COG0508	Pyruvate/2-oxoglutarate dehydrogenase complex
201	COG0513	Superfamily II DNA and RNA helicases
202	COG0517	FOG: CBS domain
203	COG0522	Ribosomal protein S4 and related proteins
204	COG0525	Valyl-tRNA synthetase
205	COG0526	Thiol-disulfide isomerase and thioredoxins
206	COG0527	Aspartokinases
207	COG0528	Uridylate kinase
208	COG0531	Amino acid transporters
209	COG0532	Translation initiation factor 2 (IF-2
210	COG0533	Metal-dependent proteases with possible chaperone activity
211	COG0536	Predicted GTPase
212	COG0537	Diadenosine tetraphosphate (Ap4A) hydrolase and other HIT family hydrolases
213	COG0539	Ribosomal protein S1
214	COG0541	Signal recognition particle GTPase

215	COG0542	ATPases with chaperone activity
216	COG0544	FKBP-type peptidyl-prolyl cis-trans isomerase (trigger factor)
217	COG0550	Topoisomerase IA
218	COG0552	Signal recognition particle GTPase
219	COG0556	Helicase subunit of the DNA excision repair complex
220	COG0558	Phosphatidylglycerophosphate synthase
221	COG0563	Adenylate kinase and related kinases
222	COG0564	Pseudouridylate synthases
223	COG0566	rRNA methylases
224	COG0567	2-oxoglutarate dehydrogenase complex
225	COG0568	DNA-directed RNA polymerase
226	COG0571	dsRNA-specific ribonuclease
227	COG0575	CDP-diglyceride synthetase
228	COG0576	Molecular chaperone GrpE (heat shock protein)
229	COG0582	Integrase
230	COG0587	DNA polymerase III
231	COG0589	Universal stress protein UspA and related nucleotide-binding proteins
232	COG0591	Na+/proline symporter
233	COG0592	DNA polymerase sliding clamp subunit (PCNA homolog)
234	COG0593	ATPase involved in DNA replication initiation
235	COG0594	RNase P protein component
236	COG0596	Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)
237	COG0597	Lipoprotein signal peptidase
238	COG0602	Organic radical activating enzymes
239	COG0603	Predicted PP-loop superfamily ATPase
240	COG0605	Superoxide dismutase
241	COG0607	Rhodanese-related sulfurtransferase

242	COG0608	Single-stranded DNA-specific exonuclease
243	COG0612	Predicted Zn-dependent peptidases
244	COG0617	tRNA nucleotidyltransferase/poly(A) polymerase
245	COG0621	2-methylthioadenine synthetase
246	COG0624	Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase and related deacylases
247	COG0628	Predicted permease
248	COG0629	Single-stranded DNA-binding protein
249	COG0632	Holliday junction resolvasome
250	COG0635	Coproporphyrinogen III oxidase and related Fe-S oxidoreductases
251	COG0636	F0F1-type ATP synthase
252	COG0642	Signal transduction histidine kinase
253	COG0653	Preprotein translocase subunit SecA (ATPase
254	COG0654	2-polyprenyl-6-methoxyphenol hydroxylase and related FAD-dependent oxidoreductases
255	COG0668	Small-conductance mechanosensitive channel
256	COG0670	Integral membrane protein
257	COG0671	Membrane-associated phospholipid phosphatase
258	COG0679	Predicted permeases
259	COG0681	Signal peptidase I
260	COG0682	Prolipoprotein diacylglyceryltransferase
261	COG0690	Preprotein translocase subunit SecE
262	COG0691	tmRNA-binding protein
263	COG0694	Thioredoxin-like proteins and domains
264	COG0695	Glutaredoxin and related proteins
265	COG0697	Permeases of the drug/metabolite transporter (DMT) superfamily
266	COG0706	Preprotein translocase subunit YidC
267	COG0707	UDP-N-acetylglucosamine:LPS N-acetylglucosamine transferase
268	COG0711	F0F1-type ATP synthase

269	COG0712	F0F1-type ATP synthase
270	COG0718	Uncharacterized protein conserved in bacteria
271	COG0720	6-pyruvoyl-tetrahydropterin synthase
272	COG0736	Phosphopantetheinyl transferase (holo-ACP synthase)
273	COG0739	Membrane proteins related to metalloendopeptidases
274	COG0740	Protease subunit of ATP-dependent Clp proteases
275	COG0742	N6-adenine-specific methylase
276	COG0745	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
277	COG0749	DNA polymerase I - 3'-5' exonuclease and polymerase domains
278	COG0750	Predicted membrane-associated Zn-dependent proteases 1
279	COG0755	ABC-type transport system involved in cytochrome c biogenesis
280	COG0760	Parvulin-like peptidyl-prolyl isomerase
281	COG0764	3-hydroxymyristoyl/3-hydroxydecanoyl-(acyl carrier protein) dehydratases
282	COG0765	ABC-type amino acid transport system
283	COG0766	UDP-N-acetylglucosamine enolpyruvyl transferase
284	COG0768	Cell division protein FtsI/penicillin-binding protein 2
285	COG0769	UDP-N-acetylmuramyl tripeptide synthase
286	COG0770	UDP-N-acetylmuramyl pentapeptide synthase
287	COG0771	UDP-N-acetylmuramoylalanine-D-glutamate ligase
288	COG0772	Bacterial cell division membrane protein
289	COG0773	UDP-N-acetylmuramate-alanine ligase
290	COG0776	Bacterial nucleoid DNA-binding protein
291	COG0779	Uncharacterized protein conserved in bacteria
292	COG0780	Enzyme related to GTP cyclohydrolase I
293	COG0781	Transcription termination factor
294	COG0782	Transcription elongation factor
295	COG0783	DNA-binding ferritin-like protein (oxidative damage protectant)

296	COG0787	Alanine racemase
297	COG0789	Predicted transcriptional regulators
298	COG0793	Periplasmic protease
299	COG0794	Predicted sugar phosphate isomerase involved in capsule formation
300	COG0799	Uncharacterized homolog of plant Iojap protein
301	COG0802	Predicted ATPase or kinase
302	COG0805	Sec-independent protein secretion pathway component TatC
303	COG0806	RimM protein
304	COG0809	S-adenosylmethionine:tRNA-ribosyltransferase-isomerase (queuine synthetase)
305	COG0812	UDP-N-acetylmuramate dehydrogenase
306	COG0816	Predicted endonuclease involved in recombination (possible Holliday junction resolvase in Mycoplasmas and B. subtilis)
307	COG0822	NifU homolog involved in Fe-S cluster formation
308	COG0828	Ribosomal protein S21
309	COG0834	ABC-type amino acid transport/signal transduction systems
310	COG0841	Cation/multidrug efflux pump
311	COG0847	DNA polymerase III
312	COG0849	Actin-like ATPase involved in cell division
313	COG0858	Ribosome-binding factor A
314	COG0861	Membrane protein TerC
315	COG1009	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na+/H+ antiporter
316	COG1024	Enoyl-CoA hydratase/carnithine racemase
317	COG1028	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)
318	COG1048	Aconitase A
319	COG1053	Succinate dehydrogenase/fumarate reductase
320	COG1054	Predicted sulfurtransferase
321	COG1066	Predicted ATP-dependent serine protease
322	COG1074	ATP-dependent exoDNAse (exonuclease V) beta subunit (contains helicase and exonuclease domains)

323	COG1104	Cysteine sulfinate desulfinase/cysteine desulfurase and related enzymes
324	COG1108	ABC-type Mn2+/Zn2+ transport systems
325	COG1109	Phosphomannomutase
326	COG1121	ABC-type Mn/Zn transport systems
327	COG1126	ABC-type polar amino acid transport system
328	COG1132	ABC-type multidrug transport system
329	COG1136	ABC-type antimicrobial peptide transport system
330	COG1158	Transcription termination factor
331	COG1159	GTPase
332	COG1160	Predicted GTPases
333	COG1171	Threonine dehydratase
334	COG1181	D-alanine-D-alanine ligase and related ATP-grasp enzymes
335	COG1185	Polyribonucleotide nucleotidyltransferase (polynucleotide phosphorylase)
336	COG1186	Protein chain release factor B
337	COG1187	16S rRNA uridine-516 pseudouridylate synthase and related pseudouridylate synthases
338	COG1195	Recombinational DNA repair ATPase (RecF pathway)
339	COG1197	Transcription-repair coupling factor (superfamily II helicase)
340	COG1198	Primosomal protein N' (replication factor Y) - superfamily II helicase
341	COG1200	RecG-like helicase
342	COG1207	N-acetylglucosamine-1-phosphate uridyltransferase (contains nucleotidyltransferase and I-patch acetyltransferase domains)
343	COG1214	Inactive homolog of metal-dependent proteases
344	COG1217	Predicted membrane GTPase involved in stress response
345	COG1219	ATP-dependent protease Clp
346	COG1220	ATP-dependent protease HslVU (ClpYQ)
347	COG1249	Pyruvate/2-oxoglutarate dehydrogenase complex
348	COG1250	3-hydroxyacyl-CoA dehydrogenase
349	COG1253	Hemolysins and related proteins containing CBS domains

350	COG1271	Cytochrome bd-type quinol oxidase
351	COG1278	Cold shock proteins
352	COG1286	Uncharacterized membrane protein
353	COG1294	Cytochrome bd-type quinol oxidase
354	COG1295	Predicted membrane protein
355	COG1301	Na+/H+-dicarboxylate symporters
356	COG1304	L-lactate dehydrogenase (FMN-dependent) and related alpha-hydroxy acid dehydrogenases
357	COG1314	Preprotein translocase subunit SecG
358	COG1381	Recombinational DNA repair protein (RecF pathway)
359	COG1385	Uncharacterized protein conserved in bacteria
360	COG1396	Predicted transcriptional regulators
361	COG1426	Uncharacterized protein conserved in bacteria
362	COG1434	Uncharacterized conserved protein
363	COG1466	DNA polymerase III
364	COG1475	Predicted transcriptional regulators
365	COG1496	Uncharacterized conserved protein
366	COG1502	Phosphatidylserine/phosphatidylglycerophosphate/cardiolipin synthases and related enzymes
367	COG1544	Ribosome-associated protein Y (PSrp-1)
368	COG1566	Multidrug resistance efflux pump
369	COG1570	Exonuclease VII
370	COG1589	Cell division septal protein
371	COG1651	Protein-disulfide isomerase
372	COG1670	Acetyltransferases
373	COG1674	DNA segregation ATPase FtsK/SpoIIIE and related proteins
374	COG1686	D-alanyl-D-alanine carboxypeptidase
375	COG1722	Exonuclease VII small subunit
376	COG1758	DNA-directed RNA polymerase

377	COG1792	Cell shape-determining protein
378	COG1806	Uncharacterized protein conserved in bacteria
379	COG1825	Ribosomal protein L25 (general stress protein Ctc)
380	COG1841	Ribosomal protein L30/L7E
381	COG1862	Preprotein translocase subunit YajC
382	COG1959	Predicted transcriptional regulator
383	COG2009	Succinate dehydrogenase/fumarate reductase
384	COG2171	Tetrahydrodipicolinate N-succinyltransferase
385	COG2226	Methylase involved in ubiquinone/menaquinone biosynthesis
386	COG2255	Holliday junction resolvasome
387	COG2271	Sugar phosphate permease
388	COG2812	DNA polymerase III
389	COG2890	Methylase of polypeptide chain release factors
390	COG2919	Septum formation initiator
391	COG3027	Uncharacterized protein conserved in bacteria
392	COG3307	Lipid A core - O-antigen ligase and related enzymes
393	COG4536	Putative Mg2+ and Co2+ transporter CorB
394	COG4974	Site-specific recombinase XerD
395	COG5405	ATP-dependent protease HslVU (ClpYQ)

Appendix Table 3.2 – List of PPIs for *Eco, Ype, Vch, Sau* and *Rpr* based on Raja *et al* interactions for the conserved COGs from Table 3.1

proteins not present in the clone set

Bait Eco	Prey Eco	Bait COG	Prey COG	Bait Vch	Prey Vch	Bait Ype	Prey Ype	Bait Sau	Prey Sau	Bait Rpr	Prey Rpr
P0ABB0	P0A805	COG0056	COG0233	Q9KNH3	Q9KPV5	Q8Z9S4	Q8ZH63	Q5HE95	Q5HGH2	O50288	Q9ZE08
P07649	P0A805	COG0101	COG0233	Q9KTA4	Q9KPV5	Q8ZD27	Q8ZH63	Q5HDY9	Q5HGH2	Q9ZCA3	Q9ZE08
P0A805	P0A805	COG0233	COG0233	Q9KPV5	Q9KPV5	Q8ZH63	Q8ZH63	Q5HGH2	Q5HGH2	Q9ZE08	Q9ZE08
P0A7J7	P0A805	COG0080	COG0233	Q9KV34	Q9KPV5	Q8ZAP1	Q8ZH63	Q5HID8	Q5HGH2	Q9ZE24	Q9ZE08
P0A763	P0A805	COG0105	COG0233	Q9KTX4	Q9KPV5	Q8ZCT2	Q8ZH63	Q5HFV4	Q5HGH2	Q9ZE91	Q9ZE08
P0ABB0	P0A6E6	COG0056	COG0355	Q9KNH3	Q9KNH6	Q8Z9S4	P58647	Q5HE95	Q5HE98	O50288	Q9ZCF3
P0ABB4	P0A6E6	COG0055	COG0355	Q9KNH5	Q9KNH6	Q7CFM8	P58647	Q5HE97	Q5HE98	O50290	Q9ZCF3
P0ABA6	P0A6E6	COG0224	COG0355	Q9KNH4	Q9KNH6	Q8Z9S5	P58647	Q5HE96	Q5HE98	O50289	Q9ZCF3
P0A6E6	P0A6E6	COG0355	COG0355	Q9KNH6	Q9KNH6	P58647	P58647	Q5HE98	Q5HE98	Q9ZCF3	Q9ZCF3
P0AB77	P0AB77	COG0156	COG0156	Q9KSZ3	Q9KSZ3	Q8CWI4	Q8CWI4	Q5HIC5	Q5HIC5	Q9ZCB8	Q9ZCB8
P0A7W7	P60422	COG0096	COG0090	Q9KNZ8	Q9KNY7	Q8ZJ98	P60436	Q5HDX2	Q5HDW1	Q9ZCR9	Q9ZCQ8
P0A7W7	P0A7W7	COG0096	COG0096	Q9KNZ8	Q9KNZ8	Q8ZJ98	Q8ZJ98	Q5HDX2	Q5HDX2	Q9ZCR9	Q9ZCR9
P0A6K3	P0AGJ2	COG0242	COG0566	Q9KVU3	Q9KTT5	Q8ZJ79	Q8ZIV4	Q5HGL7	Q5HIE3	Q9ZDV8	Q9ZCQ5
P0AA25	P0A9P4	COG0526	COG0492	P32557	Q9KSS4	Q7CKZ3	Q7CHI6	Q5HGT9	Q5HHQ4	Q9ZEE0	Q9ZD33
P0A7N4	P0A7W7	COG0333	COG0096	Q9KQH3	Q9KNZ8	Q8ZFT9	Q8ZJ98	Q5HGV6	Q5HDX2	Q9ZCH0	Q9ZCR9
P0A7W7	P0AG48	COG0096	COG0261	Q9KNZ8	Q9KUT0	Q8ZJ98	Q0WBD7	Q5HDX2	Q5HFB6	Q9ZCR9	Q9ZCI9
P0A7W7	P02413	COG0096	COG0200	Q9KNZ8	Q9KP03	Q8ZJ98	Q8ZJ93	Q5HDX2	Q5HDX7	Q9ZCR9	Q9ZCS4
P0A7W7	P0A7V8	COG0096	COG0522	Q9KNZ8	Q9KP07	Q8ZJ98	Q8ZJ88	Q5HDX2	Q5HF54	Q9ZCR9	Q9ZDI3
P0A7W7	P0A7T7	COG0096	COG0238	Q9KNZ8	Q9KUZ0	Q8ZJ98	Q8ZB83	Q5HDX2	Q5HIS7	Q9ZCR9	Q9ZEA5
P0ABB0	P0ABB4	COG0056	COG0055	Q9KNH3	Q9KNH5	Q8Z9S4	Q7CFM8	Q5HE95	Q5HE97	O50288	O50290
P0ABB4	P0ABB0	COG0055	COG0056	Q9KNH5	Q9KNH3	Q7CFM8	Q8Z9S4	Q5HE97	Q5HE95	O50290	O50288
P60716	P60716	COG0320	COG0320	Q9KTF9	Q9KTF9	Q8ZDH0	Q8ZDH0	Q5HHG0	Q5HHG0	O05959	O05959

P28305	P28305	COG0115	COG0115	Q9KVV9	Q9KVV9	Q8D1L3	Q8D1L3	Q5HIC1	Q5HIC1	O05970	O05970
P0A6U3	P25522	COG0445	COG0486	Q9KNG4	Q9KVY5	Q8Z9R8	Q8Z9U2	Q5HCI4	Q5HCI3	Q9ZE90	Q9ZCI1
P0A7W1	P0ADZ0	COG0098	COG0089	Q9KP01	Q9KNY6	Q8ZJ95	P69963	Q5HDX5	Q5HDW0	Q9ZCS2	Q9ZCQ7
P0A6F5	P09372	COG0459	COG0576	Q9KNR7	O30862	Q8ZIY3	Q7CH40	Q5HEH2	Q5HFH9	Q9ZCT7	Q9ZCT4
P06612	P07813	COG0550	COG0495	Q9KRB2	Q9KTE6	Q7CIL8	Q8ZDF8	Q5HGI2	Q5HF16	Q9ZDK2	Q9ZDB1
P08312	P08312	COG0016	COG0016	Q9KSN7	Q9KSN7	Q8ZDX0	Q8ZDX0	Q5HGU6	Q5HGU6	Q9ZDB5	Q9ZDB5
P08337	P08337	COG0494	COG0494	Q9KU53	Q9KU53	Q8ZHU8	Q8ZHU8	0	0	Q9ZDT9	Q9ZDT9
P0A7T7	P02358	COG0238	COG0360	Q9KUZ0	Q9KUZ2	Q8ZB83	Q8ZB81	Q5HIS7	Q5HIS9	Q9ZEA5	Q9ZEA6
P0A6F5	P76524	COG0459	COG0006	Q9KNR7	Q9KVS2	Q8ZIY3	Q8CLF7	Q5HEH2	Q5HF67	Q9ZCT7	Q9ZD64
P08312	P07395	COG0016	COG0072	Q9KSN7	Q9KSN6	Q8ZDX0	Q8ZDX1	Q5HGU6	Q5HGU5	Q9ZDB5	Q9ZDB4
P08337	P0AC47	COG0494	COG0479	Q9KU53	Q9KQB2	Q8ZHU8	Q7CKM7	0	0	Q9ZDT9	Q9ZEA1
P31552	P31552	COG0318	COG0318	H9L4Q1	H9L4Q1	Q8D0S3	Q8D0S3	Q5HEY2	Q5HEY2	0	0
P0A8M3	P0A8M3	COG0441	COG0441	Q9KMN7	Q9KMN7	Q8ZDW5	Q8ZDW5	Q5HF90	Q5HF90	O05947	O05947
P0AGD7	P10121	COG0541	COG0552	Q9KUG1	Q9KVJ6	Q7CK91	Q8D1J2	0	0	Q9ZDZ0	O05948
P63389	P60716	COG0488	COG0320	Q9KU31	Q9KTF9	Q8CK90	Q8ZDH0	0	Q5HHG0	Q9ZE86	O05959
P0A6L2	P0A6L2	COG0329	COG0329	Q9KR67	Q9KR67	Q8ZCD0	Q8ZCD0	Q5HG25	Q5HG25	O05969	O05969
P0AE08	P28305	COG0450	COG0115	Q9KTZ9	Q9KVV9	Q7CK47	Q8D1L3	Q5HIR5	Q5HIC1	Q9ZDK1	O05970
P0ABH7	P0ABH7	COG0372	COG0372	Q9KSC1	Q9KSC1	Q7CH41	Q7CH41	0	0	P09948	P09948
P00579	P00579	COG0568	COG0568	P50511	P50511	Q7CKW6	Q7CKW6	Q5HFJ9	Q5HFJ9	P33451	P33451
P23367	P23367	COG0323	COG0323	Q9KV13	Q9KV13	Q8ZIW4	Q8ZIW4	Q5HGD5	Q5HGD5	Q9ZC88	Q9ZC88
P0A8L1	P0A8L1	COG0172	COG0172	Q9KSZ6	Q9KSZ6	Q8ZGC4	Q8ZGC4	Q5HJY7	Q5HJY7	Q9ZCG5	Q9ZCG5
Q46939	Q46939	COG0183	COG0183	Q9KT59	Q9KT59	Q8ZAM9	Q8ZAM9	Q5HIA0	Q5HIA0	Q9ZCJ5	Q9ZCJ5
P07118	P07118	COG0525	COG0525	Q9KP73	Q9KP73	Q8ZBH1	Q8ZBH1	Q5HFA8	Q5HFA8	Q9ZCN6	Q9ZCN6
P0A8J8	P0A8J8	COG0513	COG0513	Q9KV52	Q9KV52	Q8ZAD8	Q8ZAD8	Q5HEB9	Q5HEB9	Q9ZCQ0	Q9ZCQ0
P0A9P6	P0A9P6	COG0513	COG0513	Q9KV52	Q9KV52	Q8ZAD8	Q8ZAD8	Q5HEB9	Q5HEB9	Q9ZCQ0	Q9ZCQ0
P0AGJ2	P0AGJ2	COG0566	COG0566	Q9KTT5	Q9KTT5	Q8ZIV4	Q8ZIV4	Q5HIE3	Q5HIE3	Q9ZCQ5	Q9ZCQ5
P00959	P60422	COG0073	COG0090	Q9KT69	Q9KNY7	Q8ZG01	P60436	Q5HII6	Q5HDW1	0	Q9ZCQ8
P60390	P16384	COG0275	COG0324	Q9KPF9	Q9KV12	Q8ZIF7	Q8ZIW3	Q5HGQ2	Q5HGC9	Q9ZCY2	Q9ZD37
P22939	P22939	COG0142	COG0142	Q9KUT1	Q9KUT1	Q7CK38	Q7CK38	0	0	Q9ZD65	Q9ZD65

P09980	P09980	COG0210	COG0210	Q9KVH9	Q9KVH9	Q8D1K7	Q8D1K7	Q5HEL7	Q5HEL7	Q9ZD95	Q9ZD95
P0A6S7	P0A6S7	COG0240	COG0240	Q9KNT0	Q9KNT0	Q8ZJM6	Q8ZJM6	Q5HFU9	Q5HFU9	Q9ZDA0	Q9ZDA0
P0AG90	P07395	COG0342	COG0072	Q9KTY7	Q9KSN6	Q8D163	Q8ZDX1	0	Q5HGU5	Q9ZCW8	Q9ZDB4
P07395	P07395	COG0072	COG0072	Q9KSN6	Q9KSN6	Q8ZDX1	Q8ZDX1	Q5HGU5	Q5HGU5	Q9ZDB4	Q9ZDB4
P07395	P08312	COG0072	COG0016	Q9KSN6	Q9KSN7	Q8ZDX1	Q8ZDX0	Q5HGU5	Q5HGU6	Q9ZDB4	Q9ZDB5
P0AE08	P0AE08	COG0450	COG0450	Q9KTZ9	Q9KTZ9	Q7CK47	Q7CK47	Q5HIR5	Q5HIR5	Q9ZDK1	Q9ZDK1
P0ACE7	P0ACE7	COG0537	COG0537	Q9KQV1	Q9KQV1	Q7CJ15	Q7CJ15	0	0	Q9ZDL1	Q9ZDL1
P0C058	P0C058	COG0071	COG0071	Q9KVX0	Q9KVX0	Q8Z9V6	Q8Z9V6	0	0	Q9ZDQ3	Q9ZDQ3
P20083	P20083	COG0187	COG0187	Q9KVX3	Q9KVX3	Q7CGH6	Q7CGH6	Q5HG65	Q5HG65	Q9ZDU7	Q9ZDU7
P0A698	P0A8F8	COG0178	COG0556	Q9KUW5	0	Q8ZJ07	Q8ZGW7	Q5HHQ9	Q5HHR0	Q9ZCC3	Q9ZDW2
P39452	P0A8F8	COG0209	COG0556	Q9KSK0	0	Q8D117	Q8ZGW7	0	Q5HHR0	Q9ZD34	Q9ZDW2
P0A8F8	P0A8F8	COG0556	COG0556	0	0	Q8ZGW7	Q8ZGW7	Q5HHR0	Q5HHR0	Q9ZDW2	Q9ZDW2
P0AFG3	P0AFG3	COG0567	COG0567	Q9KQB3	Q9KQB3	Q7CH46	Q7CH46	Q5HG06	Q5HG06	Q9ZDY3	Q9ZDY3
P0AFG6	P0AFG3	COG0508	COG0567	Q9KQB4	Q9KQB3	Q7CH47	Q7CH46	Q5HG07	Q5HG06	Q9ZDY4	Q9ZDY3
P0AFG6	P0AFG6	COG0508	COG0508	Q9KQB4	Q9KQB4	Q7CH47	Q7CH47	Q5HG07	Q5HG07	Q9ZDY4	Q9ZDY4
P10121	P0AGD7	COG0552	COG0541	Q9KVJ6	Q9KUG1	Q8D1J2	Q7CK91	0	0	O05948	Q9ZDZ0
P0AE08	P0A805	COG0450	COG0233	Q9KTZ9	Q9KPV5	Q7CK47	Q8ZH63	Q5HIR5	Q5HGH2	Q9ZDK1	Q9ZE08
P0A873	P0A873	COG0336	COG0336	Q9KUF8	Q9KUF8	Q8ZBU9	Q8ZBU9	Q5HGJ2	Q5HGJ2	Q9ZE37	Q9ZE37
P31600	P31600	COG0457	COG0457	Q9KTK1	Q9KTK1	Q7CJI7	Q7CJI7	0	0	Q9ZE55	Q9ZE55
P69441	P77806	COG0563	COG0436	Q9KTB7	Q9KQM1	O69172	Q7CJE7	Q5HDX9	0	Q9ZCS6	Q9ZE56
P77434	P77434	COG0436	COG0436	Q9KQM1	Q9KQM1	Q7CJE7	Q7CJE7	0	0	Q9ZE56	Q9ZE56
P77806	P77806	COG0436	COG0436	Q9KQM1	Q9KQM1	Q7CJE7	Q7CJE7	0	0	Q9ZE56	Q9ZE56
P52097	P52097	COG0037	COG0037	Q9KS29	Q9KS29	Q7CIP8	Q7CIP8	Q5HIG6	Q5HIG6	Q9ZEA3	Q9ZEA3
P05852	P05852	COG0533	COG0533	0	0	Q74RQ9	Q74RQ9	Q5HEF2	Q5HEF2	Q9ZEA8	Q9ZEA8

Appendix Table 3.3 – List of interactions found positive in *Eco, Sau, Rpr, Vch* and *Ype* using yeast two hybrid screens with pGADT7g and pGBGT7g vector combination

x indicates presence of interaction

Bait protein	Prey protein	Eco	Vch	Ype	Rpr	Sau
Top1	Syl	x	x	x	x	x
TruA	Rrf	x	x	x	x	
SyfA	SyfA	x	x	x		
MutT	MutT	x	x	x	x	x
AtpE	AtpE	x	x	x	x	
Ch60	GrpE	x	x	x		
Def	TrmH	x	x	x	x	x
MnmG	MnmE	x	x	x		x
Ndk	Rrf	x	x	x		
Rl11	Rrf	x	x	x	x	x
Rl32	Rs8	x	x	x	x	x
Rs18	Rs6	x	x	x		x
Rs5	Rl23	x	x	x	x	x
Rs8	Rs8	x	x	x	x	x
Rs8	Rl2	x	x	x	x	x
Rs8	Rl21	x	x	x	x	x
Rs8	RplO	x	x	x	x	x
Rs8	Rs18	x	x	x	x	x
Rs8	Rs4	x	x	x	x	x
Rrf	Rrf	x	x	x		x
TrxA	TrxB	x	x	x	x	
Kbl	Kbl	x	x	x		x

AtpG	AtpE	x	x	x		
AtpA	Rrf	x	x	x		
AtpA	RtpE	x	x	x	x	
AtpA	AtpB	x	x	x	x	x
AtpB	AtpE	x	x	x		
AtpB	AtpA	x	x	x	x	x
PabC	PabC	x	x	x	x	
LipA	LipA	x	x	x	x	

Appendix Table 3.4 - List of OG IDs and description (857) for OGs present in all species being studied (*Eco, Ype, Vch, Sau*)

Sr. No.	OG ID	OG Description
1	COG0001	Glutamate-1-semialdehyde aminotransferase
2	COG0002	Acetylglutamate semialdehyde dehydrogenase
3	COG0006	Xaa-Pro aminopeptidase
4	COG0007	Uroporphyrinogen-III methylase
5	COG0008	Glutamyl- and glutaminyl-tRNA synthetases
6	COG0009	Putative translation factor (SUA5)
7	COG0012	Predicted GTPase
8	COG0013	Alanyl-tRNA synthetase
9	COG0015	Adenylosuccinate lyase
10	COG0016	Phenylalanyl-tRNA synthetase alpha subunit
11	COG0017	Aspartyl/asparaginyl-tRNA synthetases
12	COG0018	Arginyl-tRNA synthetase
13	COG0019	Diaminopimelate decarboxylase
14	COG0020	Undecaprenyl pyrophosphate synthase
15	COG0021	Transketolase
16	COG0024	Methionine aminopeptidase
17	COG0025	NhaP-type Na+/H+ and K+/H+ antiporters
18	COG0026	Phosphoribosylaminoimidazole carboxylase (NCAIR synthetase)
19	COG0028	Thiamine pyrophosphate-requiring enzymes [acetolactate synthase
20	COG0030	Dimethyladenosine transferase (rRNA methylation)
21	COG0031	Cysteine synthase
22	COG0034	Glutamine phosphoribosylpyrophosphate amidotransferase
23	COG0035	Uracil phosphoribosyltransferase
24	COG0036	Pentose-5-phosphate-3-epimerase
25	COG0037	Predicted ATPase of the PP-loop superfamily implicated in cell cycle control
26	COG0039	Malate/lactate dehydrogenases
27	COG0040	ATP phosphoribosyltransferase

28	COG0041	Phosphoribosylcarboxyaminoimidazole (NCAIR) mutase
29	COG0042	tRNA-dihydrouridine synthase
30	COG0045	Succinyl-CoA synthetase
31	COG0046	Phosphoribosylformylglycinamidine (FGAM) synthase
32	COG0047	Phosphoribosylformylglycinamidine (FGAM) synthase
33	COG0048	Ribosomal protein S12
34	COG0049	Ribosomal protein S7
35	COG0050	GTPases - translation elongation factors
36	COG0051	Ribosomal protein S10
37	COG0052	Ribosomal protein S2
38	COG0053	Predicted Co/Zn/Cd cation transporters
39	COG0054	Riboflavin synthase beta-chain
40	COG0055	F0F1-type ATP synthase
41	COG0056	F0F1-type ATP synthase
42	COG0057	Glyceraldehyde-3-phosphate dehydrogenase/erythrose-4-phosphate dehydrogenase
43	COG0059	Ketol-acid reductoisomerase
44	COG0060	Isoleucyl-tRNA synthetase
45	COG0061	Predicted sugar kinase
46	COG0063	Predicted sugar kinase
47	COG0065	3-isopropylmalate dehydratase large subunit
48	COG0066	3-isopropylmalate dehydratase small subunit
49	COG0067	Glutamate synthase domain 1
50	COG0069	Glutamate synthase domain 2
51	COG0070	Glutamate synthase domain 3
52	COG0071	Molecular chaperone (small heat shock protein)
53	COG0072	Phenylalanyl-tRNA synthetase beta subunit
54	COG0073	EMAP domain
55	COG0074	Succinyl-CoA synthetase
56	COG0078	Ornithine carbamoyltransferase
57	COG0079	Histidinol-phosphate/aromatic aminotransferase and cobyric acid decarboxylase

58	COG0080	Ribosomal protein L11
59	COG0081	Ribosomal protein L1
60	COG0082	Chorismate synthase
61	COG0083	Homoserine kinase
62	COG0084	Mg-dependent DNase
63	COG0085	DNA-directed RNA polymerase
64	COG0086	DNA-directed RNA polymerase
65	COG0087	Ribosomal protein L3
66	COG0088	Ribosomal protein L4
67	COG0089	Ribosomal protein L23
68	COG0090	Ribosomal protein L2
69	COG0091	Ribosomal protein L22
70	COG0092	Ribosomal protein S3
71	COG0093	Ribosomal protein L14
72	COG0094	Ribosomal protein L5
73	COG0095	Lipoate-protein ligase A
74	COG0096	Ribosomal protein S8
75	COG0097	Ribosomal protein L6P/L9E
76	COG0098	Ribosomal protein S5
77	COG0099	Ribosomal protein S13
78	COG0100	Ribosomal protein S11
79	COG0101	Pseudouridylate synthase
80	COG0102	Ribosomal protein L13
81	COG0103	Ribosomal protein S9
82	COG0104	Adenylosuccinate synthase
83	COG0105	Nucleoside diphosphate kinase
84	COG0106	Phosphoribosylformimino-5-aminoimidazole carboxamide ribonucleotide (ProFAR) isomerase
85	COG0107	Imidazoleglycerol-phosphate synthase
86	COG0108	4-dihydroxy-2-butanone 4-phosphate synthase
87	COG0111	Phosphoglycerate dehydrogenase and related dehydrogenases

88	COG0112	Glycine/serine hydroxymethyltransferase
89	COG0113	Delta-aminolevulinic acid dehydratase
90	COG0114	Fumarase
91	COG0115	Branched-chain amino acid aminotransferase/4-amino-4-deoxychorismate lyase
92	COG0116	Predicted N6-adenine-specific DNA methylase
93	COG0117	Pyrimidine deaminase
94	COG0118	Glutamine amidotransferase
95	COG0119	Isopropylmalate/homocitrate/citramalate synthases
96	COG0120	Ribose 5-phosphate isomerase
97	COG0124	Histidyl-tRNA synthetase
98	COG0125	Thymidylate kinase
99	COG0126	3-phosphoglycerate kinase
100	COG0127	Xanthosine triphosphate pyrophosphatase
101	COG0128	5-enolpyruvylshikimate-3-phosphate synthase
102	COG0129	Dihydroxyacid dehydratase/phosphogluconate dehydratase
103	COG0130	Pseudouridine synthase
104	COG0131	Imidazoleglycerol-phosphate dehydratase
105	COG0132	Dethiobiotin synthetase
106	COG0133	Tryptophan synthase beta chain
107	COG0134	Indole-3-glycerol phosphate synthase
108	COG0135	Phosphoribosylanthranilate isomerase
109	COG0136	Aspartate-semialdehyde dehydrogenase
110	COG0137	Argininosuccinate synthase
111	COG0138	AICAR transformylase/IMP cyclohydrolase PurH (only IMP cyclohydrolase domain in Aful)
112	COG0139	Phosphoribosyl-AMP cyclohydrolase
113	COG0140	Phosphoribosyl-ATP pyrophosphohydrolase
114	COG0141	Histidinol dehydrogenase
115	COG0142	Geranylgeranyl pyrophosphate synthase
116	COG0143	Methionyl-tRNA synthetase
117	COG0144	tRNA and rRNA cytosine-C5-methylases

118	COG0147	Anthranilate/para-aminobenzoate synthases component I
119	COG0148	Enolase
120	COG0149	Triosephosphate isomerase
121	COG0150	Phosphoribosylaminoimidazole (AIR) synthetase
122	COG0151	Phosphoribosylamine-glycine ligase
123	COG0152	Phosphoribosylaminoimidazolesuccinocarboxamide (SAICAR) synthase
124	COG0156	7-keto-8-aminopelargonate synthetase and related enzymes
125	COG0159	Tryptophan synthase alpha chain
126	COG0160	4-aminobutyrate aminotransferase and related aminotransferases
127	COG0161	Adenosylmethionine-8-amino-7-oxononanoate aminotransferase
128	COG0162	Tyrosyl-tRNA synthetase
129	COG0164	Ribonuclease HII
130	COG0165	Argininosuccinate lyase
131	COG0166	Glucose-6-phosphate isomerase
132	COG0167	Dihydroorotate dehydrogenase
133	COG0168	Trk-type K+ transport systems
134	COG0169	Shikimate 5-dehydrogenase
135	COG0171	NAD synthase
136	COG0172	Seryl-tRNA synthetase
137	COG0173	Aspartyl-tRNA synthetase
138	COG0174	Glutamine synthetase
139	COG0176	Transaldolase
140	COG0177	Predicted EndoIII-related endonuclease
141	COG0178	Excinuclease ATPase subunit
142	COG0179	2-keto-4-pentenoate hydratase/2-oxohepta-3-ene-1
143	COG0180	Tryptophanyl-tRNA synthetase
144	COG0181	Porphobilinogen deaminase
145	COG0183	Acetyl-CoA acetyltransferase
146	COG0184	Ribosomal protein S15P/S13E
147	COG0187	Type IIA topoisomerase (DNA gyrase/topo II

148	COG0188	Type IIA topoisomerase (DNA gyrase/topo II
149	COG0190	10-methylene-tetrahydrofolate dehydrogenase/Methenyl tetrahydrofolate cyclohydrolase
150	COG0191	Fructose/tagatose bisphosphate aldolase
151	COG0192	S-adenosylmethionine synthetase
152	COG0193	Peptidyl-tRNA hydrolase
153	COG0194	Guanylate kinase
154	COG0195	Transcription elongation factor
155	COG0196	FAD synthase
156	COG0197	Ribosomal protein L16/L10E
157	COG0198	Ribosomal protein L24
158	COG0199	Ribosomal protein S14
159	COG0200	Ribosomal protein L15
160	COG0201	Preprotein translocase subunit SecY
161	COG0202	DNA-directed RNA polymerase
162	COG0203	Ribosomal protein L17
163	COG0204	1-acyl-sn-glycerol-3-phosphate acyltransferase
164	COG0205	6-phosphofructokinase
165	COG0206	Cell division GTPase
166	COG0207	Thymidylate synthase
167	COG0208	Ribonucleotide reductase
168	COG0209	Ribonucleotide reductase
169	COG0210	Superfamily I DNA and RNA helicases
170	COG0211	Ribosomal protein L27
171	COG0212	5-formyltetrahydrofolate cyclo-ligase
172	COG0215	Cysteinyl-tRNA synthetase
173	COG0216	Protein chain release factor A
174	COG0217	Uncharacterized conserved protein
175	COG0218	Predicted GTPase
176	COG0219	Predicted rRNA methylase (SpoU class)
177	COG0220	Predicted S-adenosylmethionine-dependent methyltransferase

178	COG0222	Ribosomal protein L7/L12
179	COG0223	Methionyl-tRNA formyltransferase
180	COG0224	F0F1-type ATP synthase
181	COG0225	Peptide methionine sulfoxide reductase
182	COG0226	ABC-type phosphate transport system
183	COG0228	Ribosomal protein S16
184	COG0231	Translation elongation factor P (EF-P)/translation initiation factor 5A (eIF-5A)
185	COG0233	Ribosome recycling factor
186	COG0234	Co-chaperonin GroES (HSP10)
187	COG0236	Acyl carrier protein
188	COG0237	Dephospho-CoA kinase
189	COG0238	Ribosomal protein S18
190	COG0239	Integral membrane protein possibly involved in chromosome condensation
191	COG0240	Glycerol-3-phosphate dehydrogenase
192	COG0242	N-formylmethionyl-tRNA deformylase
193	COG0244	Ribosomal protein L10
194	COG0246	Mannitol-1-phosphate/altronate dehydrogenases
195	COG0249	Mismatch repair ATPase (MutS family)
196	COG0250	Transcription antiterminator
197	COG0251	Putative translation initiation inhibitor
198	COG0252	L-asparaginase/archaeal Glu-tRNAGln amidotransferase subunit D
199	COG0253	Diaminopimelate epimerase
200	COG0254	Ribosomal protein L31
201	COG0255	Ribosomal protein L29
202	COG0256	Ribosomal protein L18
203	COG0258	5'-3' exonuclease (including N-terminal domain of PolI)
204	COG0260	Leucyl aminopeptidase
205	COG0261	Ribosomal protein L21
206	COG0262	Dihydrofolate reductase
207	COG0264	Translation elongation factor Ts

208	COG0265	Trypsin-like serine proteases
209	COG0266	Formamidopyrimidine-DNA glycosylase
210	COG0267	Ribosomal protein L33
211	COG0268	Ribosomal protein S20
212	COG0272	NAD-dependent DNA ligase (contains BRCT domain type II)
213	COG0274	Deoxyribose-phosphate aldolase
214	COG0275	Predicted S-adenosylmethionine-dependent methyltransferase involved in cell envelope biogenesis
215	COG0276	Protoheme ferro-lyase (ferrochelatase)
216	COG0280	Phosphotransacetylase
217	COG0281	Malic enzyme
218	COG0282	Acetate kinase
219	COG0283	Cytidylate kinase
220	COG0284	Orotidine-5'-phosphate decarboxylase
221	COG0285	Folylpolyglutamate synthase
222	COG0287	Prephenate dehydrogenase
223	COG0289	Dihydrodipicolinate reductase
224	COG0290	Translation initiation factor 3 (IF-3)
225	COG0292	Ribosomal protein L20
226	COG0294	Dihydropteroate synthase and related enzymes
227	COG0295	Cytidine deaminase
228	COG0299	Folate-dependent phosphoribosylglycinamide formyltransferase PurN
229	COG0301	Thiamine biosynthesis ATP pyrophosphatase
230	COG0303	Molybdopterin biosynthesis enzyme
231	COG0304	3-oxoacyl-(acyl-carrier-protein) synthase
232	COG0305	Replicative DNA helicase
233	COG0306	Phosphate/sulphate permeases
234	COG0307	Riboflavin synthase alpha chain
235	COG0313	Predicted methyltransferases
236	COG0314	Molybdopterin converting factor

237	COG0315	Molybdenum cofactor biosynthesis enzyme
238	COG0316	Uncharacterized conserved protein
239	COG0317	Guanosine polyphosphate pyrophosphohydrolases/synthetases
240	COG0318	Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II
241	COG0319	Predicted metal-dependent hydrolase
242	COG0320	Lipoate synthase
243	COG0322	Nuclease subunit of the excinuclease complex
244	COG0323	DNA mismatch repair enzyme (predicted ATPase)
245	COG0324	tRNA delta(2)-isopentenylpyrophosphate transferase
246	COG0325	Predicted enzyme with a TIM-barrel fold
247	COG0327	Uncharacterized conserved protein
248	COG0328	Ribonuclease HI
249	COG0329	Dihydrodipicolinate synthase/N-acetylneuraminate lyase
250	COG0331	(acyl-carrier-protein) S-malonyltransferase
251	COG0332	3-oxoacyl-[acyl-carrier-protein] synthase III
252	COG0333	Ribosomal protein L32
253	COG0335	Ribosomal protein L19
254	COG0336	tRNA-(guanine-N1)-methyltransferase
255	COG0337	3-dehydroquinate synthetase
256	COG0340	Biotin-(acetyl-CoA carboxylase) ligase
257	COG0341	Preprotein translocase subunit SecF
258	COG0342	Preprotein translocase subunit SecD
259	COG0343	Queuine/archaeosine tRNA-ribosyltransferase
260	COG0344	Predicted membrane protein
261	COG0345	Pyrroline-5-carboxylate reductase
262	COG0346	Lactoylglutathione lyase and related lyases
263	COG0350	Methylated DNA-protein cysteine methyltransferase
264	COG0351	Hydroxymethylpyrimidine/phosphomethylpyrimidine kinase
265	COG0352	Thiamine monophosphate synthase
266	COG0353	Recombinational DNA repair protein (RecF pathway)

267	COG0355	F0F1-type ATP synthase
268	COG0356	F0F1-type ATP synthase
269	COG0357	Predicted S-adenosylmethionine-dependent methyltransferase involved in bacterial cell division
270	COG0358	DNA primase (bacterial type)
271	COG0359	Ribosomal protein L9
272	COG0360	Ribosomal protein S6
273	COG0361	Translation initiation factor 1 (IF-1)
274	COG0362	6-phosphogluconate dehydrogenase
275	COG0363	6-phosphogluconolactonase/Glucosamine-6-phosphate isomerase/deaminase
276	COG0364	Glucose-6-phosphate 1-dehydrogenase
277	COG0365	Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases
278	COG0366	Glycosidases
279	COG0369	Sulfite reductase
280	COG0370	Fe2+ transport system protein B
281	COG0372	Citrate synthase
282	COG0373	Glutamyl-tRNA reductase
283	COG0381	UDP-N-acetylglucosamine 2-epimerase
284	COG0388	Predicted amidohydrolase
285	COG0389	Nucleotidyltransferase/DNA polymerase involved in DNA repair
286	COG0391	Uncharacterized conserved protein
287	COG0405	Gamma-glutamyltransferase
288	COG0406	Fructose-2
289	COG0407	Uroporphyrinogen-III decarboxylase
290	COG0413	Ketopantoate hydroxymethyltransferase
291	COG0414	Panthothenate synthetase
292	COG0415	Deoxyribodipyrimidine photolyase
293	COG0416	Fatty acid/phospholipid biosynthesis enzyme
294	COG0419	ATPase involved in DNA repair
295	COG0420	DNA repair exonuclease
296	COG0425	Predicted redox protein

297	COG0431	Predicted flavoprotein
298	COG0436	Aspartate/tyrosine/aromatic aminotransferase
299	COG0438	Glycosyltransferase
300	COG0439	Biotin carboxylase
301	COG0440	Acetolactate synthase
302	COG0441	Threonyl-tRNA synthetase
303	COG0442	Prolyl-tRNA synthetase
304	COG0443	Molecular chaperone
305	COG0444	ABC-type dipeptide/oligopeptide/nickel transport system
306	COG0445	NAD/FAD-utilizing enzyme apparently involved in cell division
307	COG0446	Uncharacterized NAD(FAD)-dependent dehydrogenases
308	COG0447	Dihydroxynaphthoic acid synthase
309	COG0449	Glucosamine 6-phosphate synthetase
310	COG0450	Peroxiredoxin
311	COG0451	Nucleoside-diphosphate-sugar epimerases
312	COG0452	Phosphopantothenoylcysteine synthetase/decarboxylase
313	COG0454	Histone acetyltransferase HPA2 and related acetyltransferases
314	COG0456	Acetyltransferases
315	COG0457	FOG: TPR repeat
316	COG0458	Carbamoylphosphate synthase large subunit (split gene in MJ)
317	COG0459	Chaperonin GroEL (HSP60 family)
318	COG0460	Homoserine dehydrogenase
319	COG0461	Orotate phosphoribosyltransferase
320	COG0462	Phosphoribosylpyrophosphate synthetase
321	COG0463	Glycosyltransferases involved in cell wall biogenesis
322	COG0465	ATP-dependent Zn proteases
323	COG0468	RecA/RadA recombinase
324	COG0469	Pyruvate kinase
325	COG0470	ATPase involved in DNA replication
326	COG0471	Di- and tricarboxylate transporters

327	COG0472	UDP-N-acetylmuramyl pentapeptide phosphotransferase/UDP-N-acetylglucosamine-1-phosphate transferase
328	COG0473	Isocitrate/isopropylmalate dehydrogenase
329	COG0476	Dinucleotide-utilizing enzymes involved in molybdopterin and thiamine biosynthesis family 2
330	COG0477	Permeases of the major facilitator superfamily
331	COG0479	Succinate dehydrogenase/fumarate reductase
332	COG0480	Translation elongation factors (GTPases)
333	COG0481	Membrane GTPase LepA
334	COG0482	Predicted tRNA(5-methylaminomethyl-2-thiouridylate) methyltransferase
335	COG0483	Archaeal fructose-1
336	COG0484	DnaJ-class molecular chaperone with C-terminal Zn finger domain
337	COG0486	Predicted GTPase
338	COG0488	ATPase components of ABC transporters with duplicated ATPase domains
339	COG0489	ATPases involved in chromosome partitioning
340	COG0491	Zn-dependent hydrolases
341	COG0492	Thioredoxin reductase
342	COG0493	NADPH-dependent glutamate synthase beta chain and related oxidoreductases
343	COG0494	NTP pyrophosphohydrolases including oxidative damage repair enzymes
344	COG0495	Leucyl-tRNA synthetase
345	COG0497	ATPase involved in DNA repair
346	COG0498	Threonine synthase
347	COG0500	SAM-dependent methyltransferases
348	COG0502	Biotin synthase and related enzymes
349	COG0503	Adenine/guanine phosphoribosyltransferases and related PRPP-binding proteins
350	COG0504	CTP synthase (UTP-ammonia lyase)
351	COG0505	Carbamoylphosphate synthase small subunit
352	COG0507	ATP-dependent exoDNAse (exonuclease V)
353	COG0508	Pyruvate/2-oxoglutarate dehydrogenase complex
354	COG0509	Glycine cleavage system H protein (lipoate-binding)
355	COG0510	Predicted choline kinase involved in LPS biosynthesis

356	COG0511	Biotin carboxyl carrier protein
357	COG0512	Anthranilate/para-aminobenzoate synthases component II
358	COG0513	Superfamily II DNA and RNA helicases
359	COG0514	Superfamily II DNA helicase
360	COG0516	IMP dehydrogenase/GMP reductase
361	COG0517	FOG: CBS domain
362	COG0518	GMP synthase - Glutamine amidotransferase domain
363	COG0519	GMP synthase
364	COG0520	Selenocysteine lyase
365	COG0521	Molybdopterin biosynthesis enzymes
366	COG0522	Ribosomal protein S4 and related proteins
367	COG0523	Putative GTPases (G3E family)
368	COG0524	Sugar kinases
369	COG0525	Valyl-tRNA synthetase
370	COG0526	Thiol-disulfide isomerase and thioredoxins
371	COG0527	Aspartokinases
372	COG0528	Uridylate kinase
373	COG0531	Amino acid transporters
374	COG0532	Translation initiation factor 2 (IF-2
375	COG0533	Metal-dependent proteases with possible chaperone activity
376	COG0534	Na+-driven multidrug efflux pump
377	COG0536	Predicted GTPase
378	COG0537	Diadenosine tetraphosphate (Ap4A) hydrolase and other HIT family hydrolases
379	COG0539	Ribosomal protein S1
380	COG0540	Aspartate carbamoyltransferase
381	COG0541	Signal recognition particle GTPase
382	COG0542	ATPases with chaperone activity
383	COG0544	FKBP-type peptidyl-prolyl cis-trans isomerase (trigger factor)
384	COG0546	Predicted phosphatases
385	COG0547	Anthranilate phosphoribosyltransferase

386	COG0548	Acetylglutamate kinase
387	COG0550	Topoisomerase IA
388	COG0551	Zn-finger domain associated with topoisomerase type I
389	COG0552	Signal recognition particle GTPase
390	COG0553	Superfamily II DNA/RNA helicases
391	COG0554	Glycerol kinase
392	COG0556	Helicase subunit of the DNA excision repair complex
393	COG0557	Exoribonuclease R
394	COG0558	Phosphatidylglycerophosphate synthase
395	COG0561	Predicted hydrolases of the HAD superfamily
396	COG0563	Adenylate kinase and related kinases
397	COG0564	Pseudouridylate synthases
398	COG0566	rRNA methylases
399	COG0567	2-oxoglutarate dehydrogenase complex
400	COG0568	DNA-directed RNA polymerase
401	COG0569	K+ transport systems
402	COG0571	dsRNA-specific ribonuclease
403	COG0572	Uridine kinase
404	COG0573	ABC-type phosphate transport system
405	COG0575	CDP-diglyceride synthetase
406	COG0576	Molecular chaperone GrpE (heat shock protein)
407	COG0577	ABC-type antimicrobial peptide transport system
408	COG0578	Glycerol-3-phosphate dehydrogenase
409	COG0581	ABC-type phosphate transport system
410	COG0582	Integrase
411	COG0583	Transcriptional regulator
412	COG0584	Glycerophosphoryl diester phosphodiesterase
413	COG0586	Uncharacterized membrane-associated protein
414	COG0587	DNA polymerase III
415	COG0589	Universal stress protein UspA and related nucleotide-binding proteins

416	COG0591	Na+/proline symporter
417	COG0592	DNA polymerase sliding clamp subunit (PCNA homolog)
418	COG0593	ATPase involved in DNA replication initiation
419	COG0594	RNase P protein component
420	COG0596	Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)
421	COG0597	Lipoprotein signal peptidase
422	COG0598	Mg2+ and Co2+ transporters
423	COG0601	ABC-type dipeptide/oligopeptide/nickel transport systems
424	COG0602	Organic radical activating enzymes
425	COG0603	Predicted PP-loop superfamily ATPase
426	COG0604	NADPH:quinone reductase and related Zn-dependent oxidoreductases
427	COG0605	Superoxide dismutase
428	COG0607	Rhodanese-related sulfurtransferase
429	COG0608	Single-stranded DNA-specific exonuclease
430	COG0609	ABC-type Fe3+-siderophore transport system
431	COG0612	Predicted Zn-dependent peptidases
432	COG0614	ABC-type Fe3+-hydroxamate transport system
433	COG0617	tRNA nucleotidyltransferase/poly(A) polymerase
434	COG0620	Methionine synthase II (cobalamin-independent)
435	COG0621	2-methylthioadenine synthetase
436	COG0622	Predicted phosphoesterase
437	COG0624	Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase and related deacylases
438	COG0626	Cystathionine beta-lyases/cystathionine gamma-synthases
439	COG0628	Predicted permease
440	COG0629	Single-stranded DNA-binding protein
441	COG0632	Holliday junction resolvasome
442	COG0634	Hypoxanthine-guanine phosphoribosyltransferase
443	COG0635	Coproporphyrinogen III oxidase and related Fe-S oxidoreductases
444	COG0636	F0F1-type ATP synthase
445	COG0637	Predicted phosphatase/phosphohexomutase

446	COG0640	Predicted transcriptional regulators
447	COG0642	Signal transduction histidine kinase
448	COG0646	Methionine synthase I (cobalamin-dependent)
449	COG0648	Endonuclease IV
450	COG0652	Peptidyl-prolyl cis-trans isomerase (rotamase) - cyclophilin family
451	COG0653	Preprotein translocase subunit SecA (ATPase
452	COG0654	2-polyprenyl-6-methoxyphenol hydroxylase and related FAD-dependent oxidoreductases
453	COG0655	Multimeric flavodoxin WrbA
454	COG0657	Esterase/lipase
455	COG0658	Predicted membrane metal-binding protein
456	COG0664	cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases
457	COG0667	Predicted oxidoreductases (related to aryl-alcohol dehydrogenases)
458	COG0668	Small-conductance mechanosensitive channel
459	COG0669	Phosphopantetheine adenylyltransferase
460	COG0670	Integral membrane protein
461	COG0671	Membrane-associated phospholipid phosphatase
462	COG0673	Predicted dehydrogenases and related proteins
463	COG0674	Pyruvate:ferredoxin oxidoreductase and related 2-oxoacid:ferredoxin oxidoreductases
464	COG0677	UDP-N-acetyl-D-mannosaminuronate dehydrogenase
465	COG0679	Predicted permeases
466	COG0681	Signal peptidase I
467	COG0682	Prolipoprotein diacylglyceryltransferase
468	COG0685	10-methylenetetrahydrofolate reductase
469	COG0687	Spermidine/putrescine-binding periplasmic protein
470	COG0690	Preprotein translocase subunit SecE
471	COG0691	tmRNA-binding protein
472	COG0692	Uracil DNA glycosylase
473	COG0693	Putative intracellular protease/amidase
474	COG0694	Thioredoxin-like proteins and domains

475	COG0695	Glutaredoxin and related proteins
476	COG0696	Phosphoglyceromutase
477	COG0697	Permeases of the drug/metabolite transporter (DMT) superfamily
478	COG0702	Predicted nucleoside-diphosphate-sugar epimerases
479	COG0703	Shikimate kinase
480	COG0704	Phosphate uptake regulator
481	COG0705	Uncharacterized membrane protein (homolog of Drosophila rhomboid)
482	COG0706	Preprotein translocase subunit YidC
483	COG0707	UDP-N-acetylglucosamine:LPS N-acetylglucosamine transferase
484	COG0711	F0F1-type ATP synthase
485	COG0712	F0F1-type ATP synthase
486	COG0714	MoxR-like ATPases
487	COG0718	Uncharacterized protein conserved in bacteria
488	COG0720	6-pyruvoyl-tetrahydropterin synthase
489	COG0725	ABC-type molybdate transport system
490	COG0726	Predicted xylanase/chitin deacetylase
491	COG0730	Predicted permeases
492	COG0735	Fe2+/Zn2+ uptake regulation proteins
493	COG0736	Phosphopantetheinyl transferase (holo-ACP synthase)
494	COG0737	5'-nucleotidase/2'
495	COG0739	Membrane proteins related to metalloendopeptidases
496	COG0740	Protease subunit of ATP-dependent Clp proteases
497	COG0742	N6-adenine-specific methylase
498	COG0744	Membrane carboxypeptidase (penicillin-binding protein)
499	COG0745	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
500	COG0746	Molybdopterin-guanine dinucleotide biosynthesis protein A
501	COG0747	ABC-type dipeptide transport system
502	COG0749	DNA polymerase I - 3'-5' exonuclease and polymerase domains
503	COG0750	Predicted membrane-associated Zn-dependent proteases 1

504	COG0753	Catalase
505	COG0755	ABC-type transport system involved in cytochrome c biogenesis
506	COG0758	Predicted Rossmann fold nucleotide-binding protein involved in DNA uptake
507	COG0760	Parvulin-like peptidyl-prolyl isomerase
508	COG0762	Predicted integral membrane protein
509	COG0764	3-hydroxymyristoyl/3-hydroxydecanoyl-(acyl carrier protein) dehydratases
510	COG0765	ABC-type amino acid transport system
511	COG0766	UDP-N-acetylglucosamine enolpyruvyl transferase
512	COG0768	Cell division protein FtsI/penicillin-binding protein 2
513	COG0769	UDP-N-acetylmuramyl tripeptide synthase
514	COG0770	UDP-N-acetylmuramyl pentapeptide synthase
515	COG0771	UDP-N-acetylmuramoylalanine-D-glutamate ligase
516	COG0772	Bacterial cell division membrane protein
517	COG0773	UDP-N-acetylmuramate-alanine ligase
518	COG0775	Nucleoside phosphorylase
519	COG0776	Bacterial nucleoid DNA-binding protein
520	COG0777	Acetyl-CoA carboxylase beta subunit
521	COG0778	Nitroreductase
522	COG0779	Uncharacterized protein conserved in bacteria
523	COG0780	Enzyme related to GTP cyclohydrolase I
524	COG0781	Transcription termination factor
525	COG0782	Transcription elongation factor
526	COG0783	DNA-binding ferritin-like protein (oxidative damage protectant)
527	COG0786	Na+/glutamate symporter
528	COG0787	Alanine racemase
529	COG0789	Predicted transcriptional regulators
530	COG0793	Periplasmic protease
531	COG0794	Predicted sugar phosphate isomerase involved in capsule formation
532	COG0796	Glutamate racemase
533	COG0799	Uncharacterized homolog of plant Iojap protein

534	COG0801	8-dihydro-6-hydroxymethylpterin-pyrophosphokinase
535	COG0802	Predicted ATPase or kinase
536	COG0805	Sec-independent protein secretion pathway component TatC
537	COG0806	RimM protein
538	COG0807	GTP cyclohydrolase II
539	COG0809	S-adenosylmethionine:tRNA-ribosyltransferase-isomerase (queuine synthetase)
540	COG0812	UDP-N-acetylmuramate dehydrogenase
541	COG0813	Purine-nucleoside phosphorylase
542	COG0816	Predicted endonuclease involved in recombination (possible Holliday junction resolvase in Mycoplasmas and B. subtilis)
543	COG0818	Diacylglycerol kinase
544	COG0820	Predicted Fe-S-cluster redox enzyme
545	COG0822	NifU homolog involved in Fe-S cluster formation
546	COG0824	Predicted thioesterase
547	COG0825	Acetyl-CoA carboxylase alpha subunit
548	COG0826	Collagenase and related proteases
549	COG0828	Ribosomal protein S21
550	COG0834	ABC-type amino acid transport/signal transduction systems
551	COG0841	Cation/multidrug efflux pump
552	COG0842	ABC-type multidrug transport system
553	COG0846	NAD-dependent protein deacetylases
554	COG0847	DNA polymerase III
555	COG0849	Actin-like ATPase involved in cell division
556	COG0858	Ribosome-binding factor A
557	COG0860	N-acetylmuramoyl-L-alanine amidase
558	COG0861	Membrane protein TerC
559	COG1009	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na+/H+ antiporter
560	COG1011	Predicted hydrolase (HAD superfamily)
561	COG1012	NAD-dependent aldehyde dehydrogenases
562	COG1013	Pyruvate:ferredoxin oxidoreductase and related 2-oxoacid:ferredoxin oxidoreductases

563	COG1014	Pyruvate:ferredoxin oxidoreductase and related 2-oxoacid:ferredoxin oxidoreductases
564	COG1015	Phosphopentomutase
565	COG1017	Hemoglobin-like flavoprotein
566	COG1018	Flavodoxin reductases (ferredoxin-NADPH reductases) family 1
567	COG1020	Non-ribosomal peptide synthetase modules and related proteins
568	COG1024	Enoyl-CoA hydratase/carnithine racemase
569	COG1028	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)
570	COG1040	Predicted amidophosphoribosyltransferases
571	COG1045	Serine acetyltransferase
572	COG1048	Aconitase A
573	COG1052	Lactate dehydrogenase and related dehydrogenases
574	COG1053	Succinate dehydrogenase/fumarate reductase
575	COG1054	Predicted sulfurtransferase
576	COG1055	Na+/H+ antiporter NhaD and related arsenite permeases
577	COG1057	Nicotinic acid mononucleotide adenylyltransferase
578	COG1058	Predicted nucleotide-utilizing enzyme related to molybdopterin-biosynthesis enzyme MoeA
579	COG1061	DNA or RNA helicases of superfamily II
580	COG1063	Threonine dehydrogenase and related Zn-dependent dehydrogenases
581	COG1066	Predicted ATP-dependent serine protease
582	COG1074	ATP-dependent exoDNAse (exonuclease V) beta subunit (contains helicase and exonuclease domains)
583	COG1080	Phosphoenolpyruvate-protein kinase (PTS system EI component in bacteria)
584	COG1092	Predicted SAM-dependent methyltransferases
585	COG1104	Cysteine sulfinate desulfinase/cysteine desulfurase and related enzymes
586	COG1105	Fructose-1-phosphate kinase and related fructose-6-phosphate kinase (PfkB)
587	COG1108	ABC-type Mn2+/Zn2+ transport systems
588	COG1109	Phosphomannomutase
589	COG1114	Branched-chain amino acid permeases
590	COG1117	ABC-type phosphate transport system
591	COG1120	ABC-type cobalamin/Fe3+-siderophores transport systems

592	COG1121	ABC-type Mn/Zn transport systems
593	COG1123	ATPase components of various ABC-type transport systems
594	COG1126	ABC-type polar amino acid transport system
595	COG1131	ABC-type multidrug transport system
596	COG1132	ABC-type multidrug transport system
597	COG1135	ABC-type metal ion transport system
598	COG1136	ABC-type antimicrobial peptide transport system
599	COG1158	Transcription termination factor
600	COG1159	GTPase
601	COG1160	Predicted GTPases
602	COG1162	Predicted GTPases
603	COG1165	2-succinyl-6-hydroxy-2
604	COG1167	Transcriptional regulators containing a DNA-binding HTH domain and an aminotransferase domain (MocR family) and their eukaryotic orthologs
605	COG1169	Isochorismate synthase
606	COG1171	Threonine dehydratase
607	COG1173	ABC-type dipeptide/oligopeptide/nickel transport systems
608	COG1175	ABC-type sugar transport systems
609	COG1176	ABC-type spermidine/putrescine transport system
610	COG1177	ABC-type spermidine/putrescine transport system
611	COG1179	Dinucleotide-utilizing enzymes involved in molybdopterin and thiamine biosynthesis family 1
612	COG1180	Pyruvate-formate lyase-activating enzyme
613	COG1181	D-alanine-D-alanine ligase and related ATP-grasp enzymes
614	COG1185	Polyribonucleotide nucleotidyltransferase (polynucleotide phosphorylase)
615	COG1186	Protein chain release factor B
616	COG1187	16S rRNA uridine-516 pseudouridylate synthase and related pseudouridylate synthases
617	COG1188	Ribosome-associated heat shock protein implicated in the recycling of the 50S subunit (S4 paralog)
618	COG1190	Lysyl-tRNA synthetase (class II)
619	COG1191	DNA-directed RNA polymerase specialized sigma subunit

620	COG1194	A/G-specific DNA glycosylase
621	COG1195	Recombinational DNA repair ATPase (RecF pathway)
622	COG1197	Transcription-repair coupling factor (superfamily II helicase)
623	COG1198	Primosomal protein N' (replication factor Y) - superfamily II helicase
624	COG1199	Rad3-related DNA helicases
625	COG1200	RecG-like helicase
626	COG1207	N-acetylglucosamine-1-phosphate uridyltransferase (contains nucleotidyltransferase and I-patch acetyltransferase domains)
627	COG1210	UDP-glucose pyrophosphorylase
628	COG1211	4-diphosphocytidyl-2-methyl-D-erithritol synthase
629	COG1214	Inactive homolog of metal-dependent proteases
630	COG1217	Predicted membrane GTPase involved in stress response
631	COG1219	ATP-dependent protease Clp
632	COG1220	ATP-dependent protease HslVU (ClpYQ)
633	COG1225	Peroxiredoxin
634	COG1242	Predicted Fe-S oxidoreductase
635	COG1249	Pyruvate/2-oxoglutarate dehydrogenase complex
636	COG1250	3-hydroxyacyl-CoA dehydrogenase
637	COG1252	NADH dehydrogenase
638	COG1253	Hemolysins and related proteins containing CBS domains
639	COG1254	Acylphosphatases
640	COG1263	Phosphotransferase system IIC components
641	COG1264	Phosphotransferase system IIB components
642	COG1271	Cytochrome bd-type quinol oxidase
643	COG1272	Predicted membrane protein
644	COG1278	Cold shock proteins
645	COG1279	Lysine efflux permease
646	COG1281	Disulfide bond chaperones of the HSP33 family
647	COG1286	Uncharacterized membrane protein
648	COG1289	Predicted membrane protein

649	COG1292	Choline-glycine betaine transporter
650	COG1294	Cytochrome bd-type quinol oxidase
651	COG1295	Predicted membrane protein
652	COG1296	Predicted branched-chain amino acid permease (azaleucine resistance)
653	COG1299	Phosphotransferase system
654	COG1301	Na+/H+-dicarboxylate symporters
655	COG1304	L-lactate dehydrogenase (FMN-dependent) and related alpha-hydroxy acid dehydrogenases
656	COG1309	Transcriptional regulator
657	COG1314	Preprotein translocase subunit SecG
658	COG1327	Predicted transcriptional regulator
659	COG1328	Oxygen-sensitive ribonucleoside-triphosphate reductase
660	COG1335	Amidases related to nicotinamidase
661	COG1346	Putative effector of murein hydrolase
662	COG1349	Transcriptional regulators of sugar metabolism
663	COG1380	Putative effector of murein hydrolase LrgA
664	COG1381	Recombinational DNA repair protein (RecF pathway)
665	COG1385	Uncharacterized protein conserved in bacteria
666	COG1387	Histidinol phosphatase and related hydrolases of the PHP family
667	COG1393	Arsenate reductase and related proteins
668	COG1396	Predicted transcriptional regulators
669	COG1399	Predicted metal-binding
670	COG1409	Predicted phosphohydrolases
671	COG1426	Uncharacterized protein conserved in bacteria
672	COG1434	Uncharacterized conserved protein
673	COG1435	Thymidine kinase
674	COG1438	Arginine repressor
675	COG1445	Phosphotransferase system fructose-specific component IIB
676	COG1447	Phosphotransferase system cellobiose-specific component IIA
677	COG1454	Alcohol dehydrogenase
678	COG1455	Phosphotransferase system cellobiose-specific component IIC

679	COG1459	Type II secretory pathway
680	COG1464	ABC-type metal ion transport system
681	COG1466	DNA polymerase III
682	COG1475	Predicted transcriptional regulators
683	COG1482	Phosphomannose isomerase
684	COG1488	Nicotinic acid phosphoribosyltransferase
685	COG1490	D-Tyr-tRNAtyr deacylase
686	COG1496	Uncharacterized conserved protein
687	COG1502	Phosphatidylserine/phosphatidylglycerophosphate/cardiolipin synthases and related enzymes
688	COG1526	Uncharacterized protein required for formate dehydrogenase activity
689	COG1528	Ferritin-like protein
690	COG1534	Predicted RNA-binding protein containing KH domain
691	COG1539	Dihydroneopterin aldolase
692	COG1540	Uncharacterized proteins
693	COG1544	Ribosome-associated protein Y (PSrp-1)
694	COG1555	DNA uptake protein and related DNA-binding proteins
695	COG1566	Multidrug resistance efflux pump
696	COG1570	Exonuclease VII
697	COG1575	4-dihydroxy-2-naphthoate octaprenyltransferase
698	COG1587	Uroporphyrinogen-III synthase
699	COG1589	Cell division septal protein
700	COG1595	DNA-directed RNA polymerase specialized sigma subunit
701	COG1600	Uncharacterized Fe-S protein
702	COG1605	Chorismate mutase
703	COG1607	Acyl-CoA hydrolase
704	COG1609	Transcriptional regulators
705	COG1611	Predicted Rossmann fold nucleotide-binding protein
706	COG1648	Siroheme synthase (precorrin-2 oxidase/ferrochelatase domain)
707	COG1651	Protein-disulfide isomerase
708	COG1654	Biotin operon repressor

709	COG1660	Predicted P-loop-containing kinase
710	COG1670	Acetyltransferases
711	COG1671	Uncharacterized protein conserved in bacteria
712	COG1674	DNA segregation ATPase FtsK/SpoIIIE and related proteins
713	COG1686	D-alanyl-D-alanine carboxypeptidase
714	COG1694	Predicted pyrophosphatase
715	COG1695	Predicted transcriptional regulators
716	COG1702	Phosphate starvation-inducible protein PhoH
717	COG1705	Muramidase (flagellum-specific)
718	COG1722	Exonuclease VII small subunit
719	COG1733	Predicted transcriptional regulators
720	COG1737	Transcriptional regulators
721	COG1738	Uncharacterized conserved protein
722	COG1739	Uncharacterized conserved protein
723	COG1758	DNA-directed RNA polymerase
724	COG1760	L-serine deaminase
725	COG1762	Phosphotransferase system mannitol/fructose-specific IIA domain (Ntr-type)
726	COG1763	Molybdopterin-guanine dinucleotide biosynthesis protein
727	COG1792	Cell shape-determining protein
728	COG1801	Uncharacterized conserved protein
729	COG1802	Transcriptional regulators
730	COG1806	Uncharacterized protein conserved in bacteria
731	COG1820	N-acetylglucosamine-6-phosphate deacetylase
732	COG1823	Predicted Na+/dicarboxylate symporter
733	COG1825	Ribosomal protein L25 (general stress protein Ctc)
734	COG1826	Sec-independent protein secretion pathway components
735	COG1841	Ribosomal protein L30/L7E
736	COG1846	Transcriptional regulators
737	COG1854	LuxS protein involved in autoinducer AI2 synthesis
738	COG1862	Preprotein translocase subunit YajC

739	COG1866	Phosphoenolpyruvate carboxykinase (ATP)
740	COG1869	ABC-type ribose transport system
741	COG1882	Pyruvate-formate lyase
742	COG1893	Ketopantoate reductase
743	COG1896	Predicted hydrolases of HD superfamily
744	COG1902	NADH:flavin oxidoreductases
745	COG1918	Fe2+ transport system protein A
746	COG1922	Teichoic acid biosynthesis proteins
747	COG1923	Uncharacterized host factor I protein
748	COG1925	Phosphotransferase system
749	COG1929	Glycerate kinase
750	COG1940	Transcriptional regulator/sugar kinase
751	COG1943	Transposase and inactivated derivatives
752	COG1947	4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate synthase
753	COG1956	GAF domain-containing protein
754	COG1959	Predicted transcriptional regulator
755	COG1960	Acyl-CoA dehydrogenases
756	COG1961	Site-specific recombinases
757	COG1968	Uncharacterized bacitracin resistance protein
758	COG1970	Large-conductance mechanosensitive channel
759	COG1972	Nucleoside permease
760	COG1974	SOS-response transcriptional repressors (RecA-mediated autopeptidases)
761	COG1977	Molybdopterin converting factor
762	COG1982	Arginine/lysine/ornithine decarboxylases
763	COG1984	Allophanate hydrolase subunit 2
764	COG1985	Pyrimidine reductase
765	COG1989	Type II secretory pathway
766	COG2003	DNA repair proteins
767	COG2008	Threonine aldolase
768	COG2009	Succinate dehydrogenase/fumarate reductase

769	COG2011	ABC-type metal ion transport system
770	COG2030	Acyl dehydratase
771	COG2049	Allophanate hydrolase subunit 1
772	COG2050	Uncharacterized protein
773	COG2077	Peroxiredoxin
774	COG2081	Predicted flavoproteins
775	COG2103	Predicted sugar phosphate isomerase
776	COG2116	Formate/nitrite family of transporters
777	COG2137	Uncharacterized protein conserved in bacteria
778	COG2165	Type II secretory pathway
779	COG2171	Tetrahydrodipicolinate N-succinyltransferase
780	COG2182	Maltose-binding periplasmic proteins/domains
781	COG2183	Transcriptional accessory protein
782	COG2188	Transcriptional regulators
783	COG2190	Phosphotransferase system IIA components
784	COG2195	Di- and tripeptidases
785	COG2197	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain
786	COG2199	FOG: GGDEF domain
787	COG2205	Osmosensitive K+ channel histidine kinase
788	COG2207	AraC-type DNA-binding domain-containing proteins
789	COG2213	Phosphotransferase system
790	COG2217	Cation transport ATPase
791	COG2226	Methylase involved in ubiquinone/menaquinone biosynthesis
792	COG2233	Xanthine/uracil permeases
793	COG2244	Membrane protein involved in the export of O-antigen and teichoic acid
794	COG2252	Permeases
795	COG2255	Holliday junction resolvasome
796	COG2256	ATPase related to the helicase subunit of the Holliday junction resolvase
797	COG2259	Predicted membrane protein
798	COG2262	GTPases

799	COG2264	Ribosomal protein L11 methylase
800	COG2265	SAM-dependent methyltransferases related to tRNA (uracil-5-)-methyltransferase
801	COG2267	Lysophospholipase
802	COG2271	Sugar phosphate permease
803	COG2333	Predicted hydrolase (metallo-beta-lactamase superfamily)
804	COG2363	Uncharacterized small membrane protein
805	COG2390	Transcriptional regulator
806	COG2391	Predicted transporter component
807	COG2501	Uncharacterized conserved protein
808	COG2606	Uncharacterized conserved protein
809	COG2610	H+/gluconate symporter and related permeases
810	COG2723	Beta-glucosidase/6-phospho-beta-glucosidase/beta-galactosidase
811	COG2771	DNA-binding HTH domain-containing proteins
812	COG2804	Type II secretory pathway
813	COG2812	DNA polymerase III
814	COG2813	16S RNA G1207 methylase RsmC
815	COG2818	3-methyladenine DNA glycosylase
816	COG2827	Predicted endonuclease containing a URI domain
817	COG2832	Uncharacterized protein conserved in bacteria
818	COG2837	Predicted iron-dependent peroxidase
819	COG2890	Methylase of polypeptide chain release factors
820	COG2891	Cell shape-determining protein
821	COG2896	Molybdenum cofactor biosynthesis enzyme
822	COG2919	Septum formation initiator
823	COG2962	Predicted permeases
824	COG2963	Transposase and inactivated derivatives
825	COG2966	Uncharacterized conserved protein
826	COG2978	Putative p-aminobenzoyl-glutamate transporter
827	COG3001	Fructosamine-3-kinase
828	COG3010	Putative N-acetylmannosamine-6-phosphate epimerase

829	COG3027	Uncharacterized protein conserved in bacteria
830	COG3037	Uncharacterized protein conserved in bacteria
831	COG3070	Regulator of competence-specific genes
832	COG3091	Uncharacterized protein conserved in bacteria
833	COG3104	Dipeptide/tripeptide permease
834	COG3152	Predicted membrane protein
835	COG3153	Predicted acetyltransferase
836	COG3221	ABC-type phosphate/phosphonate transport system
837	COG3275	Putative regulator of cell autolysis
838	COG3279	Response regulator of the LytR/AlgR family
839	COG3307	Lipid A core - O-antigen ligase and related enzymes
840	COG3414	Phosphotransferase system
841	COG3610	Uncharacterized conserved protein
842	COG3833	ABC-type maltose transport systems
843	COG3839	ABC-type sugar transport systems
844	COG3842	ABC-type spermidine/putrescine transport systems
845	COG4108	Peptide chain release factor RF-3
846	COG4123	Predicted O-methyltransferase
847	COG4148	ABC-type molybdate transport system
848	COG4149	ABC-type molybdate transport system
849	COG4166	ABC-type oligopeptide transport system
850	COG4536	Putative Mg2+ and Co2+ transporter CorB
851	COG4608	ABC-type oligopeptide transport system
852	COG4668	Mannitol/fructose-specific phosphotransferase system
853	COG4974	Site-specific recombinase XerD
854	COG4987	ABC-type transport system involved in cytochrome bd biosynthesis
855	COG4988	ABC-type transport system involved in cytochrome bd biosynthesis
856	COG4992	Ornithine/acetylornithine aminotransferase
857	COG5405	ATP-dependent protease HslVU (ClpYQ)

Appendix Table 3.5 - List of PPIs for *Eco*, *Ype*, *Vch*, and *Sau* based on Raja *et al* interactions for the conserved COGs from Table 3.4

proteins annotated as uncharacterized/predicted in *Eco*

Bait Eco	Prey Eco	Bait COG	Prey COG	Bait Vch	Prey Vch	Bait Ype	Prey Ype	Bait Sau	Prey Sau
P03004	P76514	COG0593	COG1896	Q9KVX6	Q9KQM0	Q8Z9U7	Q8ZDK3	Q5HJZ5	0
P0A9M2	Q79E92	COG0634	COG2963	Q9KUD7	Q9K344	Q8D1B3	Q8D1M5	Q5HIG5	0
P16703	P33014	COG0031	COG0425	Q9KUI4	Q9KVW4	Q7CJG9	Q7CKR7	Q5HIG2	0
P16703	P75745	COG0031	COG1984	Q9KUI4	Q9KLG6	Q7CJG9	Q7CJR2	Q5HIG2	0
Q46939	Q46939	COG0183	COG0183	Q9KT59	Q9KT59	Q8ZAM9	Q8ZAM9	Q5HIA0	Q5HIA0
P77218	P77174	COG0280	COG1475	Q9KT08	Q9KNG7	Q8D0X3	0	Q5HI88	0
P76340	P0ACZ8	COG0745	COG0745	Q9KU11	Q9KU11	Q8D168	Q8D168	Q5HI09	Q5HI09
P0A8F8	Q79E92	COG0556	COG2963	0	Q9K344	Q8ZGW7	Q8D1M5	Q5HHR0	0
P0A8F8	P77174	COG0556	COG1475	0	Q9KNG7	Q8ZGW7	0	Q5HHR0	0
P77650	P39409	COG0446	COG1180	Q9KQK5	Q9KQX8	Q7CJG4	Q8D043	Q5HHB4	Q5HJF3
P76569	P0ACJ8	COG1393	COG0664	Q9KQ54	P0C6D0	Q8D0Y4	Q8ZE82	Q5HH87	Q5HCR6
P76569	P0A9E5	COG1393	COG0664	Q9KQ54	P0C6D0	Q8D0Y4	Q8ZE82	Q5HH87	Q5HCR6
P76569	Q46864	COG1393	COG1396	Q9KQ54	Q9KVH0	Q8D0Y4	Q8CKM7	Q5HH87	0
P22634	P39409	COG0796	COG1180	Q9KVI7	Q9KQX8	Q8ZAA2	Q8D043	Q5HGT3	Q5HJF3
P0A7X6	P77174	COG0806	COG1475	Q9KUF9	Q9KNG7	Q8ZBU8	0	Q5HGJ3	0
P77308	P0AAG0	COG0601	COG0444	Q9KVH4	Q9KUA6	Q7CKF9	Q8D1C2	Q5HG38	Q5HG40
P00861	Q79E92	COG0019	COG2963	Q9KVL7	Q9K344	Q7CGZ1	Q8D1M5	Q5HG20	0
P17443	P0AGJ5	COG0707	COG0566	Q9KPG7	0	Q8ZIE9	0	Q5HG02	0
P07639	P33014	COG0337	COG0425	Q9KNV2	Q9KVW4	Q8ZJF6	Q7CKR7	Q5HFV8	0
P04994	P76168	COG1570	COG0582	Q9KTW4	Q9KUK2	Q8ZCU2	Q8D1C0	Q5HFP8	0
P0ABS5	Q79E92	COG0358	COG2963	0	Q9K344	Q7CGG5	Q8D1M5	Q5HFJ8	0
P06616	Q79E92	COG1159	COG2963	Q9KPB3	Q9K344	Q8ZD71	Q8D1M5	Q5HFJ3	0
P0AFW4	P77174	COG0782	COG1475	Q9KU89	Q9KNG7	Q7CL79	0	Q5HFF2	0

P0AED0	P37348	COG0589	COG1801	Q9KVR3	Q9KSU1	Q8ZE81	Q7CIB9	Q5HF64	0
P77615	P77615	COG1609	COG1609	Q9KTJ3	Q9KTJ3	Q7CGW3	Q7CGW3	Q5HF38	Q5HF38
P0ACN7	P77295	COG1609	COG0640	Q9KTJ3	0	Q7CGW3	0	Q5HF38	0
P36673	P33014	COG1609	COG0425	Q9KTJ3	0	Q7CGW3	0	Q5HF38	0
P0AED7	P33014	COG0624	COG0425	Q9KQ52	0	Q7CJI9	0	Q5HF23	0
P0AED7	P71239	COG0624	COG0463	Q9KQ52	0	Q7CJI9	0	Q5HF23	0
P07813	P75767	COG0495	COG0391	Q9KTE6	Q9KT82	Q8ZDF8	P58589	Q5HF16	0
P0AGB6	Q79E92	COG1595	COG2963	Q9KT60	Q9K344	Q7CJQ9	Q8D1M5	Q5HEZ8	0
P31552	P31552	COG0318	COG0318	H9L4Q1	H9L4Q1	Q8D0S3	Q8D0S3	Q5HEY2	Q5HEY2
P05852	Q79E92	COG0533	COG2963	0	Q9K344	Q74RQ9	Q8D1M5	Q5HEF2	0
P39358	P39358	COG0129	COG0129	Q9KVW0	0	Q8ZAB3	0	Q5HEE8	0
P07862	P33014	COG1181	COG0425	Q9KM17	Q9KVW4	Q8ZIE7	Q7CKR7	Q5HEB7	0
P0A749	P33014	COG0766	COG0425	Q9KP62	Q9KVW4	Q8ZB56	Q7CKR7	Q5HE76	0
P17169	P76093	COG0449	COG0671	Q9KUM8	Q9KND0	Q8Z9S8	Q8D031	Q5HE49	0
P07649	P33014	COG0101	COG0425	Q9KTA4	Q9KVW4	Q8ZD27	Q7CKR7	Q5HDY9	0
P0A7W7	P26646	COG0096	COG0604	Q9KNZ8	0	Q8ZJ98	0	Q5HDX2	0
P60422	P45759	COG0090	COG2804	Q9KNY7	Q9KUV7	P60436	Q8D1B8	Q5HDW1	0
P63284	P33014	COG0542	COG0425	Q9KU18	0	Q7CL29	0	Q5HD02	0
P0ACU0	P0ACU0	COG1309	COG1309	Q9KVJ2	Q9KVJ2	Q8ZJP6	Q8ZJP6	Q5HCN2	Q5HCN2
P0ACU0	P25534	COG1309	COG0654	Q9KVJ2	Q9KTD9	Q8ZJP6	Q7CGT0	Q5HCN2	0
P31666	P05827	COG0726	COG0583	Q9KSH6	Q9KVS1	Q8CLN7	P58510	Q5HCM9	0
P06988	P36879	COG0141	COG1131	Q9F854	Q9KUD3	Q8ZFX5	Q7CKD9	Q5HCL9	0
P42599	P0AG11	COG0673	COG1974	Q9KN66	Q9KVP9	Q8D131	Q8ZJ16	0	Q9L4P1
P0ADF8	P39409	COG0440	COG1180	Q9KP91	Q9KQX8	Q8D0H3	Q8D043	0	Q5HJF3
P0ADG4	P39409	COG0483	COG1180	Q9KTY5	Q9KQX8	Q7CJN9	Q8D043	0	Q5HJF3
P31828	P0AFJ5	COG0612	COG0745	Q9KUG7	Q9KU11	Q7CG15	Q8D168	0	Q5HI09
P39347	P29745	COG0582	COG2195	Q9KUK2	Q9KSB5	Q8D1C0	Q7CK62	0	Q5HHS7
P68688	P30750	COG0695	COG1135	Q9KSW0	Q9KTJ5	Q7CLB8	Q7CHF8	0	Q5HHK4
P63389	P60716	COG0488	COG0320	Q9KU31	Q9KTF9	Q8CK90	Q8ZDH0	0	Q5HHG0

P0ACM9	P31134	COG2188	COG3842	Q9KSQ0	Q9KUB3	Q9ZC75	Q74R28	0	Q5HGY5
P64564	P22634	COG0762	COG0796	Q9KUQ6	Q9KVI7	Q7CGR7	Q8ZAA2	0	Q5HGT3
P24203	P0AED0	COG0523	COG0589	Q9KM61	Q9KVR3	Q7CHD5	Q8ZE81	0	Q5HF64
P77174	P24242	COG1475	COG1609	Q9KNG7	Q9KTJ3	Q8ZFS3	Q7CGW3	0	Q5HF38
P76092	P15082	COG2267	COG1349	H9L4S2	Q9KUM9	Q7CKX6	Q7CGX6	0	Q5HE08
P0A8A0	P0AF03	COG0217	COG0521	0	Q9KT80	0	Q7CG72	0	Q5HDT2
P0AF34	P0ACJ8	COG3152	COG0664	Q9KNR3	P0C6D0	Q8D1T5	Q8ZE82	0	Q5HCR6
P77174	P0ACS9	COG1475	COG1309	Q9KNG7	Q9KVJ2	Q8ZFS3	Q8ZJP6	0	Q5HCN2
P77539	P77539	COG1063	COG1063	Q9KL62	Q9KL62	Q8ZJN2	Q8ZJN2	0	0
P17115	Q79E92	COG0517	COG2963	Q9KTZ3	Q9K344	Q8D1Q8	Q8D1M5	0	0
P75824	Q79E92	COG1018	COG2963	Q9KL25	Q9K344	Q8D039	Q8D1M5	0	0
P30235	Q79E92	COG0524	COG2963	Q9KSX7	Q9K344	Q7CJW4	Q8D1M5	0	0
P28630	Q79E92	COG1466	COG2963	Q9KTE9	Q9K344	Q7CJV3	Q8D1M5	0	0
P0AB01	Q79E92	COG1434	COG2963	Q9KRE4	Q9K344	Q7CJM7	Q8D1M5	0	0
P0ACS2	Q79E92	COG0789	COG2963	Q9KV79	Q9K344	Q7CIP9	Q8D1M5	0	0
P76186	Q79E92	COG1289	COG2963	Q9KLM5	Q9K344	Q7CHL6	Q8D1M5	0	0
P04993	Q79E92	COG0507	COG2963	Q9KPP7	Q9K344	Q7CH01	Q8D1M5	0	0
P21893	Q79E92	COG0608	COG2963	Q9KPF1	Q9K344	Q7CGT9	Q8D1M5	0	0
P75728	Q79E92	COG0654	COG2963	Q9KTD9	Q9K344	Q7CGT0	Q8D1M5	0	0
P0ACG8	Q79E92	COG1188	COG2963	Q9KNK3	Q9K344	Q7CFW8	Q8D1M5	0	0
P42599	P27862	COG0673	COG1739	Q9KN66	Q9KNI2	Q8D131	Q8D1H6	0	0
P0A8A0	P07000	COG0217	COG2267	0	H9L4S2	0	Q8D138	0	0
P39353	P39353	COG0673	COG0673	Q9KN66	Q9KN66	Q8D131	Q8D131	0	0
P77376	P77376	COG0673	COG0673	Q9KN66	Q9KN66	Q8D131	Q8D131	0	0
P76114	P37631	COG1802	COG2081	Q9KSC3	Q9KVQ9	Q8D1N9	Q8CZK6	0	0
P42599	Q46864	COG0673	COG1396	Q9KN66	Q9KVH0	Q8D131	Q8CKM7	0	0
P22256	P33014	COG0160	COG0425	Q9KLY6	Q9KVW4	Q8D0Y8	Q7CKR7	0	0
P0ACM9	P07014	COG2188	COG0479	Q9KSQ0	Q9KQB2	Q9ZC75	Q7CKM7	0	0
P05824	P76104	COG0497	COG0826	P0C6Q4	Q9KU72	Q8CZZ7	Q7CKH7	0	0

P0A8A0	P33358	COG0217	COG0789	0	Q9KV79	0	Q7CIP9	0	0
P37348	P21367	COG1801	COG1335	Q9KSU1	Q9KLN1	Q7CIB9	Q7CIH4	0	0
P37348	P37348	COG1801	COG1801	Q9KSU1	Q9KSU1	Q7CIB9	Q7CIB9	0	0
P76237	P32715	COG2199	COG1289	Q9KVR7	Q9KLM5	Q8CKZ6	Q7CHL6	0	0
P32669	P33919	COG0176	COG1061	Q9KLW8	Q9KTS6	Q8ZIN2	Q7CHC7	0	0
P63389	P0AEL3	COG0488	COG1918	Q9KU31	Q9KQC2	Q8CK90	Q7CFX2	0	0
P77174	P77174	COG1475	COG1475	Q9KNG7	Q9KNG7	Q8ZFS3	0	0	0
P75824	P77174	COG1018	COG1475	Q9KL25	Q9KNG7	Q8D039	0	0	0
P69816	P33014	COG1445	COG0425	Q9KR24	0	Q8CZM7	0	0	0
P68688	P77174	COG0695	COG1475	Q9KSW0	Q9KNG7	Q7CLB8	0	0	0
P0ACG8	P77174	COG1188	COG1475	Q9KNK3	Q9KNG7	Q7CFW8	0	0	0
P77493	P30750	COG0524	COG1135	Q9KSX7	Q9KTJ5	Q7CJW4	Q7CHF8	0	Q5HHK4
P0AGB6	Q46864	COG1595	COG1396	Q9KT60	Q9KVH0	Q7CJQ9	Q8CKM7	Q5HEZ8	0
P22634	P22634	COG0796	COG0796	Q9KVI7	Q9KVI7	Q8ZAA2	Q8ZAA2	Q5HGT3	Q5HGT3
P0AED0	P0AED0	COG0589	COG0589	Q9KVR3	Q9KVR3	Q8ZE81	Q8ZE81	Q5HF64	Q5HF64
P0AFW4	P30750	COG0782	COG1135	Q9KU89	Q9KTJ5	Q7CL79	Q7CHF8	Q5HFF2	Q5HHK4
P0ACM9	Q46864	COG2188	COG1396	Q9KSQ0	Q9KVH0	Q9ZC75	Q8CKM7	0	0
P0ACN7	Q46864	COG1609	COG1396	Q9KTJ3	Q9KVH0	Q7CGW3	Q8CKM7	Q5HF38	0
P0ACN7	P0ACJ8	COG1609	COG0664	Q9KTJ3	P0C6D0	Q7CGW3	Q8ZE82	Q5HF38	Q5HCR6
P0A7X6	P25534	COG0806	COG0654	Q9KUF9	Q9KTD9	Q8ZBU8	Q7CGT0	Q5HGJ3	0
P0ACG8	P0ACJ8	COG1188	COG0664	Q9KNK3	P0C6D0	Q7CFW8	Q8ZE82	0	Q5HCR6
P0ACG8	P07014	COG1188	COG0479	Q9KNK3	Q9KQB2	Q7CFW8	Q7CKM7	0	0
P06616	Q46864	COG1159	COG1396	Q9KPB3	Q9KVH0	Q8ZD71	Q8CKM7	Q5HFJ3	0
P06616	P24242	COG1159	COG1609	Q9KPB3	Q9KTJ3	Q8ZD71	Q7CGW3	Q5HFJ3	Q5HF38
P77174	P75824	COG1475	COG1018	Q9KNG7	Q9KL25	Q8ZFS3	Q8D039	0	0
P16703	P77737	COG0031	COG4608	Q9KUI4	Q9KUA7	Q7CJG9	Q7CIN0	Q5HIG2	Q5HG41
P0A8F8	P0AEL8	COG0556	COG2197	0	Q9KSP3	Q8ZGW7	Q7CHT5	Q5HHR0	Q5HDG5
P0A8F8	P0A8F8	COG0556	COG0556	0	0	Q8ZGW7	Q8ZGW7	Q5HHR0	Q5HHR0
P0A8F8	P28630	COG0556	COG1466	0	Q9KTE9	Q8ZGW7	Q7CJV3	Q5HHR0	0

P0AED7	P0A9N8	COG0624	COG0602	Q9KQ52	Q9KS94	Q7CJI9	Q7CKD0	Q5HF23	0
P0ACG8	P0ACG8	COG1188	COG1188	Q9KNK3	Q9KNK3	Q7CFW8	Q7CFW8	0	0
P0ACG8	P33030	COG1188	COG0523	Q9KNK3	Q9KM61	Q7CFW8	Q7CHD5	0	0
P0ACG8	P06710	COG1188	COG2812	Q9KNK3	Q9KT51	Q7CFW8	Q7CK10	0	0
P0AB01	P0ABB4	COG1434	COG0055	Q9KRE4	Q9KNH5	Q7CJM7	Q7CFM8	0	Q5HE97
P0A7X6	P00893	COG0806	COG0028	Q9KUF9	Q9KVV7	Q8ZBU8	Q7CL03	Q5HGJ3	0
P0A7W7	P60422	COG0096	COG0090	Q9KNZ8	Q9KNY7	Q8ZJ98	P60436	Q5HDX2	Q5HDW1
P00861	P0AE88	COG0019	COG0745	Q9KVL7	Q9KU11	Q7CGZ1	Q8D168	Q5HG20	Q5HI09
P0A7W7	P0A7V8	COG0096	COG0522	Q9KNZ8	Q9KP07	Q8ZJ98	Q8ZJ88	Q5HDX2	Q5HF54
P0A7W7	P0AG48	COG0096	COG0261	Q9KNZ8	Q9KUT0	Q8ZJ98	Q0WBD7	Q5HDX2	Q5HFB6
P0A7W7	P68679	COG0096	COG0828	Q9KNZ8	P66532	Q8ZJ98	P68686	Q5HDX2	Q5HFI5
P0A7W7	P0A7T7	COG0096	COG0238	Q9KNZ8	Q9KUZ0	Q8ZJ98	Q8ZB83	Q5HDX2	Q5HIS7
P07649	P0A805	COG0101	COG0233	Q9KTA4	Q9KPV5	Q8ZD27	Q8ZH63	Q5HDY9	Q5HGH2
P0A7W7	P0A7W7	COG0096	COG0096	Q9KNZ8	Q9KNZ8	Q8ZJ98	Q8ZJ98	Q5HDX2	Q5HDX2
P22634	P0AFW4	COG0796	COG0782	Q9KVI7	Q9KU89	Q8ZAA2	Q7CL79	Q5HGT3	Q5HFF2
P07639	P07639	COG0337	COG0337	Q9KNV2	Q9KNV2	Q8ZJF6	Q8ZJF6	Q5HFV8	Q5HFV8
P04993	P75679	COG0507	COG2963	Q9KPP7	Q9K344	Q7CH01	Q8D1M5	0	0
P05852	P05852	COG0533	COG0533	0	0	Q74RQ9	Q74RQ9	Q5HEF2	Q5HEF2
P05852	P76256	COG0533	COG1214	0	Q9KQK9	Q74RQ9	Q8D0F5	Q5HEF2	0
P75728	P75728	COG0654	COG0654	Q9KTD9	Q9KTD9	Q7CGT0	Q7CGT0	0	0
P68688	P0AC62	COG0695	COG0695	Q9KSW0	Q9KSW0	Q7CLB8	Q7CLB8	0	0
P22634	P0A8A4	COG0796	COG1806	Q9KVI7	Q9KKW4	Q8ZAA2	Q8ZDY4	Q5HGT3	Q5HFJ7
P22634	P37313	COG0796	COG4608	Q9KVI7	Q9KUA7	Q8ZAA2	Q7CIN0	Q5HGT3	Q5HG41
P06616	P0ACU2	COG1159	COG1309	Q9KPB3	Q9KVJ2	Q8ZD71	Q8ZJP6	Q5HFJ3	Q5HCN2
P28630	P10443	COG1466	COG0587	Q9KTE9	P52022	Q7CJV3	Q7CH18	0	Q5HF71
P28630	P03007	COG1466	COG0847	Q9KTE9	Q9KV58	Q7CJV3	Q8ZE08	0	Q5HFW8
P28630	P0AGJ2	COG1466	COG0566	Q9KTE9	Q9KTT5	Q7CJV3	Q8ZIV4	0	Q5HIE3
P28630	P17117	COG1466	COG0778	Q9KTE9	Q9KU15	Q7CJV3	Q7CIH7	0	Q5HIR4
P28630	P28631	COG1466	COG0470	Q9KTE9	Q9KQI3	Q7CJV3	O69170	0	0

P28630	P32718	COG1466	COG1940	Q9KTE9	Q9KV86	Q7CJV3	Q8CL97	0	0
P04994	P23857	COG1570	COG0607	Q9KTW4	Q9KVP1	Q8ZCU2	Q8CL37	Q5HFP8	0
P0AGB6	P0AGB6	COG1595	COG1595	Q9KT60	Q9KT60	Q7CJQ9	Q7CJQ9	Q5HEZ8	Q5HEZ8
P0AGB6	P18811	COG1595	COG1609	Q9KT60	Q9KTJ3	Q7CJQ9	Q7CGW3	Q5HEZ8	Q5HF38
P30749	Q79E92	COG0314	COG2963	Q9KT77	Q9K344	Q8ZGW2	Q8D1M5	Q5HDT6	0
P22188	Q79E92	COG0769	COG2963	Q9X6N4	Q9K344	Q8ZIF4	Q8D1M5	O86491	0
P0A8D3	P39409	#N/A	COG1180	#N/A	Q9KQX8	#N/A	Q8D043	#N/A	Q5HJF3
P05523	P39409	COG0266	COG1180	Q9KVC5	Q9KQX8	Q8ZJP0	Q8D043	0	Q5HJF3
P23893	Q46864	COG0001	COG1396	Q9KU97	Q9KVH0	Q8ZBL9	Q8CKM7	Q5HER0	0
P00893	Q46864	COG0028	COG1396	Q9KVV7	Q9KVH0	Q7CL03	Q8CKM7	0	0
P08312	Q46864	COG0016	COG1396	Q9KSN7	Q9KVH0	Q8ZDX0	Q8CKM7	Q5HGU6	0
P0A7M9	Q46864	COG0254	COG1396	Q9KTM4	Q9KVH0	P58472	Q8CKM7	Q5HE80	0
P30748	Q46864	COG1977	COG1396	Q9KT78	Q9KVH0	Q7CH70	Q8CKM7	0	0
P45544	P0ACJ8	COG2188	COG0664	Q9KSQ0	P0C6D0	Q9ZC75	Q8ZE82	0	Q5HCR6
Q46864	Q46864	COG1396	COG1396	Q9KVH0	Q9KVH0	Q8CKM7	Q8CKM7	0	0
P0AD57	Q46864	COG0142	COG1396	Q9KUT1	Q9KVH0	Q7CK38	Q8CKM7	0	0
P69222	Q46864	COG0361	COG1396	P65128	Q9KVH0	P65115	Q8CKM7	Q5HDY0	0
P77444	P0ACJ8	COG0520	COG0664	Q9KSS2	P0C6D0	Q8D0M6	Q8ZE82	Q5HHH0	Q5HCR6
P42641	Q46864	COG0536	COG1396	Q9KUS8	Q9KVH0	Q7CKJ6	Q8CKM7	Q5HFB9	0
P0ABH9	Q46864	COG0542	COG1396	Q9KU18	Q9KVH0	Q7CL29	Q8CKM7	Q5HD02	0
P07117	Q46864	COG0591	COG1396	Q9KV54	Q9KVH0	Q7CI08	Q8CKM7	Q5HEM0	0
P69931	Q46864	COG0593	COG1396	Q9KVX6	Q9KVH0	Q8Z9U7	Q8CKM7	Q5HJZ5	0
P52062	P0ACJ8	COG0635	COG0664	Q9KVM5	P0C6D0	Q8CZV2	Q8ZE82	0	Q5HCR6
P0A9A2	Q46864	COG1528	COG1396	Q9KVR1	Q9KVH0	Q8D099	Q8CKM7	Q5HEN0	0
P0ACQ0	Q46864	COG1609	COG1396	Q9KTJ3	Q9KVH0	Q7CGW3	Q8CKM7	Q5HF38	0
P0AFI5	Q46864	COG1686	COG1396	Q9KTF5	Q9KVH0	Q7CJV7	Q8CKM7	0	0
P29745	Q46864	COG2195	COG1396	Q9KSB5	Q9KVH0	Q7CK62	Q8CKM7	Q5HHS7	0
P69407	Q46864	COG2197	COG1396	Q9KSP3	Q9KVH0	Q7CHT5	Q8CKM7	Q5HDG5	0
P0ABK5	P30750	COG0031	COG1135	Q9KUI4	Q9KTJ5	Q7CJG9	Q7CHF8	Q5HIG2	Q5HHK4

P07395	P32715	COG0072	COG1289	Q9KSN6	Q9KLM5	Q8ZDX1	Q7CHL6	Q5HGU5	0
P77460	P0AG11	COG0500	COG1974	Q9KV96	Q9KVP9	Q7CH29	Q8ZJ16	0	Q9L4P1
P45771	P30750	COG0551	COG1135	Q9KVT9	Q9KTJ5	Q8CZI7	Q7CHF8	0	Q5HHK4
P0AFT5	P30750	COG3279	COG1135	Q9KU36	Q9KTJ5	Q8ZBV2	Q7CHF8	Q5HJB5	Q5HHK4
P60664	P30750	COG0107	COG1135	Q9KSW8	Q9KTJ5	Q8ZFY0	Q7CHF8	Q5HCM4	Q5HHK4
P76256	P0ACS9	COG1214	COG1309	Q9KQK9	Q9KVJ2	Q8D0F5	Q8ZJP6	0	Q5HCN2
P77806	P30750	COG0436	COG1135	Q9KQM1	Q9KTJ5	Q7CJE7	Q7CHF8	0	Q5HHK4
P0A858	P29745	COG0149	COG2195	Q9KNR1	Q9KSB5	Q8ZJK9	Q7CK62	Q5HHP3	Q5HHS7
P0A7M9	P0AFJ5	COG0254	COG0745	Q9KTM4	Q9KU11	P58472	Q8D168	Q5HE80	Q5HI09
P0ABQ4	P07000	COG0262	COG2267	Q9KUS5	H9L4S2	Q7CG83	Q8D138	Q5HFZ7	0
P05523	P24242	COG0266	COG1609	Q9KVC5	Q9KTJ3	Q8ZJP0	Q7CGW3	0	Q5HF38
P0A6I0	P07000	COG0283	COG2267	Q9KQT2	H9L4S2	Q8ZGB3	Q8D138	Q5HFU6	0
P30749	P0AG11	COG0314	COG1974	Q9KT77	Q9KVP9	Q8ZGW2	Q8ZJ16	Q5HDT6	Q9L4P1
P60716	P60716	COG0320	COG0320	Q9KTF9	Q9KTF9	Q8ZDH0	Q8ZDH0	Q5HHG0	Q5HHG0
P77434	P30750	COG0436	COG1135	Q9KQM1	Q9KTJ5	Q7CJE7	Q7CHF8	0	Q5HHK4
P25522	P25534	COG0486	COG0654	Q9KVY5	Q9KTD9	Q8Z9U2	Q7CGT0	Q5HCI3	0
P0AEI6	P30750	COG0494	COG1135	Q9KU53	Q9KTJ5	Q8ZHU8	Q7CHF8	0	Q5HHK4
P77444	P30750	COG0520	COG1135	Q9KSS2	Q9KTJ5	Q8D0M6	Q7CHF8	Q5HHH0	Q5HHK4
P0AEW6	P0AEW6	COG0524	COG0524	Q9KSX7	Q9KSX7	Q7CJW4	Q7CJW4	0	0
P10908	P0AG11	COG0584	COG1974	Q9KRT2	Q9KVP9	Q7CK63	Q8ZJ16	0	Q9L4P1
P0AAB8	P0AED0	COG0589	COG0589	Q9KVR3	Q9KVR3	Q8ZE81	Q8ZE81	Q5HF64	Q5HF64
P0AFJ5	P0AFJ5	COG0745	COG0745	Q9KU11	Q9KU11	Q8D168	Q8D168	Q5HI09	Q5HI09
P0ACZ8	P0ACZ8	COG0745	COG0745	Q9KU11	Q9KU11	Q8D168	Q8D168	Q5HI09	Q5HI09
P0A9Q5	P0AG11	COG0777	COG1974	Q9KTA3	Q9KVP9	Q0WDC3	Q8ZJ16	Q5HF73	Q9L4P1
P15047	P30750	COG1028	COG1135	Q56632	Q9KTJ5	Q8D0J5	Q7CHF8	0	Q5HHK4
P0AC41	P07014	COG1053	COG0479	Q9KQB1	Q9KQB2	Q7CKM6	Q7CKM7	0	0
P31135	P31134	COG1176	COG3842	Q9KS34	Q9KUB3	Q8D0Y7	Q74R28	0	Q5HGY5
P0A968	P0AAG0	COG1278	COG0444	Q9KSW4	Q9KUA6	Q8D1N6	Q8D1C2	Q5HG18	Q5HG40
P32715	P32715	COG1289	COG1289	Q9KLM5	Q9KLM5	Q7CHL6	Q7CHL6	0	0

P0ACS9	P0ACS9	COG1309	COG1309	Q9KVJ2	Q9KVJ2	Q8ZJP6	Q8ZJP6	Q5HCN2	Q5HCN2
P24178	P30750	COG1393	COG1135	Q9KQ54	Q9KTJ5	Q8D0Y4	Q7CHF8	Q5HH87	Q5HHK4
Q46864	P0ACS9	COG1396	COG1309	Q9KVH0	Q9KVJ2	Q8CKM7	Q8ZJP6	0	Q5HCN2
Q46864	P0A9E5	COG1396	COG0664	Q9KVH0	P0C6D0	Q8CKM7	Q8ZE82	0	Q5HCR6
Q46864	P07000	COG1396	COG2267	Q9KVH0	H9L4S2	Q8CKM7	Q8D138	0	0
P24242	P24242	COG1609	COG1609	Q9KTJ3	Q9KTJ3	Q7CGW3	Q7CGW3	Q5HF38	Q5HF38
P09373	P15082	COG1882	COG1349	Q9KQY1	Q9KUM9	Q8D044	Q7CGX6	Q5HJF4	Q5HE08
P45544	P0ACS9	COG2188	COG1309	Q9KSQ0	Q9KVJ2	Q9ZC75	Q8ZJP6	0	Q5HCN2
P45544	P24242	COG2188	COG1609	Q9KSQ0	Q9KTJ3	Q9ZC75	Q7CGW3	0	Q5HF38
P29745	P29745	COG2195	COG2195	Q9KSB5	Q9KSB5	Q7CK62	Q7CK62	Q5HHS7	Q5HHS7
P0A887	P0AG11	COG2226	COG1974	Q9KVQ6	Q9KVP9	Q8D1I3	Q8ZJ16	Q5HFV2	Q9L4P1
P05100	P30750	COG2818	COG1135	Q9KRH0	Q9KTJ5	Q8CZG8	Q7CHF8	0	Q5HHK4
P0ACJ8	P77398	COG0664	COG0223	P0C6D0	Q9KVU4	Q8ZE82	Q8ZJ80	Q5HCR6	Q5HGL6
P0ABB4	P0ABB0	COG0055	COG0056	Q9KNH5	Q9KNH3	Q7CFM8	Q8Z9S4	Q5HE97	Q5HE95
P0ABB0	P0ABB4	COG0056	COG0055	Q9KNH3	Q9KNH5	Q8Z9S4	Q7CFM8	Q5HE95	Q5HE97
P75966	P0A6E9	COG1187	COG0132	Q9F855	Q9KSZ1	Q8ZEG0	Q8ZGW9	0	0
P24182	P16703	COG0439	COG0031	Q9KV62	Q9KUI4	Q7CL62	Q7CJG9	0	Q5HIG2
P0A9S1	P0A6K3	COG1454	COG0242	Q9KVA0	Q9KVU3	Q7CKP0	Q8ZJ79	0	Q5HGL7
P37773	P0A9A6	COG0773	COG0206	Q9KPG8	Q9KPH1	Q7CKK2	Q7CGB3	Q5HF34	Q5HGP5
P36979	P0A9A6	COG0820	COG0206	Q9KTX3	Q9KPH1	Q7CJM9	Q7CGB3	Q5HGL4	Q5HGP5
P06136	P0A9A6	COG1589	COG0206	Q9KPG9	Q9KPH1	Q7CGB1	Q7CGB3	0	Q5HGP5
P0C058	P0C058	#N/A	COG0071	0	Q9KVX0	0	Q8Z9V6	0	0
P08142	P08142	#N/A	COG0028	0	Q9KVV7	0	Q7CL03	0	0
P00893	P00893	COG0028	COG0028	Q9KVV7	Q9KVV7	Q7CL03	Q7CL03	0	0
P30125	P00893	COG0473	COG0028	Q9KP82	Q9KVV7	Q8D0S0	Q7CL03	Q5HEE3	0
P30748	P30749	COG1977	COG0314	Q9KT78	Q9KT77	Q7CH70	Q8ZGW2	0	Q5HDT6
P0ACJ8	P0AFG0	COG0664	COG0250	P0C6D0	Q9KV35	Q8ZE82	Q8ZAP0	Q5HCR6	Q5HID9
P0AG30	P0AFG0	COG1158	COG0250	Q9KV50	Q9KV35	Q7CKZ2	Q8ZAP0	0	Q5HID9
P0AFI0	P0AFI0	COG0028	COG0028	Q9KVV7	Q9KVV7	Q7CL03	Q7CL03	0	0

P28305	P28305	COG0115	COG0115	Q9KVV9	Q9KVV9	Q8D1L3	Q8D1L3	Q5HIC1	Q5HIC1
P0A799	P28305	COG0126	COG0115	P0C6Q3	Q9KVV9	Q8ZHH3	Q8D1L3	Q5HHP4	Q5HIC1
P0AE08	P28305	COG0450	COG0115	Q9KTZ9	Q9KVV9	Q7CK47	Q8D1L3	Q5HIR5	Q5HIC1
P03018	P09980	#N/A	COG0210	0	Q9KVH9	0	Q8D1K7	0	Q5HEL7
P0A7J0	P0A7J0	COG0108	COG0108	Q9KPU3	Q9KPU3	Q8ZI56	Q8ZI56	Q5HF07	Q5HF07
P24182	P25539	COG0439	COG0117	Q9KV62	Q9KPU1	Q7CL62	Q7CK43	0	0
P0AC65	P37146	COG0695	COG0208	Q9KSW0	Q9KSK1	Q7CLB8	Q8D014	0	0
P00959	P60422	COG0073	COG0090	Q9KT69	Q9KNY7	Q8ZG01	P60436	Q5HII6	Q5HDW1
P0A7W1	P0ADZ0	COG0098	COG0089	Q9KP01	Q9KNY6	Q8ZJ95	P69963	Q5HDX5	Q5HDW0
Q46864	P0A7L8	COG1396	COG0211	Q9KVH0	Q9KUS9	Q8CKM7	Q8ZBA7	0	Q5HFB8
P11454	P0A7M9	COG1020	COG0254	Q9KTV9	Q9KTM4	Q8CKJ6	P58472	Q5HHF2	Q5HE80
P32132	P75864	COG1217	COG0116	Q9KNJ4	Q9KRZ5	Q7CG28	Q7CHK7	0	0
P45543	P0AG07	COG0524	COG0036	Q9KSX7	Q9KNV5	Q7CJW4	Q8CZJ6	0	0
P0A988	P0AG07	COG0592	COG0036	Q9KVX5	Q9KNV5	Q7CFN6	Q8CZJ6	Q5HJZ4	0
P15047	P0A805	COG1028	COG0233	Q56632	Q9KPV5	Q8D0J5	Q8ZH63	0	Q5HGH2
P24178	P0A805	COG1393	COG0233	Q9KQ54	Q9KPV5	Q8D0Y4	Q8ZH63	Q5HH87	Q5HGH2
P0A9A2	P0A805	COG1528	COG0233	Q9KVR1	Q9KPV5	Q8D099	Q8ZH63	Q5HEN0	Q5HGH2
P0AD49	P0A805	COG1544	COG0233	Q9KU23	Q9KPV5	Q8D1Q6	Q8ZH63	Q5HHR8	Q5HGH2
P37188	P0A805	COG3414	COG0233	0	Q9KPV5	Q8CZL5	Q8ZH63	0	Q5HGH2
P0A7N4	P0A7W7	COG0333	COG0096	Q9KQH3	Q9KNZ8	Q8ZFT9	Q8ZJ98	Q5HGV6	Q5HDX2
P08312	P08312	COG0016	COG0016	Q9KSN7	Q9KSN7	Q8ZDX0	Q8ZDX0	Q5HGU6	Q5HGU6
P07395	P08312	COG0072	COG0016	Q9KSN6	Q9KSN7	Q8ZDX1	Q8ZDX0	Q5HGU5	Q5HGU6
P07395	P07395	COG0072	COG0072	Q9KSN6	Q9KSN6	Q8ZDX1	Q8ZDX1	Q5HGU5	Q5HGU5
P52097	P52097	COG0037	COG0037	Q9KS29	Q9KS29	Q7CIP8	Q7CIP8	Q5HIG6	Q5HIG6
P37195	P52097	COG2771	COG0037	Q9KUW3	Q9KS29	Q8CLA3	Q7CIP8	0	Q5HIG6
P0AAB8	P00895	COG0589	COG0147	Q9KVR3	Q9KST2	Q8ZE81	Q8CL24	Q5HF64	0
P0A8F0	P0A8F0	COG0035	COG0035	Q9KPY7	Q9KPY7	Q8ZCX9	Q8ZCX9	Q5HE88	Q5HE88
P23843	P03018	COG4166	#N/A	Q9KT14	0	Q7CIR0	0	0	0
P0A6F5	P76524	COG0459	COG0006	Q9KNR7	Q9KVS2	Q8ZIY3	Q8CLF7	Q5HEH2	Q5HF67

P0ABB0	P0A6E6	COG0056	COG0355	Q9KNH3	Q9KNH6	Q8Z9S4	P58647	Q5HE95	Q5HE98
P0A6N1	P37009	#N/A	COG3842	0	Q9KUB3	0	Q74R28	0	Q5HGY5
P60664	P75728	COG0107	COG0654	Q9KSW8	Q9KTD9	Q8ZFY0	Q7CGT0	Q5HCM4	0
P08142	P0ADF8	COG0028	COG0440	Q9KVV7	Q9KP91	Q7CL03	Q8D0H3	0	0
P0A763	P0A805	COG0105	COG0233	Q9KTX4	Q9KPV5	Q8ZCT2	Q8ZH63	Q5HFV4	Q5HGH2
P0A7J7	P0A805	COG0080	COG0233	Q9KV34	Q9KPV5	Q8ZAP1	Q8ZH63	Q5HID8	Q5HGH2
P0A8V2	P60240	COG0085	COG0553	Q9KV30	Q9KR83	Q8ZAP5	Q8ZII0	Q5HID3	0
P02359	P37765	COG0049	COG1187	Q9KUZ8	Q9F855	Q8ZJB4	Q8ZEG0	Q5HIC9	0
P39362	P14900	COG0036	COG0771	Q9KNV5	Q9KPG5	Q8CZJ6	Q8ZIF1	0	Q5HGP8
P08312	P0AD99	COG0016	COG1114	Q9KSN7	Q9KU61	Q8ZDX0	Q7CK50	Q5HGU6	Q5HG12
P08312	P17117	COG0016	COG0778	Q9KSN7	Q9KU15	Q8ZDX0	Q7CIH7	Q5HGU6	Q5HIR4
P00956	P15042	COG0060	COG0272	Q9KU47	Q9KTD1	Q8ZIM0	Q7CJF3	Q5HGN8	Q5HEL8
P0A890	P0A6B7	#N/A	COG1104	0	Q9KTY2	0	Q7CJN7	0	0
P38506	P07026	#N/A	COG2771	0	Q9KUW3	0	Q9RH45	0	0
No Data	P77467	COG0500	COG1024	Q9KV96	Q9KT58	Q7CH29	Q8ZAN0	0	0
P00579	P00579	COG0568	COG0568	P50511	P50511	Q7CKW6	Q7CKW6	Q5HFJ9	Q5HFJ9
P00894	P00893	COG0440	COG0028	Q9KP91	Q9KVV7	Q8D0H3	Q7CL03	0	0
P00914	P00914	COG0415	COG0415	Q9KS67	Q9KS67	Q7CJR3	Q7CJR3	0	0
P05041	P00903	COG0512	COG0147	Q9KST3	Q9KST2	Q8CZJ3	Q8CL24	0	0
P08312	P07395	COG0016	COG0072	Q9KSN7	Q9KSN6	Q8ZDX0	Q8ZDX1	Q5HGU6	Q5HGU5
P08312	P75679	COG0016	COG2963	Q9KSN7	Q9K344	Q8ZDX0	Q8D1M5	Q5HGU6	0
P08337	P0AC47	COG0494	COG0479	Q9KU53	Q9KQB2	Q8ZHU8	Q7CKM7	0	0
P0A6J8	P0A9A6	COG1181	COG0206	Q9KM17	Q9KPH1	Q8ZIE7	Q7CGB3	Q5HEB7	Q5HGP5
P0A6U3	P25522	COG0445	COG0486	Q9KNG4	Q9KVY5	Q8Z9R8	Q8Z9U2	Q5HCI4	Q5HCI3
P0A7C2	P0A7C2	COG1974	COG1974	Q9KVP9	Q9KVP9	Q8ZJ16	Q8ZJ16	Q9L4P1	Q9L4P1
P0A805	P0A805	COG0233	COG0233	Q9KPV5	Q9KPV5	Q8ZH63	Q8ZH63	Q5HGH2	Q5HGH2
P0A8F4	P0A8F4	COG0572	COG0572	Q9KT67	Q9KT67	Q8ZFZ9	Q8ZFZ9	Q5HFF1	Q5HFF1
P0A8L1	P0A8L1	COG0172	COG0172	Q9KSZ6	Q9KSZ6	Q8ZGC4	Q8ZGC4	Q5HJY7	Q5HJY7
P0A8M3	P0A8M3	COG0441	COG0441	Q9KMN7	Q9KMN7	Q8ZDW5	Q8ZDW5	Q5HF90	Q5HF90

P0A8P8	P0A8P8	COG4974	COG4974	Q9KPE9	Q9KPE9	Q8ZHK1	Q8ZHK1	Q5HFS5	Q5HFS5
P0A988	P0A805	COG0592	COG0233	Q9KVX5	Q9KPV5	Q7CFN6	Q8ZH63	Q5HJZ4	Q5HGH2
P0A9L5	P0A9L5	COG0760	COG0760	Q9KUS0	Q9KUS0	Q7CG87	Q7CG87	Q5HET4	Q5HET4
P0A9P4	P06710	COG0492	COG2812	Q9KSS4	Q9KT51	Q7CHI6	Q7CK10	Q5HHQ4	0
P0A9P6	P0A9P6	COG0513	COG0513	Q9KV52	Q9KV52	Q8ZAD8	Q8ZAD8	Q5HEB9	Q5HEB9
P0A9W0	P76594	COG1349	COG0454	Q9KUM9	Q9KVN5	Q7CGX6	Q8D179	Q5HE08	0
P0AA25	P0A9P4	COG0526	COG0492	P32557	Q9KSS4	Q7CKZ3	Q7CHI6	Q5HGT9	Q5HHQ4
P0AB71	P0AB71	COG0191	COG0191	Q9KUN7	Q9KUN7	Q7CGW5	Q7CGW5	Q5HE75	Q5HE75
P0ABB0	P0A805	COG0056	COG0233	Q9KNH3	Q9KPV5	Q8Z9S4	Q8ZH63	Q5HE95	Q5HGH2
P0ABB4	P0A6E6	COG0055	COG0355	Q9KNH5	Q9KNH6	Q7CFM8	P58647	Q5HE97	Q5HE98
P0ABD5	P0A9Q5	COG0825	COG0777	Q9KPW8	Q9KTA3	Q8ZH52	Q0WDC3	Q5HF74	Q5HF73
P0ACP1	P0ACP1	COG1609	COG1609	Q9KTJ3	Q9KTJ3	Q7CGW3	Q7CGW3	Q5HF38	Q5HF38
P0ACP1	P16384	COG1609	COG0324	Q9KTJ3	Q9KV12	Q7CGW3	Q8ZIW3	Q5HF38	Q5HGC9
P0ADW6	P08390	COG1242	COG0136	Q9KPJ6	Q9KQG2	Q8D1R5	Q7CJA6	0	0
P0AE08	P0A805	COG0450	COG0233	Q9KTZ9	Q9KPV5	Q7CK47	Q8ZH63	Q5HIR5	Q5HGH2
P0AEC3	P75957	COG0642	COG1136	Q9KUA1	Q9KRR8	Q8D1G7	Q8ZFR4	Q5HI08	Q5HDJ6
P0AFG3	P0AFG3	COG0567	COG0567	Q9KQB3	Q9KQB3	Q7CH46	Q7CH46	Q5HG06	Q5HG06
P0AG90	P07395	COG0342	COG0072	Q9KTY7	Q9KSN6	Q8D163	Q8ZDX1	0	Q5HGU5
P10408	P0A8Z3	COG0653	COG0824	Q9KPH4	Q9KR07	Q7CGB6	Q7CH52	Q5HHR7	0
P12996	P00895	COG0502	COG0147	Q9KSZ4	Q9KST2	Q7CH65	Q8CL24	Q5HDC9	0
P14294	P21893	COG0550	COG0608	Q9KRB2	Q9KPF1	Q7CIL8	Q7CGT9	Q5HGI2	0
P14900	P0A9A6	COG0771	COG0206	Q9KPG5	Q9KPH1	Q8ZIF1	Q7CGB3	Q5HGP8	Q5HGP5
P15288	P15288	COG2195	COG2195	Q9KSB5	Q9KSB5	Q7CK62	Q7CK62	Q5HHS7	Q5HHS7
P17952	P0AD68	COG0773	COG0768	Q9KPG8	Q9KTF2	Q7CKK2	Q7CJV5	Q5HF34	0
P21507	P37765	COG0513	COG1187	Q9KV52	Q9F855	Q8ZAD8	Q8ZEG0	Q5HEB9	0
P21865	P21866	COG2205	COG0745	Q9KM57	Q9KU11	Q7CJR5	Q8D168	0	Q5HI09
P24554	P0AFW4	COG1066	COG0782	Q9KPM4	Q9KU89	Q7CG62	Q7CL79	0	Q5HFF2
P28903	P28903	COG1328	COG1328	Q9KM77	Q9KM77	Q7CKG1	Q7CKG1	0	0
P30748	P75679	COG1977	COG2963	Q9KT78	Q9K344	Q7CH70	Q8D1M5	0	0

P32715	P07395	COG1289	COG0072	Q9KLM5	Q9KSN6	Q7CHL6	Q8ZDX1	0	Q5HGU5
P37647	P37647	COG0524	COG0524	Q9KSX7	Q9KSX7	Q7CJW4	Q7CJW4	0	0
P37765	P75864	COG1187	COG0116	Q9F855	Q9KRZ5	Q8ZEG0	Q7CHK7	0	0
P46888	P28630	COG0589	COG1466	Q9KVR3	Q9KTE9	Q8ZE81	Q7CJV3	Q5HF64	0
P52106	P52106	COG2771	COG2771	Q9KUW3	Q9KUW3	Q8CLA3	Q9RH45	0	0
P60240	P75864	COG0553	COG0116	Q9KR83	Q9KRZ5	Q8ZII0	Q7CHK7	0	0
P77808	P0A7W7	COG1058	COG0096	Q9KSK4	Q9KNZ8	Q7CH93	Q8ZJ98	0	Q5HDX2
P24251	P13445	COG0060	#N/A	Q9KU47	0	Q8ZIM0	0	Q5HGN8	0
P25539	P25539	COG0117	COG0117	Q9KPU1	Q9KPU1	Q7CK43	Q7CK43	0	0
P51020	P00895	COG0119	COG0147	Q9KP83	Q9KST2	Q8ZIG8	Q8CL24	Q5HEE4	0
P06987	P06987	COG0131	COG0131	Q9KSX1	Q9KSX1	Q8ZFX7	Q8ZFX7	Q5HCM1	Q5HCM1
P08390	P21599	COG0136	COG0469	Q9KQG2	Q9KUN0	Q7CJA6	Q7CIS8	0	Q5HF76
P22939	P22939	COG0142	COG0142	Q9KUT1	Q9KUT1	Q7CK38	Q7CK38	0	0
P0A858	P09373	COG0149	COG1882	Q9KNR1	Q9KQY1	Q8ZJK9	Q8D044	Q5HHP3	Q5HJF4
P0AB77	P0AB77	COG0156	COG0156	Q9KSZ3	Q9KSZ3	Q8CWI4	Q8CWI4	Q5HIC5	Q5HIC5
P0A6T1	P0A6T1	COG0166	COG0166	Q9KUY4	Q9KUY4	Q8ZAS2	Q8ZAS2	Q5HHC2	Q5HHC2
P15770	P0AGJ2	COG0169	COG0566	Q9KVT3	Q9KTT5	Q8ZJ74	Q8ZIV4	Q5HFG5	Q5HIE3
P15770	P05052	COG0169	COG2207	Q9KVT3	Q9KVF4	Q8ZJ74	Q7CG68	Q5HFG5	Q5HJR8
P18843	P18843	COG0171	COG0171	Q9KMW1	Q9KMW1	Q7CJP7	Q7CJP7	Q5HEK9	Q5HEK9
P18843	P03004	COG0171	COG0593	Q9KMW1	Q9KVX6	Q7CJP7	Q8Z9U7	Q5HEK9	Q5HJZ5
P0A698	P0A8F8	COG0178	COG0556	Q9KUW5	0	Q8ZJ07	Q8ZGW7	Q5HHQ9	Q5HHR0
P20083	P20083	COG0187	COG0187	Q9KVX3	Q9KVX3	Q7CGH6	Q7CGH6	Q5HG65	Q5HG65
P0AB74	P0AEM6	COG0191	COG1191	Q9KUN7	Q9KQD4	Q7CGW5	Q7CI01	Q5HE75	0
P39452	P0A8F8	COG0209	COG0556	Q9KSK0	0	Q8D117	Q8ZGW7	0	Q5HHR0
P09980	P09980	COG0210	COG0210	Q9KVH9	Q9KVH9	Q8D1K7	Q8D1K7	Q5HEL7	Q5HEL7
P03018	P25665	COG0210	COG0620	Q9KVH9	Q9KRD8	Q8D1K7	Q8ZAL3	Q5HEL7	Q5HIT8
P0ABA6	P0A6E6	COG0224	COG0355	Q9KNH4	Q9KNH6	Q8Z9S5	P58647	Q5HE96	Q5HE98
P0A805	P0AD49	COG0233	COG1544	Q9KPV5	Q9KU23	Q8ZH63	Q8D1Q6	Q5HGH2	Q5HHR8
P0A805	P00805	COG0233	COG0252	Q9KPV5	Q9KQK3	Q8ZH63	Q7CIH5	Q5HGH2	0

P0A7T7	P02358	COG0238	COG0360	Q9KUZ0	Q9KUZ2	Q8ZB83	Q8ZB81	Q5HIS7	Q5HIS9
P0A6S7	P0A6S7	COG0240	COG0240	Q9KNT0	Q9KNT0	Q8ZJM6	Q8ZJM6	Q5HFU9	Q5HFU9
P0A6K3	P0AGJ2	COG0242	COG0566	Q9KVU3	Q9KTT5	Q8ZJ79	Q8ZIV4	Q5HGL7	Q5HIE3
P0A6K3	P75728	COG0242	COG0654	Q9KVU3	Q9KTD9	Q8ZJ79	Q7CGT0	Q5HGL7	0
P0A7M9	P12996	COG0254	COG0502	Q9KTM4	Q9KSZ4	P58472	Q7CH65	Q5HE80	Q5HDC9
P0A7M9	P0AEI4	COG0254	COG0621	Q9KTM4	Q9KTE0	P58472	Q0WDR2	Q5HE80	Q5HGD9
P0ABQ4	Q46948	COG0262	COG0693	Q9KUS5	Q9KPQ8	Q7CG83	Q8D159	Q5HFZ7	Q5HIC4
P15042	P15042	COG0272	COG0272	Q9KTD1	Q9KTD1	Q7CJF3	Q7CJF3	Q5HEL8	Q5HEL8
P15042	P07813	COG0272	COG0495	Q9KTD1	Q9KTE6	Q7CJF3	Q8ZDF8	Q5HEL8	Q5HF16
P60390	P16384	COG0275	COG0324	Q9KPF9	Q9KV12	Q8ZIF7	Q8ZIW3	Q5HGQ2	Q5HGC9
P0A9M8	P77625	COG0280	COG0637	Q9KT08	Q9KN63	Q8D0X3	Q7CJ95	Q5HI88	0
P0ABF6	P0ABF6	COG0295	COG0295	Q9KSM5	Q9KSM5	Q8ZG08	Q8ZG08	0	0
P08179	P08179	COG0299	COG0299	Q9KPY5	Q9KPY5	Q7CJJ9	Q7CJJ9	Q5HH12	Q5HH12
P12281	P12281	COG0303	COG0303	Q9KRV7	Q9KRV7	Q7CHP6	Q7CHP6	Q5HDT4	Q5HDT4
P30749	P30749	COG0314	COG0314	Q9KT77	Q9KT77	Q8ZGW2	Q8ZGW2	Q5HDT6	Q5HDT6
P30749	P46888	COG0314	COG0589	Q9KT77	Q9KVR3	Q8ZGW2	Q8ZE81	Q5HDT6	Q5HF64
P30749	P75824	COG0314	COG1018	Q9KT77	Q9KL25	Q8ZGW2	Q8D039	Q5HDT6	0
P0A8G0	P0A8G0	COG0322	COG0322	Q9KSP2	Q9KSP2	Q8ZF52	Q8ZF52	Q5HGT7	Q5HGT7
P23367	P23367	COG0323	COG0323	Q9KV13	Q9KV13	Q8ZIW4	Q8ZIW4	Q5HGD5	Q5HGD5
P0A6L2	P0A6L2	COG0329	COG0329	Q9KR67	Q9KR67	Q8ZCD0	Q8ZCD0	Q5HG25	Q5HG25
P0A873	P0A873	COG0336	COG0336	Q9KUF8	Q9KUF8	Q8ZBU9	Q8ZBU9	Q5HGJ2	Q5HGJ2
P0AC81	P0A805	COG0346	COG0233	Q9KT93	Q9KPV5	Q8D0L9	Q8ZH63	Q5HDM4	Q5HGH2
P06134	P00579	COG0350	COG0568	Q9KKT3	P50511	Q7CLC0	Q7CKW6	0	Q5HFJ9
P06134	P69931	COG0350	COG0593	Q9KKT3	Q9KVX6	Q7CLC0	Q8Z9U7	0	Q5HJZ5
P0A6E6	P0A6E6	COG0355	COG0355	Q9KNH6	Q9KNH6	P58647	P58647	Q5HE98	Q5HE98
P00350	P00350	COG0362	COG0362	Q9KL50	Q9KL50	Q8D078	Q8D078	Q5HFR2	Q5HFR2
P28904	P24218	COG0366	COG0582	Q9KTJ1	Q9KUK2	Q7CL80	Q8D1C0	0	0
P38038	P38038	COG0369	COG0369	Q9KUX4	Q9KUX4	Q8ZBN6	Q8ZBN6	0	0
P33650	P33650	COG0370	COG0370	Q9KQC3	Q9KQC3	Q8CZJ9	Q8CZJ9	Q5HD01	Q5HD01

P0ABH7	P0ABH7	COG0372	COG0372	Q9KSC1	Q9KSC1	Q7CH41	Q7CH41	0	0
P77806	P10443	COG0436	COG0587	Q9KQM1	P52022	Q7CJE7	Q7CH18	0	Q5HF71
P77806	P77806	COG0436	COG0436	Q9KQM1	Q9KQM1	Q7CJE7	Q7CJE7	0	0
P77434	P77434	COG0436	COG0436	Q9KQM1	Q9KQM1	Q7CJE7	Q7CJE7	0	0
P24182	P0A9Q5	COG0439	COG0777	Q9KV62	Q9KTA3	Q7CL62	Q0WDC3	0	Q5HF73
P24182	P24182	COG0439	COG0439	Q9KV62	Q9KV62	Q7CL62	Q7CL62	0	0
P24182	P0ABD8	COG0439	COG0511	Q9KV62	Q9KV61	Q7CL62	Q8D1P3	0	0
P00894	P00894	COG0440	COG0440	Q9KP91	Q9KP91	Q8D0H3	Q8D0H3	0	0
P0ABU0	P0ABU0	COG0447	COG0447	Q9KQM5	Q9KQM5	Q8D0S4	Q8D0S4	Q5HH38	Q5HH38
P0AE08	P0AE08	COG0450	COG0450	Q9KTZ9	Q9KTZ9	Q7CK47	Q7CK47	Q5HIR5	Q5HIR5
P31600	P31600	COG0457	COG0457	Q9KTK1	Q9KTK1	Q7CJI7	Q7CJI7	0	0
P00968	P00968	COG0458	COG0458	Q9KPH9	Q9KPH9	Q8ZIL4	Q8ZIL4	Q5HGM9	Q5HGM9
P00968	P0A6F1	COG0458	COG0505	Q9KPH9	Q9KPH8	Q8ZIL4	Q8ZIL5	Q5HGM9	Q5HGN0
P0A6F5	P09372	COG0459	COG0576	Q9KNR7	O30862	Q8ZIY3	Q7CH40	Q5HEH2	Q5HFH9
P0A6F5	P07913	COG0459	COG1063	Q9KNR7	Q9KL62	Q8ZIY3	Q8ZJN2	Q5HEH2	0
P21599	P21599	COG0469	COG0469	Q9KUN0	Q9KUN0	Q7CIS8	Q7CIS8	Q5HF76	Q5HF76
P21599	P19642	COG0469	COG1263	Q9KUN0	Q9KVD9	Q7CIS8	Q7CL81	Q5HF76	Q5HJI0
P60785	P17446	COG0481	COG1309	Q9KPB0	Q9KVJ2	Q8ZD74	Q8ZJP6	Q5HFH6	Q5HCN2
P60785	P64423	COG0481	COG0598	Q9KPB0	Q9KPN2	Q8ZD74	Q8ZAG5	Q5HFH6	0
P0A9P4	P0AC65	COG0492	COG0695	Q9KSS4	Q9KSW0	Q7CHI6	Q7CLB8	Q5HHQ4	0
P08337	P08337	COG0494	COG0494	Q9KU53	Q9KU53	Q8ZHU8	Q8ZHU8	0	0
P0AFG6	P0AFG3	COG0508	COG0567	Q9KQB4	Q9KQB3	Q7CH47	Q7CH46	Q5HG07	Q5HG06
P0AFG6	P0AFG6	COG0508	COG0508	Q9KQB4	Q9KQB4	Q7CH47	Q7CH47	Q5HG07	Q5HG07
P06959	P0A9P0	COG0508	COG1249	Q9KQB4	P50529	Q7CH47	Q7CL12	Q5HG07	Q5HGY8
P0AFG6	P0A9P0	COG0508	COG1249	Q9KQB4	P50529	Q7CH47	Q7CL12	Q5HG07	Q5HGY8
P0A8J8	P0A8J8	COG0513	COG0513	Q9KV52	Q9KV52	Q8ZAD8	Q8ZAD8	Q5HEB9	Q5HEB9
P60560	P15288	COG0516	COG2195	Q9KTW3	Q9KSB5	Q8ZBI2	Q7CK62	Q5HIQ7	Q5HHS7
Q46925	Q46925	COG0520	COG0520	Q9KSS2	Q9KSS2	Q8D0M6	Q8D0M6	Q5HHH0	Q5HHH0
P33030	P30014	COG0523	COG0847	Q9KM61	Q9KV58	Q7CHD5	Q8ZE08	0	Q5HFW8

P45543	P45543	COG0524	COG0524	Q9KSX7	Q9KSX7	Q7CJW4	Q7CJW4	0	0
P07118	P07118	COG0525	COG0525	Q9KP73	Q9KP73	Q8ZBH1	Q8ZBH1	Q5HFA8	Q5HFA8
P0ACE7	P0ACE7	COG0537	COG0537	Q9KQV1	Q9KQV1	Q7CJ15	Q7CJ15	0	0
P0AGD7	P10121	COG0541	COG0552	Q9KUG1	Q9KVJ6	Q7CK91	Q8D1J2	0	0
P0ABH9	P75679	COG0542	COG2963	Q9KU18	Q9K344	Q7CL29	Q8D1M5	Q5HD02	0
P06612	P15042	COG0550	COG0272	Q9KRB2	Q9KTD1	Q7CIL8	Q7CJF3	Q5HGI2	Q5HEL8
P06612	P07813	COG0550	COG0495	Q9KRB2	Q9KTE6	Q7CIL8	Q8ZDF8	Q5HGI2	Q5HF16
No Data	P46888	COG0551	COG0589	Q9KVT9	Q9KVR3	Q8CZI7	Q8ZE81	0	Q5HF64
P10121	P0AGD7	COG0552	COG0541	Q9KVJ6	Q9KUG1	Q8D1J2	Q7CK91	0	0
P69441	P77806	COG0563	COG0436	Q9KTB7	Q9KQM1	O69172	Q7CJE7	Q5HDX9	0
P0AA37	P0AG76	COG0564	COG0420	Q9KU20	Q9KM68	Q8ZBV7	Q7CK53	0	Q5HG73
P0AA41	P0AEI4	COG0564	COG0621	Q9KU20	Q9KTE0	Q8ZBV7	Q0WDR2	0	Q5HGD9
P0AGJ2	P0AGJ2	COG0566	COG0566	Q9KTT5	Q9KTT5	Q8ZIV4	Q8ZIV4	Q5HIE3	Q5HIE3
P00579	P60240	COG0568	COG0553	P50511	Q9KR83	Q7CKW6	Q8ZII0	Q5HFJ9	0
P0A8F4	P39177	COG0572	COG0589	Q9KT67	Q9KVR3	Q8ZFZ9	Q8ZE81	Q5HFF1	Q5HF64
P10908	P0A7Y0	COG0584	COG0571	Q9KRT2	Q9KPB2	Q7CK63	Q8ZD72	0	Q5HGJ9
P10908	P32662	COG0584	COG0546	Q9KRT2	Q9KNV6	Q7CK63	Q8ZJF3	0	0
P10443	P03007	COG0587	COG0847	P52022	Q9KV58	Q7CH18	Q8ZE08	Q5HF71	Q5HFW8
P0A988	P10443	COG0592	COG0587	Q9KVX5	P52022	Q7CFN6	Q7CH18	Q5HJZ4	Q5HF71
P0A988	P0A988	COG0592	COG0592	Q9KVX5	Q9KVX5	Q7CFN6	Q7CFN6	Q5HJZ4	Q5HJZ4
P0A988	P69931	COG0592	COG0593	Q9KVX5	Q9KVX6	Q7CFN6	Q8Z9U7	Q5HJZ4	Q5HJZ5
P0A988	P28630	COG0592	COG1466	Q9KVX5	Q9KTE9	Q7CFN6	Q7CJV3	Q5HJZ4	0
P69931	P69931	COG0593	COG0593	Q9KVX6	Q9KVX6	Q8Z9U7	Q8Z9U7	Q5HJZ5	Q5HJZ5
P64423	P64423	COG0598	COG0598	Q9KPN2	Q9KPN2	Q8ZAG5	Q8ZAG5	0	0
P0A9N8	P0A9N8	COG0602	COG0602	Q9KS94	Q9KS94	Q7CKD0	Q7CKD0	0	0
P0A9N8	P28903	COG0602	COG1328	Q9KS94	Q9KM77	Q7CKD0	Q7CKG1	0	0
P0A809	P0A809	COG0632	COG0632	Q9KR01	Q9KR01	Q8ZEU6	Q8ZEU6	Q5HFC1	Q5HFC1
P08401	P08368	COG0642	COG0745	Q9KUA1	Q9KU11	Q8D1G7	Q8D168	Q5HI08	Q5HI09
P10408	P0AG99	COG0653	COG1314	Q9KPH4	Q9KU83	Q7CGB6	Q7CKI9	Q5HHR7	Q5HHN9

P25535	P00805	COG0654	COG0252	Q9KTD9	Q9KQK3	Q7CGT0	Q7CIH5	0	0
P0ACJ8	P0ACN7	COG0664	COG1609	P0C6D0	Q9KTJ3	Q8ZE82	Q7CGW3	Q5HCR6	Q5HF38
P0ACJ8	P00579	COG0664	COG0568	P0C6D0	P50511	Q8ZE82	Q7CKW6	Q5HCR6	Q5HFJ9
P0AG96	P00579	COG0690	COG0568	Q9KV36	P50511	Q8D1H4	Q7CKW6	0	Q5HFJ9
P0A832	P0A832	COG0691	COG0691	P52116	P52116	Q8ZH14	Q8ZH14	Q5HHN6	Q5HHN6
P12295	P75957	COG0692	COG1136	Q9KPK8	Q9KRR8	Q8ZD85	Q8ZFR4	Q5HI95	Q5HDJ6
P0ABA4	P06710	COG0712	COG2812	Q9KNH2	Q9KT51	Q7CFM5	Q7CK10	Q5HE94	0
P0A6G7	P0A6G7	COG0740	COG0740	Q9KQS6	Q9KQS6	Q8ZC65	Q8ZC65	Q5HHQ0	Q5HHQ0
P69228	P69228	COG0745	COG0745	Q9KU11	Q9KU11	Q8D168	Q8D168	Q5HI09	Q5HI09
P08368	P08368	COG0745	COG0745	Q9KU11	Q9KU11	Q8D168	Q8D168	Q5HI09	Q5HI09
P21179	P21179	COG0753	COG0753	Q9KRQ1	Q9KRQ1	Q7CH91	Q7CH91	Q5HG86	Q5HG86
P21179	P0AEM6	COG0753	COG1191	Q9KRQ1	Q9KQD4	Q7CH91	Q7CI01	Q5HG86	0
P45768	P0A9Q5	COG0765	COG0777	Q9KVX9	Q9KTA3	Q7CJV0	Q0WDC3	0	Q5HF73
P45768	P0ABD5	COG0765	COG0825	Q9KVX9	Q9KPW8	Q7CJV0	Q8ZH52	0	Q5HF74
P0AD68	P46022	COG0768	COG0744	Q9KTF2	Q9KUC0	Q7CJV5	Q8D1A6	0	Q5HEQ0
P11880	P11880	COG0770	COG0770	Q9KPG3	Q9KPG3	Q7CGA9	Q7CGA9	0	0
P14900	P14900	COG0771	COG0771	Q9KPG5	Q9KPG5	Q8ZIF1	Q8ZIF1	Q5HGP8	Q5HGP8
P17952	P17952	COG0773	COG0773	Q9KPG8	Q9KPG8	Q7CKK2	Q7CKK2	Q5HF34	Q5HF34
P37773	P37773	COG0773	COG0773	Q9KPG8	Q9KPG8	Q7CKK2	Q7CKK2	Q5HF34	Q5HF34
P0AF12	P07026	COG0775	COG2771	Q9KPI8	Q9KUW3	Q7CKD4	Q9RH45	Q5HFG2	0
P0A6X7	P0A8F8	COG0776	COG0556	Q9KV83	0	Q7CKT4	Q8ZGW7	Q5HFV0	Q5HHR0
P0A9Q5	P46888	COG0777	COG0589	Q9KTA3	Q9KVR3	Q0WDC3	Q8ZE81	Q5HF73	Q5HF64
P0A9Q5	P0ABD5	COG0777	COG0825	Q9KTA3	Q9KPW8	Q0WDC3	Q8ZH52	Q5HF73	Q5HF74
P38489	P38489	COG0778	COG0778	Q9KU15	Q9KU15	Q7CIH7	Q7CIH7	Q5HIR4	Q5HIR4
P21507	P21507	COG0820	COG0513	Q9KTX3	Q9KV52	Q7CJM9	Q8ZAD8	Q5HGL4	Q5HEB9
P36979	P36979	COG0820	COG0820	Q9KTX3	Q9KTX3	Q7CJM9	Q7CJM9	Q5HGL4	Q5HGL4
P0ACD4	P0ACD4	COG0822	COG0822	Q9KTY1	Q9KTY1	Q7CJN6	Q7CJN6	0	0
P0ACD4	P06710	COG0822	COG2812	Q9KTY1	Q9KT51	Q7CJN6	Q7CK10	0	0
P0ABD5	P0ABD5	COG0825	COG0825	Q9KPW8	Q9KPW8	Q8ZH52	Q8ZH52	Q5HF74	Q5HF74

P0ABD5	P69874	COG0825	COG3842	Q9KPW8	Q9KUB3	Q8ZH52	Q74R28	Q5HF74	Q5HGY5
P37902	P0AAG3	COG0834	COG1126	Q9KVX8	Q9KVY0	Q7CK73	Q7CK70	0	0
P37902	P37902	COG0834	COG0834	Q9KVX8	Q9KVX8	Q7CK73	Q7CK73	0	0
P75960	P75960	COG0846	COG0846	Q9KRX4	Q9KRX4	Q8ZFR1	Q8ZFR1	Q5HE07	Q5HE07
P75960	P75869	COG0846	COG3070	Q9KRX4	Q9KSV3	Q8ZFR1	Q8D057	Q5HE07	0
P03007	P03007	COG0847	COG0847	Q9KV58	Q9KV58	Q8ZE08	Q8ZE08	Q5HFW8	Q5HFW8
P25516	P31551	COG1048	COG1024	Q9KSC0	Q9KT58	Q7CIL5	Q8ZAN0	Q5HG69	0
P08839	P0AEL8	COG1080	COG2197	Q9KTD7	Q9KSP3	Q7CJF7	Q7CHT5	Q5HH01	Q5HDG5
P23878	P23878	COG1120	COG1120	Q9KVE5	Q9KVE5	Q7CKD5	Q7CKD5	0	0
P10346	P10346	COG1126	COG1126	Q9KVY0	Q9KVY0	Q7CK70	Q7CK70	0	0
P75957	P31802	COG1136	COG2197	Q9KRR8	Q9KSP3	Q8ZFR4	Q7CHT5	Q5HDJ6	Q5HDG5
P75957	P07821	COG1136	COG1120	Q9KRR8	Q9KVE5	Q8ZFR4	Q7CKD5	Q5HDJ6	0
P0AG30	P0AG30	COG1158	COG1158	Q9KV50	Q9KV50	Q7CKZ2	Q7CKZ2	0	0
P17109	P17109	COG1165	COG1165	Q9KQM3	Q9KQM3	Q7CJ76	Q7CJ76	Q5HH40	Q5HH40
P17109	P37647	COG1165	COG0524	Q9KQM3	Q9KSX7	Q7CJ76	Q7CJW4	Q5HH40	0
P05055	P21507	COG1185	COG0513	Q9KU76	Q9KV52	Q0WBF9	Q8ZAD8	Q5HGF7	Q5HEB9
P07012	P07026	COG1186	COG2771	0	Q9KUW3	Q8ZHK4	Q9RH45	Q5HHR5	0
P0ACC7	P0ACC7	COG1207	COG1207	Q9KNH7	Q9KNH7	Q8Z9S7	Q8Z9S7	Q5HIH6	Q5HIH6
P76256	P05852	COG1214	COG0533	Q9KQK9	0	Q8D0F5	Q74RQ9	0	Q5HEF2
P76256	P0AF67	COG1214	COG0802	Q9KQK9	Q9KV15	Q8D0F5	Q7CKM2	0	0
P76256	P0A944	COG1214	COG0456	Q9KQK9	Q9KU66	Q8D0F5	Q8CZN5	0	0
P76256	P76256	COG1214	COG1214	Q9KQK9	Q9KQK9	Q8D0F5	Q8D0F5	0	0
P32132	P32132	COG1217	COG1217	Q9KNJ4	Q9KNJ4	Q7CG28	Q7CG28	0	0
P0A6H1	P0A6H1	COG1219	COG1219	Q9KQS7	Q9KQS7	Q8ZC66	Q8ZC66	Q5HF98	Q5HF98
P08722	P0A8A8	COG1263	COG0779	Q9KVD9	Q9KU82	Q7CL81	Q8ZBC0	Q5HJI0	Q5HGG6
P0A9X9	P25888	COG1278	COG0513	Q9KSW4	Q9KV52	Q8D1N6	Q8ZAD8	Q5HG18	Q5HEB9
P0A9Y6	P0A9Y6	COG1278	COG1278	Q9KSW4	Q9KSW4	Q8D1N6	Q8D1N6	Q5HG18	Q5HG18
P0ACT6	P0ACT6	COG1309	COG1309	Q9KVJ2	Q9KVJ2	Q8ZJP6	Q8ZJP6	Q5HCN2	Q5HCN2
P0A8D0	P0A8D0	COG1327	COG1327	Q9KPU0	Q9KPU0	Q8ZC39	Q8ZC39	Q5HF87	Q5HF87

P28903	P30138	COG1328	COG0476	Q9KM77	Q9KVS6	Q7CKG1	Q7CKT8	0	0
P24178	P24178	COG1393	COG1393	Q9KQ54	Q9KQ54	Q8D0Y4	Q8D0Y4	Q5HH87	Q5HH87
P24178	P0A9N4	COG1393	COG1180	Q9KQ54	Q9KQX8	Q8D0Y4	Q8D043	Q5HH87	Q5HJF3
P24178	P75824	COG1393	COG1018	Q9KQ54	Q9KL25	Q8D0Y4	Q8D039	Q5HH87	0
P0A9S1	P0A9S1	COG1454	COG1454	Q9KVA0	Q9KVA0	Q7CKP0	Q7CKP0	0	0
P04846	P0AGJ2	COG1464	COG0566	Q9KTJ7	Q9KTT5	Q7CHF7	Q8ZIV4	0	Q5HIE3
P0A9A2	P33593	COG1528	COG0444	Q9KVR1	Q9KUA6	Q8D099	Q8D1C2	Q5HEN0	Q5HG40
P0A9A2	P75679	COG1528	COG2963	Q9KVR1	Q9K344	Q8D099	Q8D1M5	Q5HEN0	0
P0AD49	P0AD49	COG1544	COG1544	Q9KU23	Q9KU23	Q8D1Q6	Q8D1Q6	Q5HHR8	Q5HHR8
P27303	P27303	COG1566	COG1566	Q9KS50	Q9KS50	Q8D177	Q8D177	0	0
P03024	P03024	COG1609	COG1609	Q9KTJ3	Q9KTJ3	Q7CGW3	Q7CGW3	Q5HF38	Q5HF38
P03024	P25748	COG1609	COG1609	Q9KTJ3	Q9KTJ3	Q7CGW3	Q7CGW3	Q5HF38	Q5HF38
P0ADR8	P28304	COG1611	COG0604	Q9KTK3	Q9KVW2	Q7CH10	Q8D0P7	0	Q5HE19
P0AFI5	P0A6P5	COG1686	COG1160	Q9KTF5	Q9KTW7	Q7CJV7	Q8ZCT9	0	Q5HFU8
P0AFI5	P69931	COG1686	COG0593	Q9KTF5	Q9KVX6	Q7CJV7	Q8Z9U7	0	Q5HJZ5
P0AEY3	P06616	COG1694	COG1159	Q9KPC2	Q9KPB3	Q7CKD1	Q8ZD71	0	Q5HFJ3
P69829	P0ADG4	COG1762	COG0483	Q9KR27	Q9KTY5	Q7CJ97	Q7CJN9	0	0
P32125	P32125	COG1763	COG1763	0	0	Q8CZM2	Q8CZM2	0	0
P0A8A4	P0A8A4	COG1806	COG1806	Q9KKW4	Q9KKW4	Q8ZDY4	Q8ZDY4	Q5HFJ7	Q5HFJ7
P0AG51	P12996	COG1841	COG0502	Q9KP02	Q9KSZ4	Q7CFT2	Q7CH65	Q5HDX6	Q5HDC9
P69811	P69811	COG1925	COG1925	Q9KTD6	Q9KTD6	Q7CHE8	Q7CHE8	Q5HH02	Q5HH02
P32718	P32718	COG1940	COG1940	Q9KV86	Q9KV86	Q8CL97	Q8CL97	0	0
P0AGK8	P0AGK8	COG1959	COG1959	Q9KTY3	Q9KTY3	Q0WJT1	Q0WJT1	0	0
P46837	P46837	COG2183	COG2183	Q9KNL6	Q9KNL6	Q7CFX1	Q7CFX1	0	0
P69407	P69407	COG2197	COG2197	Q9KSP3	Q9KSP3	Q7CHT5	Q7CHT5	Q5HDG5	Q5HDG5
P31802	P75957	COG2197	COG1136	Q9KSP3	Q9KRR8	Q7CHT5	Q8ZFR4	Q5HDG5	Q5HDJ6
P21865	P69829	COG2205	COG1762	Q9KM57	Q9KR27	Q7CJR5	Q7CJ97	0	0
P21865	P21865	COG2205	COG2205	Q9KM57	Q9KM57	Q7CJR5	Q7CJR5	0	0
P06710	P10443	COG2812	COG0587	Q9KT51	P52022	Q7CK10	Q7CH18	0	Q5HF71

P06710	P03007	COG2812	COG0847	Q9KT51	Q9KV58	Q7CK10	Q8ZE08	0	Q5HFW8
P06710	P06710	COG2812	COG2812	Q9KT51	Q9KT51	Q7CK10	Q7CK10	0	0
P75869	P75960	COG3070	COG0846	Q9KSV3	Q9KRX4	Q8D057	Q8ZFR1	0	Q5HE07
P75742	P76594	COG3104	COG0454	Q9KTB5	Q9KVN5	Q0WEG0	Q8D179	0	0
P0AFT5	P0AFT5	COG3279	COG3279	Q9KU36	Q9KU36	Q8ZBV2	Q8ZBV2	Q5HJB5	Q5HJB5
P09833	P42630	COG4148	COG1760	Q9KLL9	Q9KSF6	Q8ZGX6	Q7CHW1	0	0
P77737	P77467	COG4608	COG1024	Q9KUA7	Q9KT58	Q8CZL6	Q8ZAN0	Q5HG41	0
P28304	P28304	0	COG0604	0	Q9KVW2	0	Q8D0P7	0	Q5HE19
P00579	P0A7C2	0	COG1974	0	Q9KVP9	0	Q8ZJ16	0	Q9L4P1

Appendix Table 3.6 – List of interactions found positive in *Eco, Sau, Vch* and *Ype* using yeast two hybrid screens using pGADT7g and pGBGT7g vector combination

x indicates presence of interaction

Bait Protein	Prey protein	Eco	Vch	Ype	Sau
acrR	ybdM	x	x		
acuI	rs8	x			
ascG	ybdM	x			
caiC	caiC	x			
crp	yiiR	x			
crp	yfgD	x	x		x
cusR	yedW	x		x	
dppD	ddpB	x		x	
feoA	yheS	x		x	
fnr	yfgD	x			
gngF	syl	x			
gspE	rl2	x			
ilvY	yadE	x		x	
ingK	gngF	x	x		
intQ	ex7L	x			
lipA	yheS	x			
mdtO	yeaJ	x			
metN	ydjH	x			
mlrA	yebC	x			
mog	yebC	x	x		
mqsA	ygjR				
mqsA	yfgD	x	x		
murI	yggT	x		x	
pepT	intB	x	x		

phoB	pqqL	x			
pldB	yebC				
potG	yihL	x			
raDD	fsaB	x			
sdhB	yihL	x			
srlR	ynbC	x			
ubiH	ybiH	x			
umuD	ygjR	x			
uspA	yjiA	x			
wcaE	dapE	x			
yadG	hisX	x	x		
ybdM	rimM	x			
ybdM	uvrB		x		
ybdM	hslR		x		
ybdM	rnk	x			
ybdM	glrX1	x			
ybdM	hcr	x	x		
ybdM	ybdM	x			
ybdM	eutD	x	x	x	
ybgK	cysM	x	x	x	x
ybiH	ybiH	x	x		x
ycaC	yecE		x		
ycjW	ycjW	x			
ydcP	recN	x	x		
ydgJ	ydgJ	x			
ydjL	ydjL	x			
yecE	uspA	x	x		
yecE	yecE				
yeeD	aroB	x	x		
yeeD	clpB	x	x	x	x

yeeD	cysM	x			
yeeD	dapE	x			
yeeD	ddlB	x	x	x	x
yeeD	gabT	x			
yeeD	murA	x	x	x	x
yeeD	yahK	x			
yeeD	treR	x	x	x	x
yeeD	truA	x	x	x	x
yfdR	dnaA	x			
yfiF	murG	x	x		
ygaV	cytR	x			
yhiN	mcbR	x			
yigZ	ygjR	x	x		
yjhC	yjhC	x			
yjhG	yjhG	x			
yjjW	ilvN	x	x		
yjjW	suhB	x			
yjjW	murI	x	x		
yjjW	hcaD	x			
ykgN	dcdA	x	x		
ykgN	recD	x	x	x	x
ykgN	tsaD	x			
ykgN	era	x	x	x	
ykgN	uvrB	x			
ykgN	hprT	x			
ykgN	elyC	x	x	x	x
ykgN	dnaG	x			
ykgN	hslR		x		
ykgN	soxR				
ykgN	rpoE	x			

ykgN	gutQ	x	x		
ykgN	recJ			x	
ykgN	holA	x		x	x
ykgN	psuK				
ykgN	ubiF	x			
ykgN	hcr	x	x	x	
ykgN	ydhK	x			
ynbD	glmS				
yqeF	yqeF	x	x		

Appendix Table 4.1 – Hubs identified from all *Eco* Interactions (Raja *et al* dataset). Hubs defined as proteins having >= 8 interactions

unknown function hub proteins

unknown function hub proteins present in all spp.

UniProt ID	Protein	Protein names
P24178	YFFB_ECOLI	Protein YffB
P75767	GNGF_ECOLI	Putative gluconeogenesis factor
Q47147	YAFJ_ECOLI	Putative glutamine amidotransferase YafJ (EC 2.4.2.-)
P39409	YJJW_ECOLI	Putative glycyl-radical enzyme activating enzyme YjjW (GRE activating enzyme YjjW) (EC 1.97.1.-)
P77396	YPDC_ECOLI	Uncharacterized HTH-type transcriptional regulator YpdC
P75680	INSO1_ECOLI	Putative transposase InsO for insertion sequence element IS911A
Q79E92	YKGN_ECOLI	Putative transposase YkgN
P77460	YBCY_ECOLI	Putative uncharacterized protein YbcY
P39391	YJIT_ECOLI	Putative uncharacterized protein YjiT
P0ABW5	SFMA_ECOLI	Uncharacterized fimbrial-like protein SfmA (Type-1A pilin)
P37909	YBGD_ECOLI	Uncharacterized fimbrial-like protein YbgD
P76500	YFCQ_ECOLI	Uncharacterized fimbrial-like protein YfcQ
P39834	YGIL_ECOLI	Uncharacterized fimbrial-like protein YgiL
P45579	YBDH_ECOLI	Uncharacterized oxidoreductase YbdH (EC 1.1.-.-)
P77174	YBDM_ECOLI	Uncharacterized protein YbdM
P64455	YDCY_ECOLI	Uncharacterized protein YdcY
P64503	YEBV_ECOLI	Uncharacterized protein YebV
Q2M7R5	YIBT_ECOLI	Uncharacterized protein YibT
P39351	YJGZ_ECOLI	Uncharacterized protein YjgZ
P33014	YEED_ECOLI	UPF0033 protein YeeD
P0AFT8	YEIW_ECOLI	UPF0153 protein YeiW
P67553	YNFC_ECOLI	UPF0257 lipoprotein YnfC
P0C037	YAIE_ECOLI	UPF0345 protein YaiE

P75728	UBIF_ECOLI	2-octaprenyl-3-methyl-6-methoxy-1,4-benzoquinol hydroxylase (EC 1.14.13.-)
P0A7W7	RS8_ECOLI	30S ribosomal protein S8
P77580	ACDH_ECOLI	Acetaldehyde dehydrogenase (EC 1.2.1.10) (Acetaldehyde dehydrogenase [acetylating])
P0A9Q5	ACCD_ECOLI	Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta (ACCase subunit beta) (Acetyl-CoA carboxylase carboxyltransferase subunit beta) (EC 6.4.1.2)
P00936	CYAA_ECOLI	Adenylate cyclase (EC 4.6.1.1) (ATP pyrophosphate-lyase) (Adenylyl cyclase)
P04825	AMPN_ECOLI	Aminopeptidase N (EC 3.4.11.2) (Alpha-aminoacylpeptide hydrolase)
Q46864	MQSA_ECOLI	Antitoxin MqsA
P0A9E0	ARAC_ECOLI	Arabinose operon regulatory protein
P0ABB0	ATPA_ECOLI	ATP synthase subunit alpha (EC 3.6.3.14) (ATP synthase F1 sector subunit alpha) (F-ATPase subunit alpha)
P69432	PGAD_ECOLI	Biofilm PGA synthesis protein PgaD
P12996	BIOB_ECOLI	Biotin synthase (EC 2.8.1.6)
P0ACJ8	CRP_ECOLI	cAMP-activated global transcriptional regulator CRP (Catabolite activator protein) (CAP) (Catabolite gene activator) (cAMP receptor protein) (CRP) (cAMP regulatory protein)
P30131	HYPF_ECOLI	Carbamoyltransferase HypF (EC 2.1.3.-) (Carbamoyl phosphate-converting enzyme HypF) ([NiFe]-hydrogenase maturation factor HypF) (Hydrogenase maturation protein HypF)
P0A9A6	FTSZ_ECOLI	Cell division protein FtsZ
P75862	ZAPC_ECOLI	Cell division protein ZapC (FtsZ-associated protein C) (Z-ring-associated protein C)
P0A9I1	CITE_ECOLI	Citrate lyase subunit beta (Citrase beta chain) (EC 4.1.3.6) (Citrate (pro-3S)-lyase subunit beta) (Citryl-CoA lyase subunit) (EC 4.1.3.34)
Q46896	CAS1_ECOLI	CRISPR-associated endonuclease Cas1 (EC 3.1.-.-)
P11989	BGLG_ECOLI	Cryptic beta-glucoside bgl operon antiterminator
P16703	CYSM_ECOLI	Cysteine synthase B (CSase B) (EC 2.5.1.47) (O-acetylserine (thiol)-lyase B) (OAS-TL B) (O-acetylserine sulfhydrylase B)
P0AFI5	PBP7_ECOLI	D-alanyl-D-alanine endopeptidase (DD-endopeptidase) (EC 3.4.21.-) (Penicillin-binding protein 7) (PBP-7)
P28638	YHDJ_ECOLI	DNA adenine methyltransferase YhdJ (EC 2.1.1.72)

P15042	DNLJ_ECOLI	DNA ligase (EC 6.5.1.2) (Polydeoxyribonucleotide synthase [NAD(+)])
P0A988	DPO3B_ECOLI	DNA polymerase III subunit beta (EC 2.7.7.7) (Beta sliding clamp) (Beta clamp)
P28630	HOLA_ECOLI	DNA polymerase III subunit delta (EC 2.7.7.7)
P03007	DPO3E_ECOLI	DNA polymerase III subunit epsilon (EC 2.7.7.7)
P06710	DPO3X_ECOLI	DNA polymerase III subunit tau (EC 2.7.7.7) (DNA polymerase III subunit gamma)
P0ABE2	BOLA_ECOLI	DNA-binding transcriptional regulator BolA
P0AGB6	RPOE_ECOLI	ECF RNA polymerase sigma-E factor (RNA polymerase sigma-E factor) (Sigma-24)
P0AAK4	HYDN_ECOLI	Electron transport protein HydN
P19636	EUTC_ECOLI	Ethanolamine ammonia-lyase light chain (EC 4.3.1.7) (Ethanolamine ammonia-lyase small subunit)
P0AEJ8	EUTN_ECOLI	Ethanolamine utilization protein EutN
P15028	FECB_ECOLI	Fe(3+) dicitrate-binding periplasmic protein (Iron(III) dicitrate-binding periplasmic protein)
P68646	FIXX_ECOLI	Ferredoxin-like protein FixX
P39405	FHUF_ECOLI	Ferric iron reductase protein FhuF
P0ABZ1	FLIG_ECOLI	Flagellar motor switch protein FliG
P06974	FLIM_ECOLI	Flagellar motor switch protein FliM
P61949	FLAV_ECOLI	Flavodoxin-1
P0ACL5	GLCC_ECOLI	Glc operon transcriptional activator
P22634	MURI_ECOLI	Glutamate racemase (EC 5.1.1.3)
P00960	SYGA_ECOLI	Glycine--tRNA ligase alpha subunit (EC 6.1.1.14) (Glycyl-tRNA synthetase alpha subunit) (GlyRS)
P27254	ARGK_ECOLI	GTPase ArgK (EC 3.6.5.-) (G-protein chaperone)
P06616	ERA_ECOLI	GTPase Era (ERA) (GTP-binding protein Era)
P0ACG8	HSLR_ECOLI	Heat shock protein 15 (HSP15)
P0ACS9	ACRR_ECOLI	HTH-type transcriptional regulator AcrR (Potential acrAB operon repressor)
P24242	ASCG_ECOLI	HTH-type transcriptional regulator AscG (Cryptic asc operon repressor)
P17410	CHBR_ECOLI	HTH-type transcriptional regulator ChbR (Chb operon repressor)
P06993	MALT_ECOLI	HTH-type transcriptional regulator MalT (ATP-dependent transcriptional activator MalT)

P19930	HYAD_ECOLI	Hydrogenase 1 maturation protease (EC 3.4.23.-)
P24193	HYPE_ECOLI	Hydrogenase isoenzymes formation protein HypE
P69741	MBHT_ECOLI	Hydrogenase-2 small chain (HYD2) (EC 1.12.99.6) (Membrane-bound hydrogenase 2 small subunit) (NiFe hydrogenase)
P0A7C2	LEXA_ECOLI	LexA repressor (EC 3.4.21.88)
P75957	LOLD_ECOLI	Lipoprotein-releasing system ATP-binding protein LolD (EC 3.6.3.-)
P37680	SGBE_ECOLI	L-ribulose-5-phosphate 4-epimerase SgbE (EC 5.1.3.4) (Phosphoribulose isomerase)
P07000	PLDB_ECOLI	Lysophospholipase L2 (EC 3.1.1.5) (Lecithinase B)
P0A8N0	MATP_ECOLI	Macrodomain Ter protein
P41052	MLTB_ECOLI	Membrane-bound lytic murein transglycosylase B (EC 4.2.2.n1) (35 kDa soluble lytic transglycosylase) (Murein hydrolase B) (Slt35)
P76344	ZINT_ECOLI	Metal-binding protein ZinT (Cadmium-induced protein ZinT)
P77806	YBDL_ECOLI	Methionine aminotransferase (EC 2.6.1.88) (Methionine-oxo-acid transaminase)
P30750	METN_ECOLI	Methionine import ATP-binding protein MetN (EC 3.6.3.-)
P30749	MOAE_ECOLI	Molybdopterin synthase catalytic subunit (EC 2.8.1.12) (MPT synthase subunit 2) (Molybdenum cofactor biosynthesis protein E) (Molybdopterin-converting factor large subunit) (Molybdopterin-converting factor subunit 2)
P75824	HCR_ECOLI	NADH oxidoreductase HCR (EC 1.-.-.-)
P0AFD1	NUOE_ECOLI	NADH-quinone oxidoreductase subunit E (EC 1.6.5.11) (NADH dehydrogenase I subunit E) (NDH-1 subunit E) (NUO5)
P0A6Z6	NIKR_ECOLI	Nickel-responsive regulator
P0AFB5	NTRB_ECOLI	Nitrogen regulation protein NR(II) (EC 2.7.13.3)
P0AB38	LPOB_ECOLI	Penicillin-binding protein activator LpoB (PBP activator LpoB) (Lipoprotein activator of PBP from the outer membrane B)
P08312	SYFA_ECOLI	Phenylalanine--tRNA ligase alpha subunit (EC 6.1.1.20) (Phenylalanyl-tRNA synthetase alpha subunit) (PheRS)
P76502	SIXA_ECOLI	Phosphohistidine phosphatase SixA (EC 3.1.3.-) (RX6)
P0ACZ4	EVGA_ECOLI	Positive transcription regulator EvgA
P23862	PRIC_ECOLI	Primosomal replication protein N"
P19317	NARW_ECOLI	Probable nitrate reductase molybdenum cofactor assembly chaperone NarW (Redox enzyme maturation protein NarW)

Q46953	YPJF_ECOLI	Probable toxin YpjF
P77609	FLXA_ECOLI	Protein FlxA
P0AFW4	RNK_ECOLI	Regulator of nucleoside diphosphate kinase
P04983	RBSA_ECOLI	Ribose import ATP-binding protein RbsA (EC 3.6.3.17)
P02925	RBSB_ECOLI	Ribose import binding protein RbsB
P0A805	RRF_ECOLI	Ribosome-recycling factor (RRF) (Ribosome-releasing factor)
P45577	PROQ_ECOLI	RNA chaperone ProQ
P00579	RPOD_ECOLI	RNA polymerase sigma factor RpoD (Sigma-70)
P18196	MINC_ECOLI	Septum site-determining protein MinC
P09153	TFAE_ECOLI	Tail fiber assembly protein homolog from lambdoid prophage e14
P0A9P4	TRXB_ECOLI	Thioredoxin reductase (TRXR) (EC 1.8.1.9)
P0AA25	THIO_ECOLI	Thioredoxin-1 (Trx-1)
P0A9F1	MNTR_ECOLI	Transcriptional regulator MntR (Manganese transport regulator)
P0AEF4	DPIA_ECOLI	Transcriptional regulatory protein DpiA (Destabilizer of plasmid inheritance)
P75679	INSN1_ECOLI	Transposase InsN for insertion sequence element IS911A
P0AGJ2	TRMH_ECOLI	tRNA (guanosine(18)-2'-O)-methyltransferase (EC 2.1.1.34) (tRNA [Gm18] methyltransferase)
P17993	UBIG_ECOLI	Ubiquinone biosynthesis O-methyltransferase (2-octaprenyl-6-hydroxyphenol methylase) (EC 2.1.1.222) (3-demethylubiquinone-8 3-O-methyltransferase) (EC 2.1.1.64)
P09147	GALE_ECOLI	UDP-glucose 4-epimerase (EC 5.1.3.2) (Galactowaldenase) (UDP-galactose 4-epimerase)
P0A8F8	UVRB_ECOLI	UvrABC system protein B (Protein UvrB) (Excinuclease ABC subunit B)

Appendix Table 4.2 – *Eco* binary interactions for hub proteins YeeD, YbhK, YpdC, YffB and YjjW and their homologs in *Vch*, *Sau*, and *Ype*

protein not present in the cloneset

Prey Name	Bait Name	Eco Prey	Eco Bait	Vch Prey	Vch Bait	Ype Prey	Ype Bait	Sau Prey	Sau Bait
yjgL	ybhK	P39336	P75767	0	Q9KT82	0	P58589	0	A0A0H2WWK1
htrE	ybhK	P33129	P75767	0	Q9KT82	Q7CIW3	P58589	0	A0A0H2WWK1
ingK	ybhK	P0AEW6	P75767	Q9KSX7	Q9KT82	Q7CK02	P58589	0	A0A0H2WWK1
ybhK	ybhK	P75767	P75767	Q9KT82	Q9KT82	P58589	P58589	A0A0H2WWK1	A0A0H2WWK1
ligA	ybhK	P15042	P75767	Q9KTD1	Q9KT82	Q7CJF3	P58589	Q5HEL8	A0A0H2WWK1
leuS	ybhK	P07813	P75767	Q9KTE6	Q9KT82	Q8ZDF8	P58589	Q5HF16	A0A0H2WWK1
hyaD	ybhK	P19930	P75767	0	Q9KT82	0	P58589	Q5HF34	A0A0H2WWK1
rbsB	ybhK	P02925	P75767	Q9KNQ4	Q9KT82	Q8D133	P58589	Q5HF38	A0A0H2WWK1
top1	ybhK	P06612	P75767	Q9KRB2	Q9KT82	Q7CIL8	P58589	Q5HGI2	A0A0H2WWK1
ileS	ybhK	P00956	P75767	Q9KU47	Q9KT82	Q8ZIM0	P58589	Q5HGN8	A0A0H2WWK1
hiuH	yeeD	P76341	P33014	0	Q9KVW4	0	Q7CKR7	0	A0A0H2WWL7
yicS	yeeD	Q2M7X4	P33014	0	Q9KVW4	0	Q7CKR7	0	A0A0H2WWL7
yijK	yeeD	P69816	P33014	Q9KR26	Q9KVW4	Q7CG43	Q7CKR7	0	A0A0H2WWL7
dapE	yeeD	P0AED7	P33014	Q9KQ52	Q9KVW4	Q7CJI9	Q7CKR7	0	A0A0H2WWL7
thiL	yeeD	P0AGG0	P33014	Q9KPU6	Q9KVW4	Q7CK42	Q7CKR7	0	A0A0H2WWL7
hyi	yeeD	P30147	P33014	0	Q9KVW4	Q8D019	Q7CKR7	0	A0A0H2WWL7
garL	yeeD	P23522	P33014	0	Q9KVW4	Q8D095	Q7CKR7	0	A0A0H2WWL7
insO1	yeeD	P75680	P33014	Q9K3D5	Q9KVW4	Q8D1M4	Q7CKR7	0	A0A0H2WWL7

clpB	yeeD	P63284	P33014	Q9KU18	Q9KVW4	Q74X11	Q7CKR7	Q5HD02	A0A0H2WWL7
evgA	yeeD	P0ACZ4	P33014	Q9KSP3	Q9KVW4	Q9ZC35	Q7CKR7	Q5HDG5	A0A0H2WWL7
truA	yeeD	P07649	P33014	Q9KTA4	Q9KVW4	Q8ZD27	Q7CKR7	Q5HDY9	A0A0H2WWL7
murA	yeeD	P0A749	P33014	Q9KP62	Q9KVW4	Q8ZB56	Q7CKR7	Q5HEA0	A0A0H2WWL7
ddlB	yeeD	P07862	P33014	Q9KV62	Q9KVW4	Q8ZIE7	Q7CKR7	Q5HEB7	A0A0H2WWL7
treR	yeeD	P36673	P33014	Q9KTJ3	Q9KVW4	Q7CL82	Q7CKR7	Q5HF38	A0A0H2WWL7
glxR	yeeD	P77161	P33014	0	Q9KVW4	Q7CL58	Q7CKR7	Q5HFR2	A0A0H2WWL7
aroB	yeeD	P07639	P33014	Q9KNV2	Q9KVW4	Q8ZJF6	Q7CKR7	Q5HFV8	A0A0H2WWL7
rsmJ	yeeD	P68567	P33014	Q9KVR6	Q9KVW4	Q8ZA46	Q7CKR7	Q5HH01	Q5HJF3
gabT	yeeD	P22256	P33014	Q9KNW2	Q9KVW4	Q8D0Y8	Q7CKR7	Q5HHC8	A0A0H2WWL7
yahK	yeeD	P75691	P33014	Q9KUG9	Q9KVW4	Q8D070	Q7CKR7	Q5HI63	A0A0H2WWL7
cysM	yeeD	P16703	P33014	Q9KUI4	Q9KVW4	Q7CJG9	Q7CKR7	Q5HIG2	A0A0H2WWL7
hcr	yffB	P75824	P24178	Q9KVY6	Q9KR22	Q7CL82	Q7CJI8	0	A0A0H2WW65
proQ	yffB	P45577	P24178	Q9KVF4	Q9KR22	Q8D095	Q7CJI8	0	A0A0H2WW65
eutC	yffB	P19636	P24178	Q9KR22	Q9KR22	Q8ZB56	Q7CJI8	0	A0A0H2WW65
rrf	yffB	P0A805	P24178	Q9KQ54	Q9KQ54	Q8ZIM0	Q7CJI8	Q5HGH2	A0A0H2WW65
yffB	yffB	P24178	P24178	Q9KPV5	Q9KQ54	Q7CJI9	Q7CJI8	Q5HH87	A0A0H2WW65
metN	yffB	P30750	P24178	Q9KUW4	Q9KR22	Q7CK42	Q7CJI8	Q5HHK4	A0A0H2WW65
pflA	yffB	P0A9N4	P24178	Q9KMT8	Q9KR22	Q8ZDF8	Q7CJI8	Q5HJF3	A0A0H2WW65
ldcA	yjjW	P76008	P39409	0	Q9KQX8	0	Q8D043	0	Q5HJF3
modE	yjjW	P0A9G8	P39409	Q9KQX5	Q9KQX8	P58497	Q8D043	0	Q5HJF3
ymbA	yjjW	P0AB10	P39409	Q9KR87	Q9KQX8	Q7CHK9	Q8D043	0	Q5HJF3

ycgL	yjjW	P0AB43	P39409	Q9KQP1	Q9KQX8	Q7CID0	Q8D043	0	Q5HJF3
suhB	yjjW	P0ADG4	P39409	Q9KTY5	Q9KQX8	Q7CJN9	Q8D043	0	Q5HJF3
bolA	yjjW	P0ABE2	P39409	Q9KPS0	Q9KQX8	Q7CK31	Q8D043	0	Q5HJF3
yabP	yjjW	P39220	P39409	0	Q9KQX8	Q7CKK5	Q8D043	0	Q5HJF3
ftsN	yjjW	P29131	P39409	Q9KNQ5	Q9KQX8	Q7CL21	Q8D043	0	Q5HJF3
ilvN	yjjW	P0ADF8	P39409	Q9KP91	Q9KQX8	Q8D0H3	Q8D043	0	Q5HJF3
glrR	yjjW	P0AFU4	P39409	Q9KNI9	Q9KQX8	Q8D0Z8	Q8D043	0	Q5HJF3
yaiE	yjjW	P0C037	P39409	Q9KRM2	Q9KQX8	Q8ZC17	Q8D043	0	Q5HJF3
minC	yjjW	P18196	P39409	Q9KQN9	Q9KQX8	Q8ZES5	Q8D043	0	Q5HJF3
fpg	yjjW	P05523	P39409	Q9KVC5	Q9KQX8	Q8ZJP0	Q8D043	0	Q5HJF3
galE	yjjW	P09147	P39409	H9L4R0	Q9KQX8	Q9F7D4	Q8D043	0	Q5HJF3
rsmJ	yjjW	P68567	P39409	Q9KVR6	Q9KQX8	Q8ZA46	Q8D043	0	A0A0H2WWL7
dpiA	yjjW	P0AEF4	P39409	Q9KTU7	Q9KQX8	Q8ZBV2	Q8D043	A0A0H2WXP1	Q5HJF3
feoC	yjjW	P64638	P39409	Q9KR16	Q9KQX8	Q7CFX5	Q8D043	Q5HE73	Q5HJF3
gnsA	yjjW	P0AC92	P39409	Q9KU40	Q9KQX8	0	Q8D043	Q5HEF5	Q5HJF3
xerC	yjjW	P0A8P6	P39409	Q9KVL4	Q9KQX8	Q8D1K0	Q8D043	Q5HFS5	Q5HJF3
murI	yjjW	P22634	P39409	Q9KVI7	Q9KQX8	Q8ZAA2	Q8D043	Q5HGT3	Q5HJF3
hcaD	yjjW	P77650	P39409	Q9KQV8	Q9KQX8	Q7CFW0	Q8D043	Q5HHB4	Q5HJF3
yaiI	yjjW	P0A8D3	P39409	Q9KTM1	Q9KQX8	Q8ZCF8	Q8D043	Q5HHY7	Q5HJF3
ydaG	yjjW	P76061	P39409	0	Q9KQX8	Q7CG15	Q8D043	Q5HJI1	Q5HJF3
rhaS	ypdC	P09377	P77396	Q9KRY6	Q9KQ54	Q8ZES5	Q7CG42	0	Q5HJR8
mioC	ypdC	P03817	P77396	0	Q9KQ54	Q8D0H3	Q7CG42	A0A0H2X0T1	Q5HJR8
rpoD	ypdC	P00579	P77396	Q9KR22	Q9KR22	Q7CFW0	Q7CG42	Q5HFJ9	Q5HJR8
araC	ypdC	P0A9E0	P77396	0	Q9KQ54	0	Q7CG42	Q5HJR8	Q5HJR8
chbR	ypdC	P17410	P77396	Q9KQX8	Q9KQ54	Q8ZJ00	Q7CG42	Q5HJR8	Q5HJR8
ypdC	ypdC	P77396	P77396	Q9KUK1	Q9KR22	Q7CJR9	Q7CG42	Q5HJR8	Q5HJR8
lexA	ypdC	P0A7C2	P77396	Q9KTJ5	Q9KQ54	Q7CKK5	Q7CG42	Q9L4P1	Q5HJR8

Appendix Table 4.3 – Function of the proteins interacting with hub protein YeeD in *Escherichia coli* based on UniProt.

Prey	**Ype**	**Vch**	**Sau**	**Protein name**	**Function**
AroB	x	x	x	3-dehydroquinate synthase	This protein is involved in step 2 of the subpathway that synthesizes chorismate from D-erythrose 4-phosphate and phosphoenolpyruvate.
ClpB	x	x	x	Chaperone protein ClpB	Part of a stress-induced multi-chaperone system, it is involved in the recovery of the cell from heat-induced damage, in cooperation with DnaK, DnaJ and GrpE. Acts before DnaK, in the processing of protein aggregates. Protein binding stimulates the ATPase activity; ATP hydrolysis unfolds the denatured protein aggregates, which probably helps expose new hydrophobic binding sites on the surface of ClpB-bound aggregates, contributing to the solubilization and refolding of denatured protein aggregates by DnaK.
CysM	x	x	x	Cysteine synthase B	Two cysteine synthase enzymes are found. Both catalyze the same reaction. Cysteine synthase B can also use thiosulfate in place of sulfide to give cysteine thiosulfonate as a product.
DapE	x	x		Succinyl-diaminopimelate desuccinylase	Catalyzes the hydrolysis of N-succinyl-L,L-diaminopimelic acid (SDAP), forming succinate and LL-2,6-diaminoheptanedioate (DAP), an intermediate involved in the bacterial biosynthesis of lysine and meso-diaminopimelic acid, an essential component of bacterial cell walls.
DdlB	x	x	x	D-alanine--D-alanine ligase B	Cell wall formation.
EvgA	x	x	x	Positive transcription regulator EvgA	Member of the two-component regulatory system EvgS/EvgA. Regulates the expression of emrKY operon and yfdX. Seems also to control expression of at least one other multidrug efflux operon.
GabT	x	x	x	4-aminobutyrate aminotransferase GabT	Catalyzes the transfer of the amino group from gamma-aminobutyrate (GABA) to alpha-ketoglutarate (KG) to yield succinic semialdehyde (SSA).

GarL	x			5-keto-4-deoxy-D-glucarate aldolase	Catalyzes the reversible retro-aldol cleavage of both 5-keto-4-deoxy-D-glucarate and 2-keto-3-deoxy-D-glucarate to pyruvate and tartronic semialdehyde.
GlxR	x		x	2-hydroxy-3-oxopropionate reductase	This protein is involved in step 3 of the subpathway that synthesizes 3-phospho-D-glycerate from glycolate.
HiuH				5-hydroxyisourate hydrolase	Catalyzes the hydrolysis of 5-hydroxyisourate (HIU) to 2-oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline (OHCU).
Hyi	x			Hydroxypyruvate isomerase	Catalyzes the reversible isomerization between hydroxypyruvate and 2-hydroxy-3-oxopropanoate (also termed tartronate semialdehyde). Does not catalyze the isomerization of D-fructose to D-glucose or that of D-xylulose to D-xylose. Also does not catalyze racemization of serine, alanine, glycerate or lactate.
InsO1	x	x		Putative transposase InsO for insertion sequence element IS911A	
MurA	x	x	x	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	Cell wall formation
RsmJ	x	x		Ribosomal RNA small subunit methyltransferase	Specifically methylates the guanosine in position 1516 of 16S rRNA.
ThiL	x	x		Thiamine-monophosphate kinase	Catalyzes the ATP-dependent phosphorylation of thiamine-monophosphate (TMP) to form thiamine-pyrophosphate (TPP), the active form of vitamin B1. Cannot use thiamine as substrate. Is highly specific for ATP as phosphate donor.
TreR	x	x	x	HTH-type transcriptional regulator TreR	Repressor of the treBC operon. It is able to bind trehalose-6-phosphate and trehalose.2 Publications
TruA	x	x	x	tRNA pseudouridine synthase A	Formation of pseudouridine at positions 38, 39 and 40 in the anticodon stem and loop of transfer RNAs.1 Publication

YahK	x	x	x	Aldehyde reductase YahK	Catalyzes the reduction of a wide range of aldehydes into their corresponding alcohols. Has a strong preference for NADPH over NADH as the electron donor. Cannot use a ketone as substrate. Is a major source of NADPH-dependent aldehyde reductase activity in E.coli. The in vivo functions of YahK has yet to be determined.
YicS				Uncharacterized protein YicS	
YijK	x	x		PTS system fructose-like EIIB component 2	The phosphoenolpyruvate-dependent sugar phosphotransferase system (sugar PTS), a major carbohydrate active transport system, catalyzes the phosphorylation of incoming sugar substrates concomitantly with their translocation across the cell membrane. The enzyme II FrwABC PTS system is involved in fructose transport.

Appendix Table 4.4 – Function of the proteins interacting with hub protein YjjW in *Escherichia coli* based on UniProt.

Prey	Ype	Vch	Sau	Protein Name	Function
BolA	x	x		DNA-binding transcriptional regulator BolA	Transcriptional regulator that plays an important role in general stress response. Has many effects on cell morphology, cell growth and cell division. Acts by regulating the transcription of many genes, including dacA (PBP-5), dacC (PBP-6), ampC and mreB. Probably involved in the coordination of genes that adapt the cell physiology in order to enhance cell adaptation and survival under stress conditions. Essential for normal cell morphology in stationary phase and under conditions of starvation
DpiA	x	x	x	Transcriptional regulatory protein ...	Member of the two-component regulatory system DpiA/DpiB, which is essential for expression of citrate-specific fermentation genes and genes involved in plasmid inheritance. Could be involved in response to both the presence of citrate and external redox conditions. Regulates the transcription of citCDEFXGT, dpiAB, mdh and exuT. Binds specifically to the dpiB-citC intergenic region.
FeoC	x	x	x	Probable [Fe-S]-dependent transcriptional repressor	May function as a transcriptional regulator that controls feoABC expression.
FtsN	x	x		Cell division protein FtsN	Essential cell division protein that activates septal peptidoglycan synthesis and constriction of the cell. Acts on both sides of the membrane, via interaction with FtsA in the cytoplasm and interaction with the FtsQBL complex in the periplasm. These interactions may induce a conformational switch in both FtsA and FtsQBL, leading to septal peptidoglycan synthesis by FtsI and associated synthases (Probable).
GalE	x	x		UDP-glucose 4-epimerase	Involved in the metabolism of galactose. Catalyzes the conversion of UDP-galactose (UDP-Gal) to UDP-glucose (UDP-Glc) through a mechanism involving the transient reduction of NAD. It is only active on UDP-galactose and UDP-glucose.

GlrR	x	x	x	Transcriptional regulatory protein GlrR	Member of the two-component regulatory system GlrR/GlrK that up-regulates transcription of the glmY sRNA when cells enter the stationary growth phase. Regulates glmY transcription by binding to three conserved sites in the purL-glmY intergenic region.
GnsA		x	x	Protein GnsA	Overexpression increases levels of unsaturated fatty acids and suppresses both the temperature-sensitive fabA6 mutation and cold-sensitive secG null mutation.
HcaD	x	x	x	3-phenylpropionate/cinnamic acid dioxygenase ferredoxin--NAD(+) reductase component	Part of the multicomponent 3-phenylpropionate dioxygenase, that converts 3-phenylpropionic acid (PP) and cinnamic acid (CI) into 3-phenylpropionate-dihydrodiol (PP-dihydrodiol) and cinnamic acid-dihydrodiol (CI-dihydrodiol), respectively.
IlvN	x	x		Acetolactate synthase isozyme 1 small subunit	This protein is involved in step 1 of the subpathway that synthesizes L-isoleucine from 2-oxobutanoate.
LdcA	x	x		Murein tetrapeptide carboxypeptidase	Releases the terminal D-alanine residue from the cytoplasmic tetrapeptide recycling product L-Ala-gamma-D-Glu-meso-Dap-D-Ala. To a lesser extent, can also cleave D-Ala from murein derivatives containing the tetrapeptide, i.e. MurNAc-tetrapeptide, UDP-MurNAc-tetrapeptide, GlcNAc-MurNAc-tetrapeptide, and GlcNAc-anhMurNAc-tetrapeptide. Does not act on murein sacculi or cross-linked muropeptides. The tripeptides produced by the LcdA reaction can then be reused as peptidoglycan building blocks; LcdA is thereby involved in murein recycling. Is also essential for viability during stationary phase.2
MinC	x	x		Septum site-determining protein MinC	Cell division inhibitor that blocks the formation of polar Z ring septums. Rapidly oscillates between the poles of the cell to destabilize FtsZ filaments that have formed before they mature into polar Z rings. Prevents FtsZ polymerization.

ModE	x	x		Transcriptional regulator ModE	The ModE-Mo complex acts as a repressor of the modABC operon, involved in the transport of molybdate. Upon binding molybdate, the conformation of the protein changes, promoting dimerization of ModE-Mo. The protein dimer is then competent to bind a DNA region, upstream of the modABC operon, which contains an 8-base inverted repeat 5'-TAACGTTA-3' flanked by two CAT boxes. Acts also as an enhancer of the expression of genes coding for molybdoenzymes, both directly and indirectly. ModE also interacts with tungstate.
MurI	x	x	x	Glutamate racemase	Provides the (R)-glutamate required for cell wall biosynthesis.
MutM	x	x		Formamidopyrimidine-DNA glycosylase	Involved in base excision repair of DNA damaged by oxidation or by mutagenic agents. Acts as DNA glycosylase that recognizes and removes damaged bases. Has a preference for oxidized purines, such as 7,8-dihydro-8-oxoguanine (8-oxoG) and its derivatives such as guanidinohydantoin:C and spiroiminodihydantoin:C, however it also acts on thymine glycol:G, 5,6-dihydrouracil:G and 5-hydroxyuracil:G. Has AP (apurinic/apyrimidinic) lyase activity and introduces nicks in the DNA strand. Cleaves the DNA backbone by beta-delta elimination to generate a single-strand break at the site of the removed base with both 3'- and 5'-phosphates. Cleaves ssDNA containing an AP site.
RsmJ	x	x	x	Ribosomal RNA small subunit methyltransferase J	Specifically methylates the guanosine in position 1516 of 16S rRNA.
SuhB	x	x		Inositol-1-monophosphatase	The suhB gene has been shown to encode a protein with inositol monophosphatase and glycerol-2-phosphatase activity.

XerC	x	x	x	Tyrosine recombinase XerC	Site-specific tyrosine recombinase, which acts by catalyzing the cutting and rejoining of the recombining DNA molecules. Binds cooperatively to specific DNA consensus sequences that are separated from XerD binding sites by a short central region, forming the heterotetrameric XerC-XerD complex that recombines DNA substrates. The complex is essential to convert dimers of the bacterial chromosome into monomers to permit their segregation at cell division. It also contributes to the segregational stability of plasmids at ColE1 xer (or cer) and pSC101 (or psi) sites. In the complex XerC specifically exchanges the top DNA strands (By similarity).
YabP	x			Putative uncharacterized protein Ya...	
YaiE	x	x		UPF0345 protein YaiE	
YaiI	x	x	x	UPF0178 protein YaiI	
YcgL	x	x		Protein YcgL	
YdaG	x		x	Uncharacterized protein YdaG	

YmbA	x			Uncharacterized lipoprotein YmbA	

Appendix Table 4.5 – Function of the proteins interacting with hub protein YbhK in *Escherichia coli* based on UniProt.

Prey	Ype	Vch	Sau	Protein Name	Function
HtrE	x			Outer membrane usher protein HtrE	Part of the yadCKLM-htrE-yadVN fimbrial operon. Could contribute to adhesion to various surfaces in specific environmental niches. Probably involved in the export and assembly of fimbrial subunits across the outer membrane.
HyaD			x	Hydrogenase 1 maturation protease	Protease involved in the C-terminal processing of HyaB, the large subunit of hydrogenase 1.(Curated)
IleS	x	x	x	Isoleucine--tRNA ligase	Catalyzes the attachment of isoleucine to tRNA(Ile). As IleRS can inadvertently accommodate and process structurally similar amino acids such as valine, to avoid such errors it has two additional distinct tRNA(Ile)-dependent editing activities. One activity is designated as 'pretransfer' editing and involves the hydrolysis of activated Val-AMP. The other activity is designated 'posttransfer' editing and involves deacylation of mischarged Val-tRNA(Ile).
IngK	x	x		Inosine-guanosine kinase	This protein is involved in step **1** of the subpathway that synthesizes IMP from inosine. This subpathway is part of the pathway IMP biosynthesis via salvage pathway, which is itself part of Purine metabolism.
LeuS	x	x	x	Leucine--tRNA ligase	Leucine-tRNA ligase (LeuRS) is a member of the family of aminoacyl tRNA synthetases, which interpret the genetic code by covalently linking amino acids to their specific tRNA molecules. The reaction is driven by ATP hydrolysis. LeuRS belongs to the Class IB aminoacyl tRNA synthetases; apart from sequence motifs within the active site, the different enzymes show little similarity in their primary amino acid sequences.

LigA	x	x	x	DNA ligase	DNA ligase that catalyzes the formation of phosphodiester linkages between 5'-phosphoryl and 3'-hydroxyl groups in double-stranded DNA using NAD as a coenzyme and as the energy source for the reaction. It is essential for DNA replication and repair of damaged DNA.
RbsB	x	x	x	Ribose import binding protein RbsB	Part of the ABC transporter complex RbsABC involved in ribose import. Binds ribose. Also serves as the primary chemoreceptor for chemotaxis.
TopA	x	x	x	DNA topoisomerase 1	Releases the supercoiling and torsional tension of DNA, which is introduced during the DNA replication and transcription, by transiently cleaving and rejoining one strand of the DNA duplex. Introduces a single-strand break via transesterification at a target site in duplex DNA. The scissile phosphodiester is attacked by the catalytic tyrosine of the enzyme, resulting in the formation of a DNA-(5'-phosphotyrosyl)-enzyme intermediate and the expulsion of a 3'-OH DNA strand. The free DNA strand then undergoes passage around the unbroken strand, thus removing DNA supercoils. Finally, in the religation step, the DNA 3'-OH attacks the covalent intermediate to expel the active-site tyrosine and restore the DNA phosphodiester backbone.
YjgL				Uncharacterized protein YjgL	

Appendix Table 4.6 – Function of the proteins interacting with hub protein YffB in *Escherichia coli* based on UniProt.

Prey	Ype	Vch	Sau	Protein Name	Function
EutC				Ethanolamine ammonia-lyase light chain	This protein is involved in the pathway ethanolamine degradation, which is part of Amine and polyamine degradation.
Frr	x	x	x	Ribosome-recycling factor	Responsible for the release of ribosomes from messenger RNA at the termination of protein biosynthesis. May increase the efficiency of translation by recycling ribosomes from one round of translation to another.
Hcr	x			NADH oxidoreductase HCR	NADH oxidoreductase acting in concert with HCP.
InsN1				Transposase InsN for insertion sequence element IS911A	Involved in the transposition of the insertion sequence IS911.

MetN	x	x	x	Methionine import ATP-binding protein MetN	Part of the ABC transporter complex MetNIQ involved in methionine import. Responsible for energy coupling to the transport system (Probable). It has also been shown to be involved in formyl-L-methionine transport.
PflA	x	x	x	Pyruvate formate-lyase 1-activating...	Activation of pyruvate formate-lyase 1 under anaerobic conditions by generation of an organic free radical, using S-adenosylmethionine and reduced flavodoxin as cosubstrates to produce 5'-deoxy-adenosine.
ProQ	x	x		RNA chaperone ProQ	RNA chaperone with significant RNA binding, RNA strand exchange and RNA duplexing activities. May regulate ProP activity through an RNA-based, post-transcriptional mechanism.
YffB	x	x	x	Protein YffB	

VITA

Neha Sakhawalkar was born on December 4th 1987 in Bhopal, Madhya Pradesh. She is an Indian citizen. She graduated with a Bachelors' degree in Biotechnology from University of Pune in 2008. After that she enrolled for Masters' of Science in Biotechnology at University of Pune and graduated in 2010. She pursued a Master of Bioinformatics from Virginia Commonwealth University in 2012.