

**To Understand The Functional Co-relation between Proteome set of
Microbes by Phylogenetic Profiling**



Co-ordinator
PROF. INDIRA GHOSH
DIRECTOR
BIOINFORMATICS CENTRE
UNIVERSITY OF PUNE
PUNE 411007, INDIA

Submitted by:

AMITABH GUPTA
Enrollment No. - 040002

Course
B.Tech in Biotechnology
Jaypee Institute of Information Technology University.
(JIITU)



DECLARATION

I hereby declare that the project titled "**To Understand the Functional Correlation Between Proteome Set of Microbes by Phylogenetic Profiling**" is submitted in partial fulfillment of B.Tech in Biotechnology has been carried out by me at Bioinformatics Centre, University of Pune under the guidance of **Dr. Indira Ghosh**, Director, Bioinformatics Centre, University of Pune, Pune.

Any further extension, continuation or use of this project has to be undertaken with prior express written consent from the Director, Bioinformatics Center, University of Pune & Supervisor.

I further declare that the project work or any part thereof has not been previously submitted for any degree or diploma in any university.

Date:

Amitabh Gupta
Bioinformatics Centre
University of Pune,
Pune, India

Place:

CERTIFICATE

This is to certify that the report entitled, "To Understand the Functional Co-relation Between Proteome set of Microbes by Phylogenetic Profiling" submitted by Amitabh Gupta to the University of Pune, Pune 411007 in partial fulfillment of the requirements of the degree of B.Tech in Biotechnology, embodies work carried out by him in the Bioinformatics Centre, University of Pune during 1 June 2007 – 15 July 2007 under my guidance.

Materials obtained from other sources have been duly acknowledged.

Date:

Place:

Prof. Indira Ghosh
Bioinformatics Centre,
University of Pune,
Pune, India

Official Seal of Institute

ACKNOWLEDGEMENT

I extend my warm and sincere thanks to my project supervisor **Prof. Indira Ghosh** who was a staunch supporter and motivator throughout the project.

Right from the inception of this project work, my mentor Prachi Mam guided me till the very end in the true sense of the word. She always came up with innovative ways and creative terms thus also helping me to instill and enhance the quality of creative thinking within myself.

I would also like to express my gratitude to this alma mater Bioinformatics Centre, University of Pune for providing proper resources as and when required such as an all time internet facility and an excellent library.

I would also take this opportunity to thank Dr. Sanjay Londe, Rahul sir for making the necessary computer laboratory arrangements for me, Mr. Limbkar for making all arrangements required for work here.

Hence without giving a warm thanks to all of them who made this project work a reality my work would be incomplete.

Contents

	Page No.
1. Declaration.....	2
2. Certificate.....	3.
3. Acknowledgement.....	4.
4. Abstract.....	6
5. Objective.....	7
6. Introduction:.....	8
6.1 What is phylogenetic profiling?.....	9
6.2 Literature survey-	
• Phylogeny among microbes.....	10
• Molecular evolution in protein families.....	12
6.3 Assumption.....	14
7. Materials and Methods.....	15
7.1 Materials	
• Databases Used.....	16
• Programmes Used.....	19
7.2 Methodology.....	20
7.3 Detail Description.....	22.
8. Findings.....	24
8.1 Results.....	24
8.2 Discussion.....	27
8.3 Final extract.....	32
9. Inference.....	33
10. Future work.....	34
11. References.....	35.
12. Appendix.....	37

ABSTRACT:

To drive the protein function from the genome sequences is now the central goal of computational biology and evolution play the key role in this process. Phylogenetic profiling is based upon the phylogeny and homology among the protein sequences of organism. According to homology the functionally important sites in protein sequences conserve or co-related during evolutionary process and we represent these proteins as a string that encodes presence or absence of a protein in every known genome.

Here we applied phylogenetic profiling on the proteome set of microbes from different taxonomical hierarchy and analyze 5267 proteins of different microbes in reference of *Pseudomonas aeruginosa*. Here we assumed that proteins involved in metabolic pathways, cell transport machinery and cell division supposed to be evolved in correlated fashion during the evolutionary process or they may be eliminated or preserved in new species.

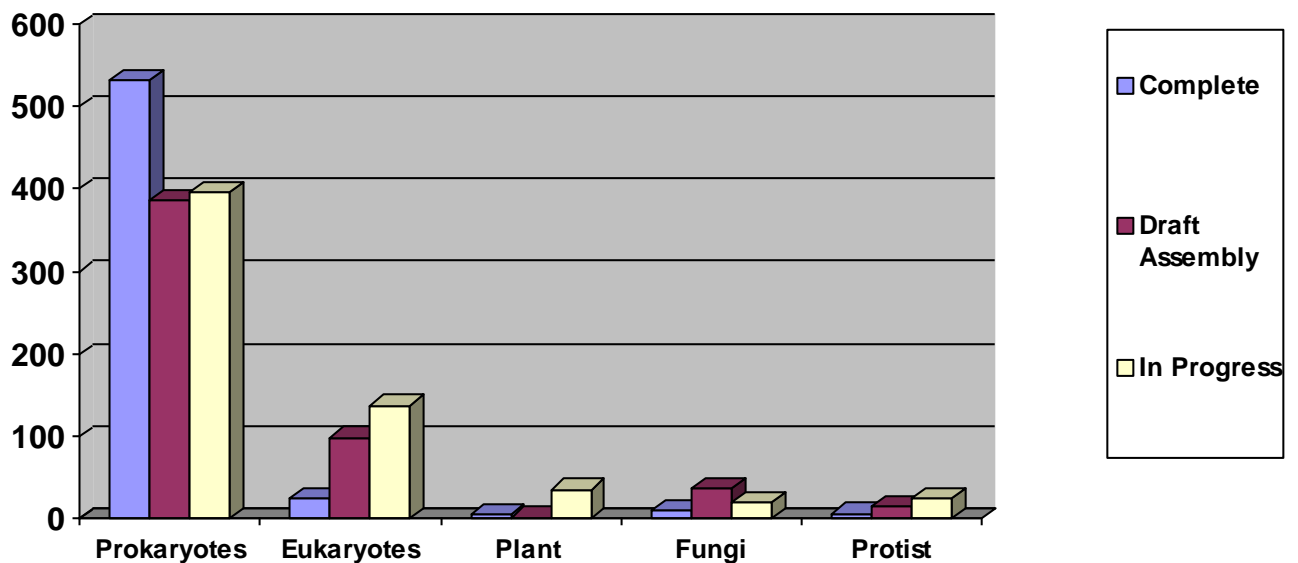
Objective:

Construct the database to create phylogenetic profile of whole proteome set for 49 microbes and analyze the similar phylogenetic profile proteins. With the help of that establish functional relationship between proteins of different classes and annotate function of uncharacterized protein, identified the factor responsible for multidrug resistance.

6.Introduction:

Now 0 scientists' sequenced complete genome of many organisms: 532 in case of prokaryotes: 491 **bacteria**, 41 **archaea**, 26 in case of **eukaryotes** and 485 genome sequenced as a draft assembly, 532 genome project in progress [9]. These statistics shows our biological database grows at exponential pace but our meaningful knowledge about these sequences is very limited. The central goal of computational biology is to determine protein function from genomic sequences. In last decade scientists discovered new techniques in genomics to functionally annotate uncharacterized protein sequences derived from the genome of organisms.

These techniques based upon the evolutionary process: **Rosetta stone method** based upon the gene fusion events between the two functionally related gene sequences during evolutionary process, **Gene neighbor method** based upon the conservation in location of two functionally related gene during evolutionary process but **Phylogenetic profiling** is the most sensitive technique discovered for functional annotation of uncharacterized protein [14].



Graph 6.1- shows the graphical representation of statistical data for genome projects.

6.1-What is Phylogenetic Profiling?

This method detects proteins that participate in a common structural complex or metabolic pathway. Proteins within these groups are defined as **functionally linked**. The underlying hypothesis is that functionally linked proteins evolve in a correlated fashion, and, therefore, they have homologs in the same subset of organisms [13].

In short, we show that if two proteins have homologs in the same subset of fully sequenced organisms, they are likely to be functionally linked. By exploiting this property, we can systematically map links between all the proteins coded by a genome.

In general, pairs of functionally linked proteins have no amino acid sequence similarity with each other and, therefore, cannot be linked by conventional sequence-alignment techniques.

We can make a phylogenetic profile of organisms by two approaches-

- ✓ With the help of genome of organisms.
- ✓ With the help of proteome of organisms.

In genomic approach, first we identified all the functional ORF's coded by genome for given subset of organisms, these ORF's (blastx) compare with protein set of reference organism, select significant ORF's corresponding to each protein.

Profile is a string with **n** entries: **n** corresponds to number of genomes. In profile the presence of homolog (ORF) to a given protein in the **n**th genome with an entry of **unity** at **n**th position. If no homolog is found the entry is **zero**.

In proteomic approach similar procedure is followed but take proteins rather than ORF's of organisms.

Proteins are **clustered** according to the similarity of their phylogenetic profiles. Similar profiles show a correlated pattern of inheritance and, by implication, **functional linkage**. The method predicts that the functions of uncharacterized proteins are likely to be similar to characterized proteins within a cluster.

In figure 2.1 method of analyzing protein by phylogenetic profiles is illustrated schematically for the hypothetical case of four fully sequenced genomes (from *E. coli*, *Saccharomyces cerevisiae*, *Haemophilus influenzae*, and *Bacillus subtilis*) in which they focus on seven proteins (P1–P7). For each *E. coli* protein, they construct a profile, indicating which genomes code for homologs of the protein. They next cluster the profiles to determine which proteins share the same profiles. Proteins with identical (or similar) profiles are boxed to indicate that they are likely to be functionally linked. Boxes connected by lines have phylogenetic profiles that differ by one bit and are termed **neighbors**.

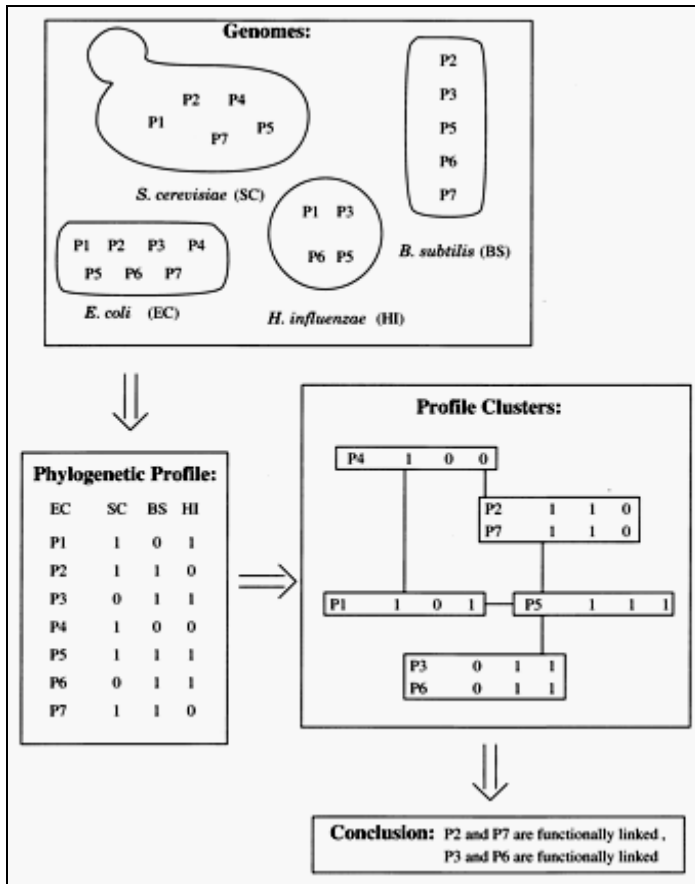


Figure 6.1 [14]

6.2-Literature survey:

For understanding phylogenetic profiling and its application on complete genome sequences of microorganism, it's mandatory to know about following prerequisite-

- ✓ Phylogeny among microbes.
- ✓ Molecular evolution in protein families.

I-Phylogeny among Microbes:

Phylogeny is the description of biological relationships, usually expressed as a **tree**. Phylogeny states as topology of the relationships based on the classification according to similarity of one or more sets of characters, or on a model of evolutionary process.

Molecular phylogenetics is the use of molecular sequences to construct evolutionary trees typically, we study a family of related sequences that we know evolved from a common ancestor and we want to know in which order these sequences diverged from one another. There are two approaches to deriving phylogenetic trees: **phenetic approach** is based on similarity; **cladistic approach** is based upon **genealogy** [2].

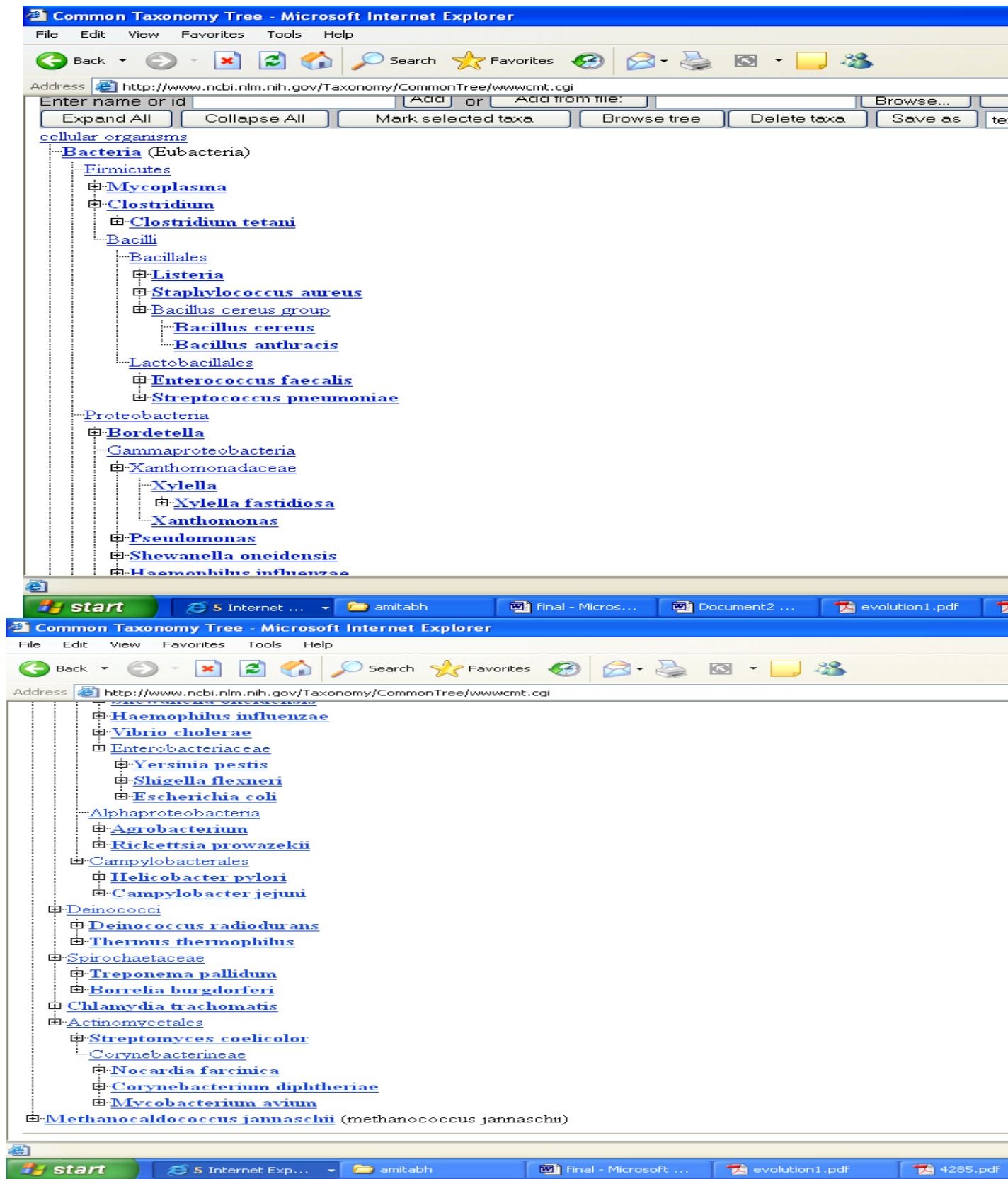


Figure-6.2- snapshot of taxonomic tree

In figure 6.2, we draw a phylogenetic tree for microbes of different **taxonomical hierarchy** [10], which are used in our method of phylogenetic profiling. Maximum number of bacteria belongs to class **proteobacteria**; in our study we select *Pseudomonas aeruginosa* as a reference organism, which belongs to genus **gamma proteobacteria**. Due to following reasons [11] we select *P.aeruginosa* as a **reference organism**-

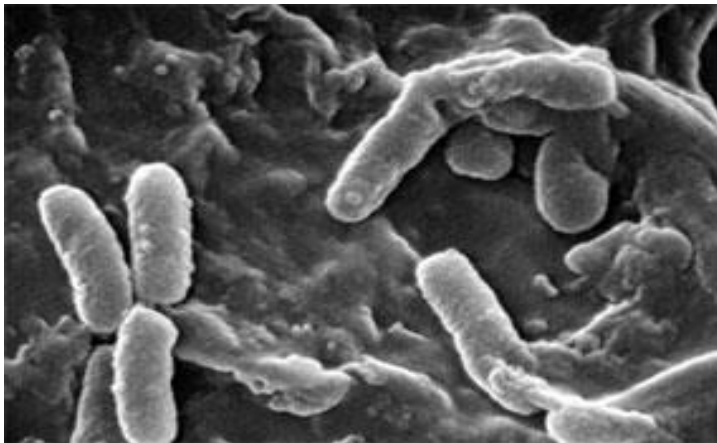


Figure 6.3- *Pseudomonas aeruginosa* [11].

- It is a **Gram-negative** bacterium that is noted for its **environmental versatility**, ability to cause disease in particular susceptible individuals, and its resistance to antibiotics.
- The bacterium is capable of utilizing a wide range of organic compounds as food sources, thus giving it an exceptional ability to colonize **ecological niches** where nutrients are limited.
- *P. aeruginosa* is widely studied by scientists who are interested in not only its ability to cause disease and resist antibiotics, but also its **metabolic capability and environmental versatility**.

Due to selecting microbes of different taxonomical hierarchy, we able to generate profile of complete proteome for closely as well as distantly related microbes and get wider idea which proteins conserve and functionally related during evolutionary process.

II-Molecular evolution in protein families:

In our experiment we used whole proteome of microbes, here I explain some basic concept regarding **protein evolution** to give more sense for result we obtained through phylogenetic profiling. **SCOP** (structural classification of protein database) is most popular classification of protein in evolutionary aspect and also governs **3D structure**. The unit of classification is protein **domain** [17]. The classification is on hierarchical levels that embody the evolutionary and structural relationships-

Family: Proteins are clustered together into families on the basis of one of two criteria that imply their having a common evolutionary origin: first, all proteins that have **residue identities** of 30% and greater; second, proteins with lower **sequence identities** but whose functions and structures are very similar; for example, globins with sequence identities of 15%.

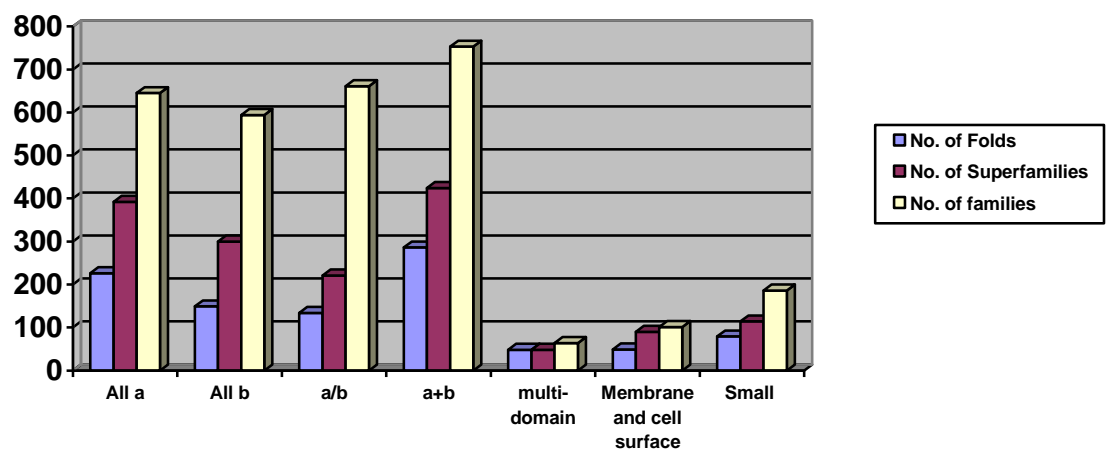
SUPERFAMILY: Families, whose proteins have low sequence identities but whose **structures** and, in many cases, **functional features** suggest that a **common evolutionary origin** is probable, are placed together in superfamilies; for example, actin, the ATPase domain of the heat-shock protein and hexokinase.

Common fold: Superfamilies and families are defined as having a common fold if their proteins have same major secondary structures in same arrangement with the same topological connections. There may, however, be cases where a common evolutionary origin is obscured by the extent of the **divergence** in sequence, **structure** and **function**. In these cases, it is possible that the discovery of new structures, with folds between those of the previously known structures, will make clear their common evolutionary relationship.

Class: The different folds have been grouped into classes. Most of the folds are assigned to one of the five structural classes on the basis of the secondary structures of which they composed:

- ✓ All α ,
- ✓ All β ,
- ✓ α Plus β ,
- ✓ α And β ,
- ✓ Multi-domain.

Graph 6.2 shows the statistics [12] of SCOP database classification elements-



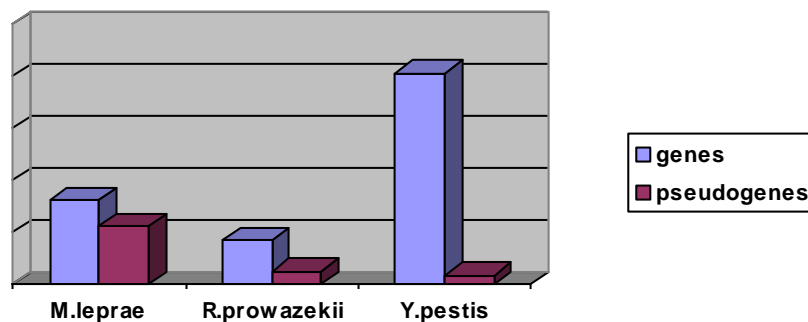
Graph 6.2

Here we give emphasis upon protein families, because the link between genome and proteome can be well established by the help of **evolution** in **protein families**.

Protein families can be used to understand many aspects of genomes, both their “live” and their “dead” parts (i.e. genes and pseudo genes). Some microbes have large protein families to perform cellular function and others have small protein families to perform similar function. The reduction of protein families **redundancy** observed in many pathogenic bacteria like *Mycobacterium leprae*, *Y.pestis*, *R.prowazekii* [15,16].

The *M. leprae* genome contains about, 1100 apparent pseudogenes and 1600 genes. This is a considerable reduction when compared to the, 4000 proteins encoded in the genome of the related bacterium *Mycobacterium tuberculosis* and involves decrease in the redundancy of almost all protein families, with loss of substantial parts of pathways, such as the **anaerobic respiratory chain**. This protein decay may be due to change in **environmental condition** and **ecological niche**.

Graph 6.3 showing the relative protein decay in reference to genome of organism-



Graph 6.3

After understanding reductive protein evolution in protein families, we can identify the function of uncharacterized or hypothetical protein which are present in normal microbes and absent in the microbes that undergo massive **proteome decay** with the help of phylogenetic analysis.

6.3-Assumption:

In this approach, we assumed that the **functional importance sites** of proteins are inversely related to the **evolutionary rate** of amino acid replacement [18]. This belief arises from one interpretation of the **neutral theory** in which sites of greatest functional significance are under the strongest selective constraints. An organism that experiences a replacement at one of these sites is less likely to survive and reproduce.

For example, the observation that a histidine is highly conserved during the evolution of a protein family is frequently taken as an indication that the residue is in the active site of the enzyme and is directly involved in catalytic function.

7. Materials AND Methodology:

Materials:

Following things used for creating phylogenetic profile of 49 microorganisms-

7.1- Database Used:

We used Expasy 's **HAMAP: High-quality Automated and Manual Annotation of microbial Proteomes**, for collecting whole proteome of 49 microbes of different taxonomical hierarchy including reference microbe *P.aeruginosa*. HAMAP database [5] contains total **566 proteomes** of Bacteria (421), archaea (39) and plastids (106).

HAMAP proteomes: Bacteria

Proteomes by kingdom: [Summary] **Bacteria** Archaea Plastids

UniProtKB/Swiss-Prot Release 53.1 of 12-Jun-2007: 270778 entries
UniProtKB/TrEMBL Release 36.1 of 12-Jun-2007: 4448557 entries

The UniProt Knowledgebase currently contains **421 complete proteomes of the kingdom Bacteria** that are covered by the HAMAP project. Whole-genome shotgun (WGS) and draft sequences are not covered by the HAMAP project. Only closed genomes are entered into the HAMAP database and are considered for annotation.

To retrieve information from a particular organism (description, genome reference paper(s), taxonomy or a file containing all the UniProt Knowledgebase entries)
or to perform a BLAST search against all potential proteins from this organism, click on the species code.

Species code	Species	Total number of entries	Number of UniProtKB/ Swiss-Prot entries	Number of UniProtKB/ TrEMBL entries	Status in UniProtKB/ Swiss-Prot
ACIAC	Acidovorax avenae subsp. citrulli (strain AAC00-1)	4602	15	4587	0%
ACIAD	Acinetobacter sp. (strain ADP1)	3310	383	2927	11%
ACIBL	Acidobacteria bacterium (strain Ellin345)	4771	83	4688	1%
ACIBT	Acinetobacter baumannii (strain ATCC 17978 / NCDC KC 755)	3365	1	3364	0%
ACIC1	Acidithiobacillus caldophilus (strain ATCC 43068 / 11B)	2157	0	2148	0%

Snapshot of Expasy's HAMAP database

Proteome size varies in correlation with genome size of microbe, reference organism *P.aeruginosa* proteome consists of 5568 proteins. *Mycoplasma genitalium* proteome consists of minimum number of proteins 483. Web address of database is <http://ca.expasy.org/sprot/hamap/>.

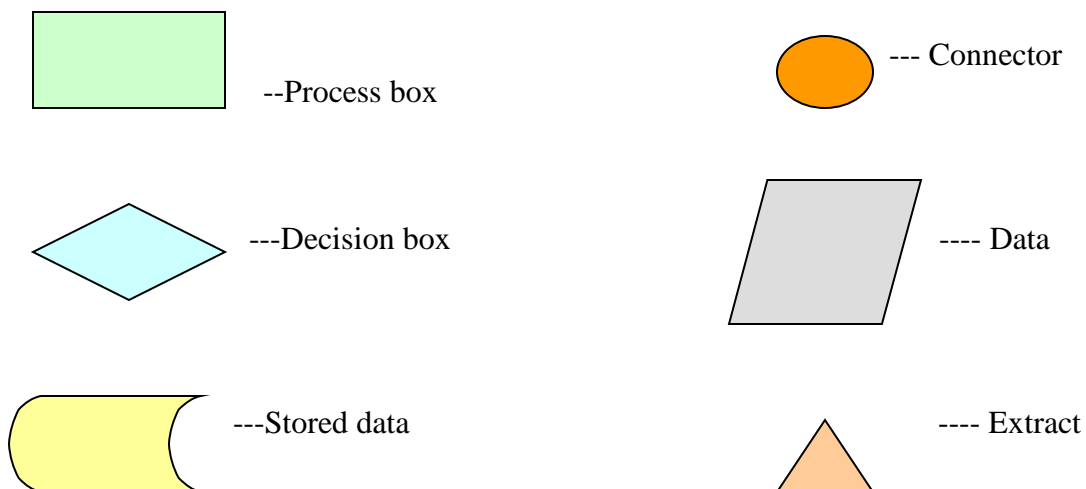
7.2- Programme Used:

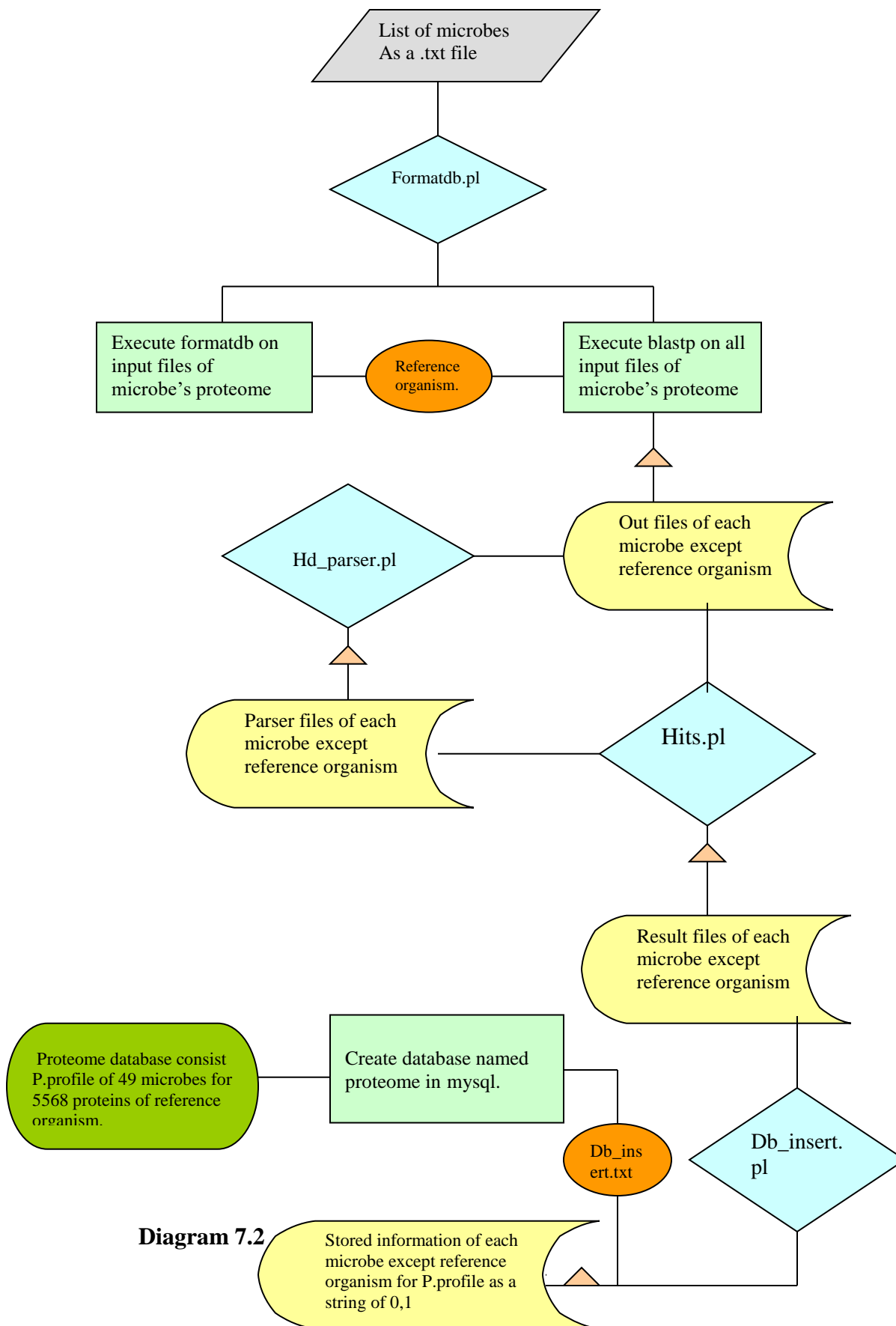
We used following programme for creating phylogenetic profile of selected microorganisms-

- BLAST (basic local sequence alignment tool)
- My sql server.
- Programme written in perl script –
 - **Formatdb.pl:** For creating blast output files for each microbe's proteome file in reference of *P.aeruginosa* proteome file.
 - **Hd_parser.pl:** For parsing the significant information from blast output files and creates parsing files for each microbe except reference organism.
 - **Hits.pl:** For creating result files for each microbe from parsing files
 - **Db_insert.pl:** Create database for phylogenic profiling in mysql.

7.3-Methodology:

Flow diagram explain the process of database creation for phylogenetic profiling:





7.4 - Detail Description:

In following steps, we are giving detail procedure which used to create database for phylogenetic profiling ---

- 1) First, we downloaded proteome of 50 microbes of different taxonomical hierarchy from expasy's HAMAP database in FASTA format.
- 2) Stored these proteome as a .txt file specific for each bacteria in input files folder, made another .txt file consist a list of names for all microbes selected with there four letter specific code.
- 3) With the help of blastp algorithm, we compare each microbe's input file with reference organism *P.aeruginosa* input file, which are listed in bacteria list text file. This process execute with the help of programme formatdb.pl.
- 4) After execution of programme we got blast output files for the listed microbes, with the help of these output files, we generate query files and result files for each microbe.
- 5) In query file, we parsing the query for each blast hit and its corresponding significant similar protein match, if significant match didn't found it parsing the only query pattern. We decided the **significant value** ($E \leq 10^{-10}$) for parsing in hd_parser.pl.
- 6) After parsing the value in query files we generate result files which containing 1bit value corresponding to each protein accession number: value 1 signifies the **significant match** or **probable homolog** and 0 signifies the **mismatch** or **insignificant similarity**.
- 7) After generating query and result file made a proteome database in **MySQL server** which containing one table **phylobact** consist of **5568 rows** and **51 columns**: 49 columns specific for each microbe, 1 for id (primary key), and accession number of proteins corresponding to each gene id.
- 8) After creation of database we can identify the phylogenetic profile of each protein, this profile is a string of 49 entries, each one bit. We indicate the presence of protein in all microbes by value 1 or absent by value 0. In database we generate the phylogenetic profile of 5567 proteins, which represent the string of 49 entries.

MySQL Control Center 0.9.4-beta - [[root@localhost:3306_1] Query Window]

Console Options HotKeys Window Help

File Edit View Query Options HotKeys

id	gene	abau	atum	aaeo	aful	bant	bcer	bhal
1	AAG03391.1	1	1	1	0	1	1	1
2	AAG03392.1	1	1	1	0	1	1	1
3	AAG03393.1	1	1	0	0	1	1	1
4	AAG03394.1	1	1	1	1	1	1	1
5	AAG03395.1	0	0	1	0	1	1	1
6	AAG03396.1	0	0	0	0	0	0	1
7	AAG03397.1	0	0	0	0	0	0	0
8	AAG03398.1	1	1	1	0	0	0	0
9	AAG03399.1	1	1	1	0	0	0	0
10	AAG03400.1	1	1	0	0	0	0	0
11	AAG03401.1	1	1	0	0	0	0	0
12	AAG03402.1	0	0	0	0	0	0	0
13	AAG03403.1	0	0	0	0	0	0	0
14	AAG03404.1	0	0	0	0	0	0	0
15	AAG03405.1	0	0	0	0	0	0	0
16	AAG03406.1	0	1	0	1	0	0	0
17	AAG03407.1	1	1	0	1	1	1	1
18	AAG03408.1	1	1	1	0	1	1	1
19	AAG03409.1	1	1	1	0	1	1	1
20	AAG03410.1	1	0	0	0	0	0	0
21	AAG03411.1	1	1	1	0	0	0	1
22	AAG03412.1	1	0	0	1	1	1	1
23	AAG03413.1	1	1	1	0	1	1	1

Result 1

5568 rows in set (0.20) sec

Messages History Explain

Executing Query Read Only

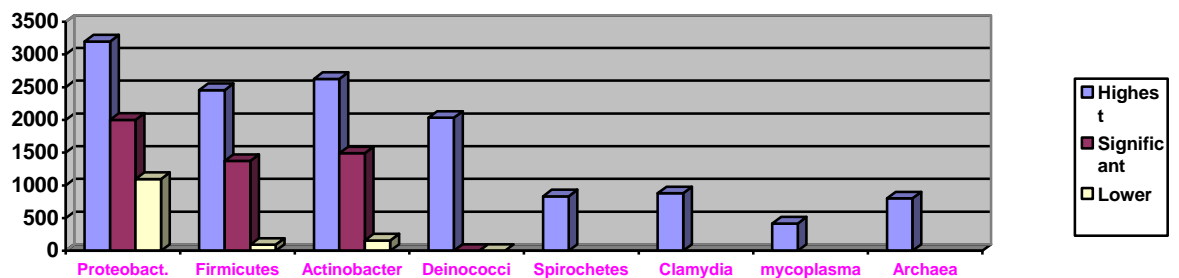
start 3 Windows Ex... final - Microsoft ... report - Microso... 2 Adobe Read... 2 Internet Expl... MySQL Control ... 6:16 PM

Snapshot of proteome database in mysql

8- Result and Discussion:

Before describing result in detail, we are giving some statistics in form of graph, which are quite useful to get general idea about result-

- ✓ In graph 8.1, we are giving information about **significant proteome set** among microbes present at different taxonomical hierarchy with respect to reference organism *P.aeruginosa*.



Graph-8.1

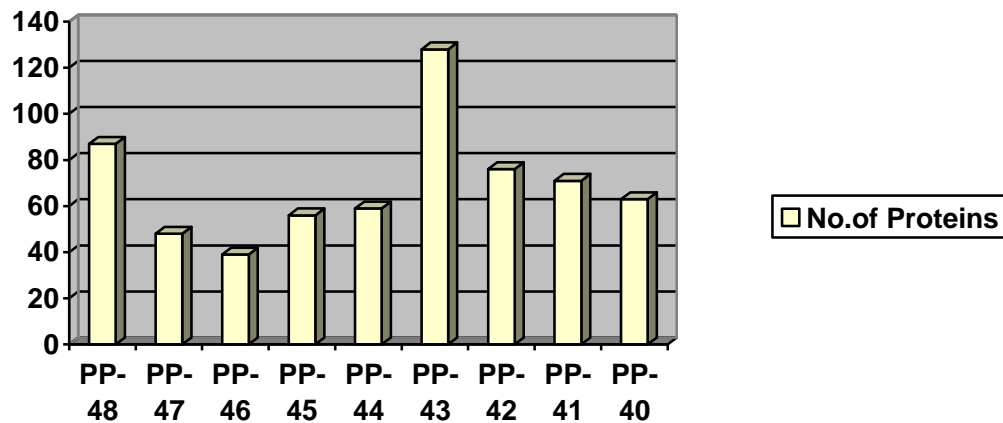
- ✓ In graph 8.1 there is a consistent pattern among first 4 classes of bacteria: **proteobacteria, firmicutes, actinobacter, deinococci**. Maximum number of hits observed in class proteobacteria which same for the reference organism *P.aeruginosa*.

Table 8.1 gives detail description about proteome set of microbes present at different taxonomical hierarchy-

No. of Proteins	Classes →	Proteobacteria		Firmicutes		Actinobacter		Deinococci	
Maximum		<i>E.colik12</i>	3199	<i>B. anthracis</i>	2456	<i>M. smegmatis</i>	2627	<i>D. radiodurans</i>	2034
Intermediate		<i>H.influenjae</i>	2002	<i>S. pneumoniae</i>	1373	<i>M.leprae</i>	1491		
Minimum		<i>R.prowazekii</i>	1099	<i>C. perfringes</i>	94	<i>S. coelicolor</i>	160		

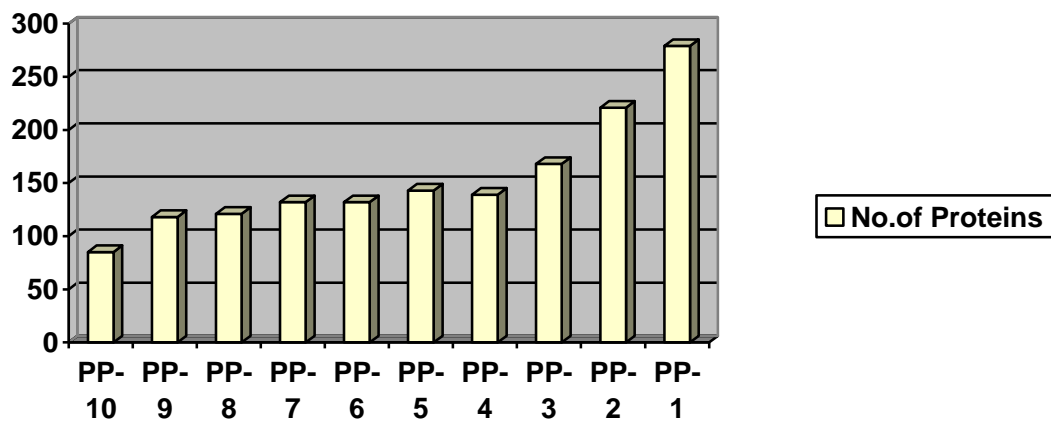
Table 8.1

- ✓ In graph 8.2 on **x-axis** type of phylogenetic profiling is represented and on **y-axis** no. of proteins are displayed. Numbers of proteins, which are conserved or functionally related, are maximum in PPT-43 and minimum in PPT-46.



Graph-8.2

- ✓ In graph 7.3-axis pattern are same. Numbers of proteins, which are conserved or functionally related, are maximum in PPT-1, that is present only in one microbe; these microorganisms mainly belongs to order **gamma proteobacteria** and closely related to *P.aeruginosa*.



Graph 8.3

- ✓ **Distribution pattern** for number of proteins is quite different in these two graphs which indirectly signnifighes the evolutionary pattern of conserved proteins between different taxonomical hierarchy.

To test whether similar phylogenetic profile proteins are functionally related we selected three type of phylogenetic profile –

In type I we choose those proteins, which are present in 48 out of 49 organism except *Streptomyces coelicolor*, in type II, and III we choose those proteins, which are, differ from type I by **one bit** and **two bit** respectively.

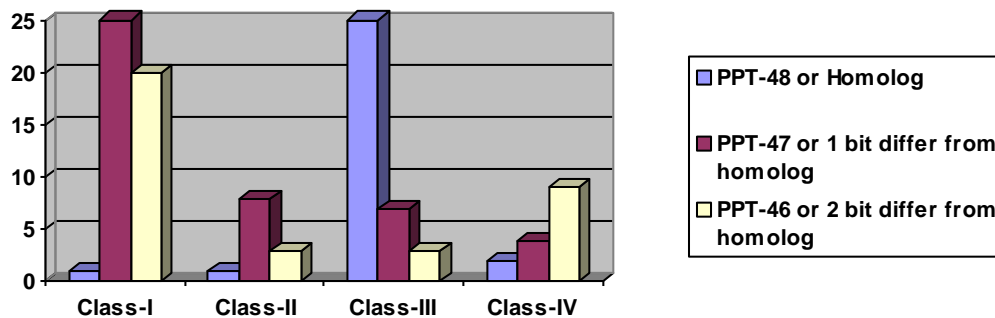
Type II proteins absent in *S.coelicolor* and *Clostridium perfringes* but type III proteins absent in *S.coelicolor*, *C. perfringes*, and *Methanococcus jannaschii*.

We divided these proteins into following four classes --

- Class I- Proteins related to DNA replication and protein synthesis.
- Class II- Proteins related to cell division.
- Class III-Proteins related to cell transport-membrane proteins.
- Class IV-proteins related to metabolic pathways.

We assumed that protein involved in these classes are functionally linked and evolved in co-related fashion during evolutionary process.

- ✓ In figure 8.1 we represent **functionally related proteins** of similar phylogenetic profile in **four boxes** which signifies the **classes**, proteins which are enclosed in double line boxes supposed to be homologous and linked by single line boxes by double arrow, these single line boxes are present in two column which are differ by one bit and two bits from double line box proteins respectively. Boxes of similar phylogenetic profile proteins connected by single line.
- ✓ In graph 8.4, we give the statistics for distribution of homolog, 1-bit and 2-bit differ proteins respectively which are arranged in four classes ---



Graph-8.4

- ✓ According to graph 8.4 class III proteins **significantly conserve** in microbes during evolutionary process after that class-I proteins show good conservancy among one bit and 2 bit differ protein set of microbes, class IV proteins exponentially increases from homolog to 2 bit differ protein set of microbes.
- ✓ Class-II proteins follow same pattern as class-I because these two classes are **closely related** to each other than other classes.

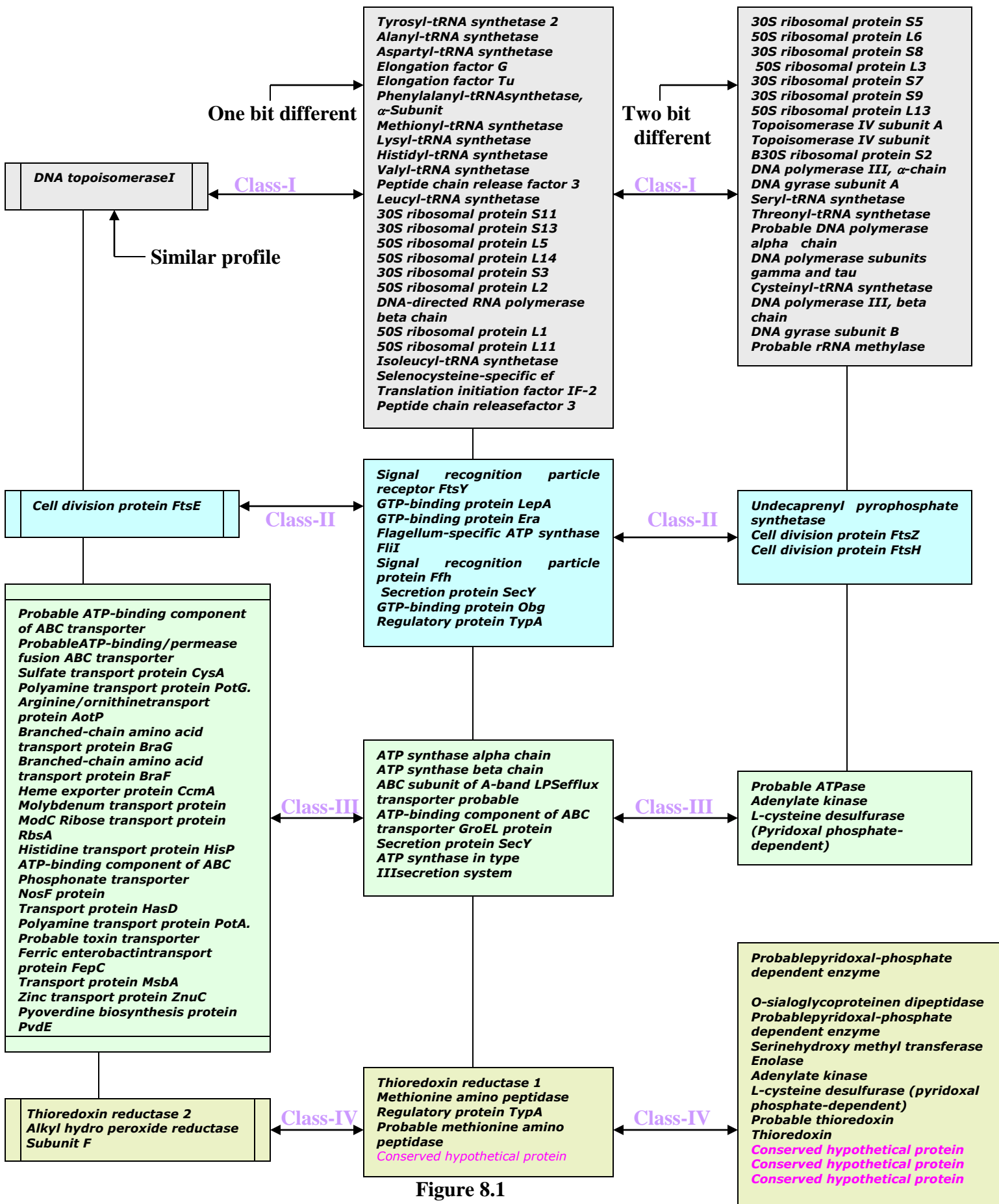


Figure 8.1

8.2 Discussion:

Only those proteins conserved during evolutionary process which are involved in **related cellular process** like **cell division** which is coordinated with other cell cycle events such as **genomic DNA synthesis** that leads to **chromosomal replication** and partition, increase of cell mass, and cell expansion by cell wall synthesis.

- ✓ In figure 8.1 maximum numbers of proteins, which are conserved or evolved in correlated fashion during evolutionary process are belongs to these cellular processes.

We can prove this point with the help of following evidences-

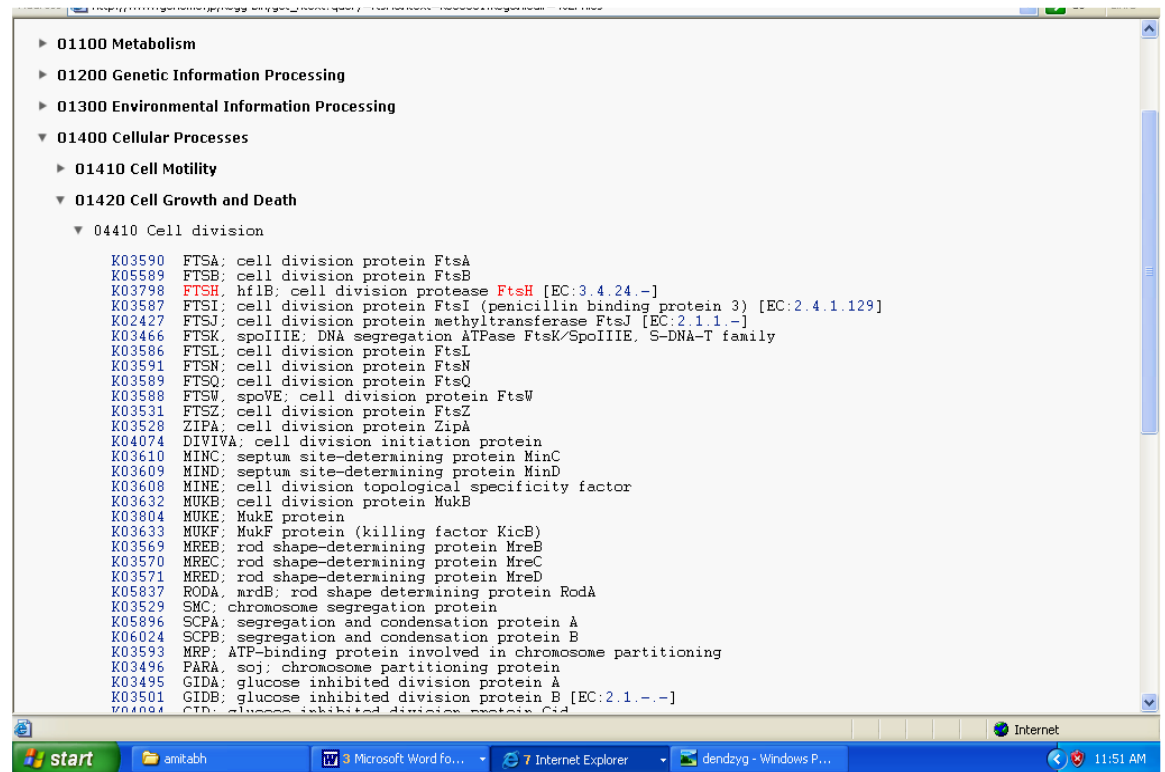
Table-8.2

Pathway		Proteins involved	Cellular process
Major	Genetic Information Processing: Replication & Repair		
Replication complex		DNA topoisomerase II, IV, DNA gyraseA, B, Helicases, Recombination protein-RecA, RuvA, RuvB.	DNA synthesis
DNA polymerases		DNA polymerase α & β chain, DNA polymerase subunit γ and ϕ .	DNA synthesis
Major	Genetic Information Processing: Translation		
Amino acyl -tRNA biosynthesis		Tyrosyl, Alanyl, Aspartyl, Leucyl, Isoleucyl, scryl... t-RNA synthase.	Protein synthesis
Ribosome		Ribosomal proteins	Protein synthesis and antibiotic resistance.
Translation related protein		Elongation factor Tu, elongation factor G, protein release factor -3, ftsY.	Protein synthesis
Major	Environmental information processing: Membrane transport		
ABC transporters		FtsZ, ftsE, ftsH, LepA, Era, Obg, TypA, cysA, potG, AotP, braG, braF, ccmA, modC, rbsA, HisP.	Cell division, antibiotic resistance and nutrients transport.
Type III secretion system		SecY, ATP synthase, ATP ase, Adenylate kinase	Cellular metabolism
Major	Metabolism:		
Carbohydrate		Enolase	Energy generation
Nucleotide/ Antioxidant		Thioredoxin reductase I, methionine amino peptidase, kinases.	Cellular integrity & Energy generation

In table 8.2[4], we categorized proteins according to their concerned pathway and cellular processes; all observed proteins present in those cellular processes, which are linked to each other. For example- **DNA synthesis, protein synthesis, cellular energetics** is linked to **cell division**.

Here we gave focus upon the functionally related **fts proteins**, which shows significant co-related evolution during evolutionary process-

- ✓ The ratios of **FtsA** to **FtsZ** in the cytoplasm seem to be important to initiate cell division. **FtsW** is an **integral membrane protein** that plays an important role in Z-ring stability and **FtsI** (penicillin-binding protein 3) is a **transpeptidase** required for the final step of peptidoglycan synthesis in the septum.

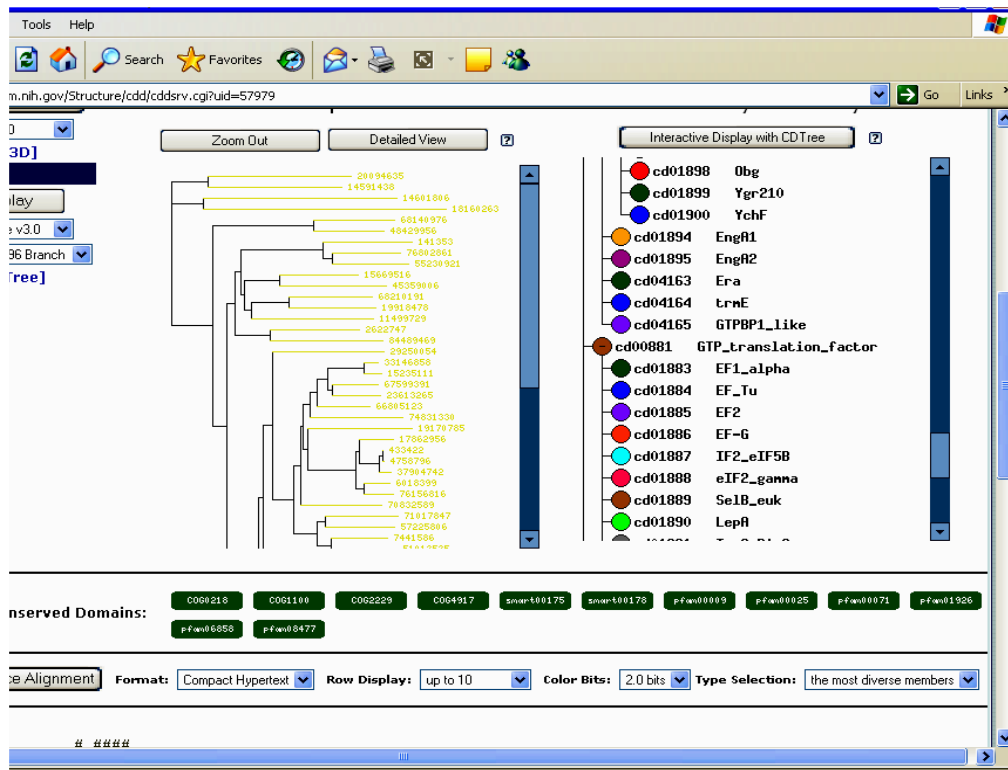


Snapshot of KEGG orthology for cell division proteins [4]

- ✓ **FtsH** is an **ATP-dependent zinc protease** with a membrane-spanning segment in the N terminus. This protein is known to degrade SecY, the F₀ subunit of ATPase, heat shock transcription factor sigma 32, and EnvA (LpxC), a key enzyme for **lipid A biosynthesis** [3].
- ✓ **FtsY** belongs to the family of "**signal recognition particle (SRP)-type GTPases**." SRP is a ribonucleoprotein particle that binds to short nascent polypeptides, exposing a hydrophobic targeting signal just outside the ribosome that is released by the action of the receptor FtsY.

- ✓ **FtsE** and **FtsX** form a complex in the inner membrane that bears the characteristics of an **ATP-binding cassette (ABC)-type transporter**. A mutant lacking functional FtsE has filamentous growth and is only viable on high-salt medium, indicating a role for FtsE in cell division and/or salt transport.

After analyzing the result we can say that the GTPase super family [7], present in all domains of life, is related to many functions such as protein synthesis, cell cycle and differentiation.



Snapshot for phylogenetic tree of GTPase superfamily proteins.

- ✓ **Era**, **Obg** and **LepA** is a membrane-associated GTP-binding protein which is essential for cell growth in bacteria [20].
- ✓ In order to examine the physiological role of Era, strains in which Era was expressed at 40 degrees C but completely repressed at 27 degrees C were constructed. The growth of these strains was inhibited at the non-permissive temperature, and cells became elongated. Under such conditions, no constrictions or septum formation could be detected by phase-contrast microscopy, and DNA segregation was apparently normal as revealed by fluorescence staining.

These data demonstrate that Era has an essential function in cell growth rate control in liquid media and that depletion of Era blocks cell division either directly or indirectly.

✓ It's proved by the experiment that the expression of a “**probable ATP-binding component of ATP binding cassette (ABC) transporter**” existed in ciprofloxacin-intermediate and -resistant strains, but not in sensitive strain. [19] The changes of mRNA levels of the probable ATP-binding component of ABC transporter were detected by virtue of RT-PCR and showed that this protein did not express in the **sensitive strains** but expressed increasingly in the **intermediate** and **resistant** strains.

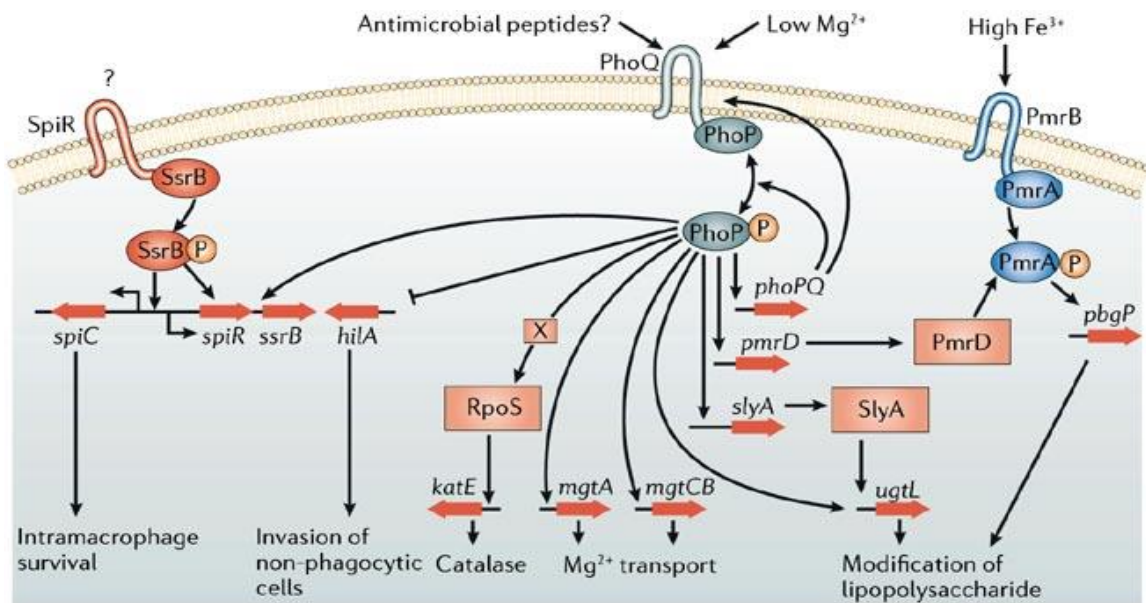


After analyzing phylogenetic profile, which are differ from homolog by **4 to 9 bit**, **two-component regulatory signal transduction systems** are widely distributed among these microbes and it's responsible for enable the organisms to make coordinated changes in gene expression in response to a variety of environmental stimuli.

Mainly following proteins show significant evolutionary relatedness in these microbes-

- **Heat shock proteins** –Rho proteins, Proteins of OmpR family-KdpE, PhoQ, PhoP and PQ .
- **Molecular chaperons** - ClpB/DnaK/DnaJ/GrpE.

- ✓ Heat shock proteins help to adopt microorganism during stress response or diverse environmental conditions.



Copyright © 2006 Nature Publishing Group
Nature Reviews | Microbiology

Figure 8.2

- ✓ In figure 8.2 [8], we can see low Mg⁺⁺ concentration activate terminal domain attached protein PhoP of trans membrane protein PhoQ, which further activate cascade of transcription factors which activate expression of genes to cope up the effects of low Mg⁺⁺ concentration.
- ✓ Molecular chaperons mainly ClpB/DnaK/DnaJ/GrpE system is involved in **protein monomerization** and in the activation of a **DNA replication factor I** [21].

8.3-Final Extract:

- ❖ After analyzing result and discussion, **probable ATP binding component protein** show highest conservance among 49 proteome sets of microbes and it's proved by expeiment this is responsible for **multidrug resistance**. So, it can be act as a novel drug target for *P.aeruginosa* and closely linked microbes which show multi drug resistance.
- ❖ In figure 8.1 ,we find 3 conserved hypothetical protein ,first in one bit different profile proteins and 2nd and 3rd in two bit diffeent profile proteins. According to our discussion we can say that first protein specifically involve in **cell devision** and 2nd and 3rd involved in **genomic information processing**.

9- Inference:

Following inferences draw from the result and discussion-

- Analyzing the result we can say that phylogenetic profiling is the best approach to characterize the functionally related protein, which are evolved in co-related fashion during evolutionary process.
- With the help of phylogenetic profile of homolog, 1 bit and 2 bit different profile proteins from homolog; we characterize the function of three-conserved hypothetical protein.
- There are very large number of proteins 939 which are specific for *P.aeruginosa*, show no significant match with any proteins of 49-proteome sets and 280 proteins show significant match with only one protein of 49- proteome sets. This signifies that during evolutionary process *P.aeruginosa* change their genome for key features like **environmental versality, multi- drug resistance and nutrient independency**.
- After analyzing significant match among microbes with the help of database created, *C.perfringes*, *S.coelicolor*, *M.genitalium*, *M.pneumoniae* show less than 500 significant matches, which indicate significant **proteome reduction** during evolutionary process may be due to change in ecological niche or adaptation in host cells.
- Analyzing the discussion part, maximum numbers of conserved proteins are belonging to **ABC transporters** family. The design and adaptability of ABC transporters have apparently served life well throughout evolution. Not surprisingly, loss of ABC transporter function is implicated in many diseases, so they are therefore important targets for therapeutics.
- So, we can conclude that functionally related proteins, which are involved in cell division, DNA synthesis, protein synthesis, and carbohydrate metabolism evolved in co-related fashion during **evolutionary process**. These findings indicate that comparing profiles is a useful tool for identifying the complex or pathway in which a protein participates. Finally, we were able to make **functional assignments** of **uncharacterized proteins** by examining the function of proteins with identical phylogenetic profiles.

1 0. Future work:

- By analyzing phylogenetic profile of different taxonomical hierarchy, one can study the process of proteome reduction in respect to protein family evolution and role of pseudo genes in pathogenic microbes [16].
- Many viral genomes are completed till date, so one can apply the same concept of phylogenetic profiling for viruses and identified the function of uncharacterized protein.

11. References:

Books-

1-Molecular Evolution: Wen- Hsiung Li

Chapter 14: Role of mutation and selection in molecular evolution [419-424].

2-Bioinformatics and Molecular Evolution: Paul G. Higgs and Teresa K. Attwood

Chapter 8: Phylogenetic Methods [158-162]

3-Helicobacter pylori -physiology and genetics: Harry L. T. Mobley, Mendz, and Hazell

Chapter 23: Chromosomal replication, Plasmid replication, and Cell division.

Websites-

4-<http://www.genome.jp/kegg/brite.html>

5-<http://ca.expasy.org/sprot/hamap/ACIAC.html>

6-<http://biowulf.nih.gov/apps/blast/doc/formatdb.html>

7-<http://www.ncbi.nlm.nih.gov/Structure/cdd/>

8-<http://www.nature.com/nrmicro/journal/v4/n9/images/nrmicro1478-f1.jpg>

9-http://www.ncbi.nlm.nih.gov/projects/Gene/gentrez_stats.cgi

10-<http://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/>

11-<http://www.pseudomonas.com/>

12-<http://scop.mrc-lmb.cam.ac.uk/scop/statstics>

Research papers-

13- Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D. & Yeates, T. O. assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl Acad. Sci. USA* **96**, 4285–4288 (1999).

14-Protein function in the post-genomic era: David Eisenberg, Edward M. Marcotte, and Ioannis Xenarios & Todd O. Yeates. *Nature*, 823-826(2000).

15-Studying Genomes Through the Aeons: Protein families, Pseudogenes and Proteome Evolution -Paul M. Harrison and Mark Gerstein*. *J. Mol. Biol.* (0000) 00, 1–20(in press).

16-The Real life of pseudogenes: Mark Gerstein and Deyou Zheng. *Scientific American*, 49-55 (2006).

- 17- SCOP: A Structural Classification of Proteins Database for the Investigation of Sequences and Structures: Alexey G. Murzin, Steven E. Brenner, Tim Hubbard and Cyrus Chothia*. *J. Mol. Biol.* (1995) **247**, 536–540.
- 18- Predicting functional divergence in protein evolution by site-specific rate shifts: Eric A. Gaucher, Xun Gu, Michael M. Miyamoto and Steven A. Benner. *TRENDS in Biochemical Sciences* Vol.27 No.6 June 2002.
- 19- An important role of a probable ATP-binding component of ABC transporter during the process of *Pseudomonas aeruginosa* resistance to fluoroquinolone: Jinsong Zhou, Dr. *, Dongyun Hao 2, Xudong Wang 1, Teimei Liu 1, Chengyan He 1, Feng Xie 1, Yanhong Sun 2, Jin Zhang 2. *Willey Intersciences* (2005), Abstract.
- 20- A GTP-binding protein (Era) has an essential role in growth rate and cell cycle control in *Escherichia coli*: N Gollop and P E March. *J Bacteriol.* 1991 April, 173(7): 2265–2270, Abstract.
- 21- Cooperative action of *Escherichia coli* ClpB protein and DnaK chaperone in the activation of a replication initiation protein: Igor Konieczny and Krzysztof Liberek. *J. Biol. Chem*, 10.1074. (2002), Abstract.

12. Appendix:

- 1) 1-All the information about homolog and 1-9 bit different proteins from homolog are given in the form of tables 12.1, 12.2, and 12.5.
- 2) 2- Information regarding proteins, which are present in only one microbe, and specific proteins for *P.aeruginosa* given in tables 12.3 and 12.4 respectively.
- 3) 3- List of microbes is used for phylogenetic profile database creation given on page no. 82.
- 4) 4- Perl scripts used in database creation given on page no. 84.

Tables:

Following tables constructed with the help of information extracted from proteome database:

- ✓ **Table 12.1 and Table 12.2** – These tables containing different classes of protein that are differ by one bit. Table 13.1 proteins present in 48 microbes out of 49 except *Streptomyces coelicolor*, table B proteins present in 47 microbes out of 49 except *S.coelicolor* and *Clostridium perfringes*.
- ✓ **Table 12.3 and 12.4** – These tables containing proteins that are different by one bit, table 13.3 proteins present in only one microbe out of 49 and table 13.4 proteins absent in all the 49 microbes, these proteins are specific for reference microbe *Pseudomonas aeruginosa*. [Page no. 3,10]
- ✓ **Table 12.5** – It consist annotated information about similar profile proteins (PPT-48) and proteins, which are, differ by 1 to 9 bit from these proteins. [Page no. 34]

Table 12.1:

CLASS I: MEMBRANE PROTEINS INVOLVED IN SECRETION AND TRANSPORT
Probable ATP-binding component of ABC transporter
Sulfate transport proteinCysA
Polyamine transport proteinPotG
ProbableATP-binding/permease fusion ABC transporter
Arginine/ornithinetransport protein AotP
Branched-chain amino acidtransport protein BraG
Branched-chain amino acidtransport protein BraF
Alkaline protease secretionprotein AprD
Heme exporter protein CcmA
Molybdenum transportprotein ModC
Ribose transport proteinRbsA
Histidine transportprotein HisP
ATP-binding component of ABC phosphonate transporter
NosF protein-abc trabsport related
Transport protein HasD
Polyamine transport proteinPotA0
Probable toxin transporter
Ferric enterobactintransport protein FepC
Transport protein MsbA
ATP-binding component ofABC phosphate transporter
Zinc transport protein ZnuC

CLASS II: PROTEINS INVOLVED IN CELL CYCLE AND DNA REPLICATION
Cell division protein FtsE
DNA topoisomerase I

CLASS III: PROTEINS INVOLVED IN METABOLIC PATHWAYS
Thioredoxin reductase 2
Pyoverdine biosynthesis protein PvdE
Alkyl hydroperoxidoreductase subunit F

Table 12.2:

CLASS I: MEMBRANE PROTEINS INVOLVED IN SECRETION AND TRANSPORT
SIGNAL RECOGNITION PARTICLE RECEPTOR FTSY
GTP-binding protein LepA
GTP-binding protein Era
Probable ATP-binding component of ABC transporter
ATP synthase in type III secretion system
Signal recognition particle protein Ffh
Secretion protein SecY
GTP-binding protein Obg
Probable ATP-binding component of ABC transporter
ABC subunit of A-band LPS efflux transporter

CLASS II: PROTEINS INVOLVED IN METABOLIC PATHWAYS (MAINLY IN T-RNA SYNTHESIS)
rRNA (adenine-N6, N6)-dimethyltransferase
Tyrosyl-tRNA synthetase 2
Alanyl-tRNA synthetase
Aspartyl-tRNA synthetase
Flagellum-specific ATP synthase FliI
Thioredoxin reductase 1
Phenylalanyl-tRNA synthetase, alpha-subunit

Probable methionineaminopeptidase
Methionyl-tRNA synthetase
Methionine amino peptidase
Lysyl-tRNA synthetase
Histidyl-tRNA synthetase
Valyl-tRNA synthetase
Leucyl-tRNA synthetase
ATP sulfurylase GTP-binding subunit/APS kinase
Isoleucyl-tRNA synthetase
ATP synthase beta chain

CLASSIII: PROTEINS INVOLVED IN DNA METABOLISM (REPLICATION, TRANSLATION, TRANSCRIPTION)
DNA-directed RNA polymerase beta* chain
DNA-directed RNA polymerase beta chain
Regulatory protein TypA
Elongation factor G
Peptide chain release factor 3
Elongation factor Tu
Elongation factor G
Translation initiation factor IF-2
Selenocysteine-specificelongation factor

CLASS IV: STRUCTURAL PROTEINNS (mainly involved in ribosome synthesis)
30S ribosomal protein S11
30S ribosomal protein S13
50S ribosomal protein L5
50S ribosomal protein L14
30S ribosomal protein S3
50S ribosomal protein L2
50S ribosomal protein L1
50S ribosomal protein L11

CLASS V: OTHER CONSERVED PROTEIN
GroEL protein--chaperon
Conserved hypothetical protein

Table 12.3:

	Probable proteins
	Conserved protein
	Functionally annotated protein

Gene ID	Accession No.	Protein Name
---------	---------------	--------------

15	AAG03405.1	Hypothetical protein PA0015
49	AAG03439.1	Hypothetical protein PA0049
55	AAG03445.1	Hypothetical protein PA0055
92	AAG03482.1	Hypothetical protein PA0092
216	AAG03604.1	Probable transporter
217	AAG03605.1	Probable transporter
344	AAG03732.1	Hypothetical protein PA0343
384	AAG03772.1	Conserved hypothetical protein
446	AAG03834.1	Probable transposase
458	AAG03846.1	Hypothetical protein PA0457
463	AAG03851.1	Hypothetical protein PA0462
491	AAG03879.1	Hypothetical protein PA0490
526	AAG03914.1	Probable dinitrificationprotein NorD
545	AAG03933.1	Hypothetical proteinPA0544
558	AAG03946.1	Hypothetical protein PA0557
571	AAG03959.1	Hypothetical protein PA0570
573	AAG03961.1	Hypothetical protein PA0572
616	AAG04004.1	Hypothetical proteinPA0615
671	AAG04059.1	Hypothetical protein PA0670
690	AAG04078.1	Hypothetical protein PA0689
693	AAG04081.1	Hypothetical protein PA0692
702	AAG04090.1	Probable transcriptional regulator
727	AAG04115.1	Hypothetical protein of bacteriophage Pf1
778	AAG04166.1	Hypothetical protein PA0777
789	AAG04177.1	Hypothetical protein PA0788
801	AAG04189.1	Hypothetical protein PA0800
858	AAG04245.1	Hypothetical protein PA0856
864	AAG04251.1	Hypothetical protein PA0862
869	AAG04256.1	Hypothetical protein PA0867
880	AAG04267.1	Hypothetical protein PA0878
893	AAG04280.1	Hypothetical protein PA0891
910	AAG04297.1	Hypothetical protein PA0908
911	AAG04298.1	Hypothetical protein PA0909
927	AAG04314.1	Hypothetical protein PA0925
993	AAG04380.1	Hypothetical protein PA0991

1031	AAG04418.1	Hypothetical protein PA1029
1032	AAG04419.1	Hypothetical protein PA1030
1040	AAG04427.1	Hypothetical protein PA1038
1046	AAG04433.1	Hypothetical protein PA1044
1048	AAG04435.1	Hypothetical protein PA1046
1066	AAG04453.1	Hypothetical protein PA1064
1071	AAG04458.1	Hypothetical protein PA1069
1095	AAG04482.1	Hypothetical protein PA1093
1108	AAG04495.1	Hypothetical protein PA1106
1138	AAG04525.1	Probable transcriptional regulator
1193	AAG04580.1	Hypothetical protein PA1191
1215	AAG04602.1	Hypothetical protein PA1213
1220	AAG04607.1	Hypothetical protein PA1218
1228	AAG04615.1	Probable transcriptional regulator
1241	AAG04628.1	Hypothetical protein PA1239
1243	AAG04630.1	Probable transcriptional regulator
1316	AAG04703.1	Hypothetical protein PA1314
1325	AAG04712.1	Hypothetical protein PA1323
1327	AAG04714.1	Conserved hypothetical protein
1403	AAG04790.1	Hypothetical protein PA1401
1405	AAG04792.1	Probable transcriptional regulator
1408	AAG04795.1	Hypothetical protein PA1406
1436	AAG04823.1	Hypothetical protein PA1434
1467	AAG04854.1	Hypothetical protein PA1465
1486	AAG04873.1	Probable transcriptional regulator
1501	AAG04888.1	Conserved hypothetical protein
1508	AAG04895.1	Hypothetical protein PA1506
1515	AAG04902.1	Hypothetical protein PA1513
1561	AAG04948.1	Hypothetical protein PA1559
1562	AAG04949.1	Hypothetical protein PA1560
1574	AAG04961.1	Conserved hypothetical protein
1595	AAG04982.1	Hypothetical protein PA1593
1596	AAG04983.1	Hypothetical protein PA1594
1606	AAG04993.1	Hypothetical protein PA1604
1623	AAG05010.1	Probable hydrolase
1658	AAG05045.1	Hypothetical protein PA1656
1661	AAG05048.1	Hypothetical protein PA1659
1667	AAG05054.1	Hypothetical protein PA1665
1674	AAG05061.1	Hypothetical protein PA1672
1735	AAG05122.1	Conserved hypothetical protein
1742	AAG05129.1	Hypothetical protein PA1740
1743	AAG05130.1	Hypothetical protein PA1741
1765	AAG05152.1	Hypothetical protein PA1763
1768	AAG05155.1	Hypothetical protein PA1766

1769	AAG05156.1	Hypothetical protein PA1767
1770	AAG05157.1	Hypothetical protein PA1768
1827	AAG05214.1	Hypothetical protein PA1825
1843	AAG05230.1	Hypothetical protein PA1841
1874	AAG05261.1	Hypothetical protein PA1872
1923	AAG05309.1	Hypothetical protein PA1921
1926	AAG05312.1	Hypothetical protein PA1924
1927	AAG05313.1	Hypothetical protein PA1925
1941	AAG05327.1	Hypothetical protein PA1939
1942	AAG05328.1	Hypothetical protein PA1940
1953	AAG05339.1	Hypothetical protein PA1951
2040	AAG05426.1	Hypothetical protein PA2038
2046	AAG05432.1	Hypothetical protein PA2044
2048	AAG05434.1	Hypothetical protein PA2046
2051	AAG05437.1	Hypothetical protein PA2049
2068	AAG05454.1	Hypothetical protein PA2066
2163	AAG05549.1	Hypothetical protein PA2161
2182	AAG05568.1	Hypothetical protein PA2180
2189	AAG05575.1	Hypothetical protein PA2187
2235	AAG05621.1	Probable glycosyl transferase
2241	AAG05627.1	Probable transferase
2244	AAG05630.1	Hypothetical protein PA2242
2266	AAG05652.1	Conserved hypothetical protein
2273	AAG05659.1	Probable acetyltransferase
2283	AAG05669.1	Probable transcriptional regulator
2284	AAG05670.1	Hypothetical protein PA2282
2320	AAG05706.1	Hypothetical protein PA2318
2321	AAG05707.1	Probable transposase
2338	AAG05724.1	Hypothetical protein PA2336
2376	AAG05762.1	Hypothetical protein PA2374
2379	AAG05765.1	Hypothetical protein PA2377
2454	AAG05841.1	Hypothetical protein PA2453
2456	AAG05843.1	Hypothetical protein PA2455
2503	AAG05890.1	Hypothetical protein PA2502
2539	AAG05926.1	Hypothetical protein PA2538
2545	AAG05932.1	Hypothetical protein PA2544
2592	AAG05979.1	Probable transcriptional regulator
2637	AAG06024.1	Hypothetical protein PA2636
2659	AAG06046.1	Hypothetical protein PA2658
2691	AAG06078.1	Probable transposase
2695	AAG06082.1	Probable thioredoxin
2698	AAG06085.1	Hypothetical protein PA2697
2757	AAG06144.1	Hypothetical protein PA2756
2790	AAG06177.1	Hypothetical protein PA2789

2798	AAG06185.1	Hypothetical protein PA2797
2808	AAG06195.1	Hypothetical protein PA2807
2834	AAG06221.1	Conserved hypothetical protein
2870	AAG06257.1	Hypothetical protein PA2869
2872	AAG06259.1	Hypothetical protein PA2871
2947	AAG06334.1	Hypothetical protein PA2946
3022	AAG06409.1	Hypothetical protein PA3021
3032	AAG06419.1	Hypothetical protein PA3031
3065	AAG06452.1	Hypothetical protein PA3064
3070	AAG06457.1	Hypothetical protein PA3069
3081	AAG06468.1	Hypothetical protein PA3080
3082	AAG06469.1	Conserved hypothetical protein
3094	AAG06481.1	Hypothetical protein PA3093
3131	AAG06518.1	Hypothetical protein PA3130
3145	AAG06532.1	Hypothetical protein PA3144
3181	AAG06568.1	Hypothetical protein PA3180
3232	AAG06619.1	Hypothetical protein PA3231
3252	AAG06639.1	Hypothetical protein PA3251
3282	AAG06669.1	Hypothetical protein PA3281
3289	AAG06676.1	Hypothetical protein PA3288
3290	AAG06677.1	hypothetical protein PA3289
3318	AAG06705.1	hypothetical protein PA3317
3352	AAG06739.1	hypothetical protein PA3351
3360	AAG06747.1	hypothetical protein PA3359
3369	AAG06756.1	Probable acetyltransferase
3414	AAG06801.1	Conserved hypothetical protein
3422	AAG06809.1	Conserved hypothetical protein
3423	AAG06810.1	Hypothetical protein PA3422
3435	AAG06822.1	Probable transposase
3484	AAG06871.1	Hypothetical protein PA3483
3511	AAG06898.1	Hypothetical protein PA3510
3600	AAG06987.1	Probable transcriptional regulator
3715	AAG07102.1	Hypothetical protein PA3715
3716	AAG07103.1	hypothetical protein PA3716
3756	AAG07143.1	Hypothetical protein PA3756
3848	AAG07235.1	Hypothetical protein PA3848
3885	AAG07272.1	hypothetical protein PA3885
3902	AAG07289.1	Hypothetical protein PA3902
3922	AAG07309.1	Conserved hypothetical protein
3923	AAG07310.1	Hypothetical protein PA3923
3943	AAG07330.1	Conserved hypothetical protein
3955	AAG07342.1	Hypothetical protein PA3955
3966	AAG07353.1	Hypothetical protein PA3966
3988	AAG07375.1	Hypothetical protein PA3988

3993	AAG07380.1	Probable transposase
4008	AAG07395.1	Probable hydrolase
4009	AAG07396.1	Hypothetical protein PA4009
4014	AAG07401.1	Hypothetical protein PA4014
4035	AAG07422.1	Hypothetical protein PA4035
4049	AAG07436.1	Hypothetical protein PA4049
4060	AAG07447.1	Hypothetical protein PA4060
4063	AAG07450.1	Hypothetical proteinPA4063
4074	AAG07461.1	Probable transcriptional regulator
4086	AAG07473.1	Probable fimbrial subunitCupB1
4095	AAG07482.1	Hypothetical proteinPA4095
4105	AAG07492.1	hypothetical protein PA4105
4107	AAG07494.1	hypothetical protein PA4107
4118	AAG07505.1	hypothetical protein PA4118
4164	AAG07551.1	hypothetical protein PA4164
4299	AAG07687.1	hypothetical protein PA4299
4300	AAG07688.1	hypothetical protein PA4300
4303	AAG07691.1	hypothetical protein PA4303
4305	AAG07693.1	hypothetical proteinPA4305
4317	AAG07705.1	hypothetical protein PA4317
4337	AAG07725.1	hypothetical protein PA4337
4340	AAG07728.1	hypothetical protein PA4340
4345	AAG07733.1	hypothetical protein PA4345
4349	AAG07737.1	hypothetical protein PA4349
4357	AAG07745.1	Conserved hypothetical protein
4364	AAG07752.1	Hypothetical protein PA4364
4384	AAG07772.1	Hypothetical protein PA4384
4390	AAG07778.1	Hypothetical proteinPA4390
4391	AAG07779.1	Hypothetical protein PA4391
4459	AAG07847.1	conserved hypothetical protein
4518	AAG07906.1	Hypothetical protein PA4518
4533	AAG07921.1	hypothetical protein PA4533
4534	AAG07922.1	Hypothetical protein PA4534
4577	AAG07965.1	Hypothetical protein PA4577
4579	AAG07967.1	hypothetical protein PA4579
4636	AAG08024.1	hypothetical protein PA4636
4643	AAG08031.1	hypothetical protein PA4644
4656	AAG08044.1	hypothetical proteinPA4657
4685	AAG08073.1	hypothetical protein PA4686
4701	AAG08089.1	hypothetical protein PA4703
4702	AAG08090.1	hypothetical protein PA4704
4703	AAG08091.1	hypothetical protein PA4705
4716	AAG08104.1	hypothetical protein PA4718
4791	AAG08179.1	hypothetical proteinPA4793

4793	AAG08181.1	Hypothetical proteinPA4795
4795	AAG08183.1	Probable transposase
4800	AAG08188.1	hypothetical protein PA4802
4804	AAG08192.1	Probable transcriptional regulator
4815	AAG08202.1	hypothetical protein PA4817
4840	AAG08227.1	hypothetical protein PA4842
4861	AAG08248.1	hypothetical protein PA4863
4875	AAG08262.1	Hypothetical protein PA4877
4919	AAG08306.1	Hypothetical protein PA4921
4956	AAG08343.1	hypothetical protein PA4958
4963	AAG08350.1	Hypothetical protein PA4965
5001	AAG08388.1	Hypothetical protein PA5003
5024	AAG08411.1	Hypothetical protein PA5026
5046	AAG08433.1	Probable nuclease
5060	AAG08447.1	Conserved hypothetical protein
5062	AAG08449.1	Hypothetical proteinPA5064
5086	AAG08473.1	Hypothetical proteinPA5088
5087	AAG08474.1	Hypothetical proteinPA5089
5111	AAG08498.1	Hypothetical protein PA5113
5149	AAG08536.1	Hypothetical protein PA5151
5194	AAG08581.1	Hypothetical protein PA5196
5200	AAG08587.1	Hypothetical protein PA5202
5210	AAG08597.1	Hypothetical protein PA5212
5224	AAG08611.1	Hypothetical protein PA5226
5249	AAG08636.1	Hypothetical protein PA5251
5268	AAG08655.1	Hypothetical protein PA5270
5269	AAG08656.1	Hypothetical protein PA5271
5282	AAG08669.1	Hypothetical protein PA5284
5305	AAG08692.1	Hypothetical protein PA5307
5399	AAG08786.1	Hypothetical protein PA5401
5410	AAG08797.1	Hypothetical proteinPA5412
5439	AAG08826.1	Hypothetical protein PA5441
5459	AAG08846.1	Hypothetical protein PA5461
5460	AAG08847.1	Hypothetical protein PA5462
5461	AAG08848.1	Hypothetical protein PA5463
5465	AAG08852.1	Hypothetical protein PA5467
5479	AAG08866.1	Hypothetical protein PA5481
5511	AAG08898.1	Hypothetical protein PA5513
5518	AAG08905.1	Hypothetical protein PA5520
5525	AAG08912.1	Hypothetical protein PA5527

Proteins with specific annotation

44	AAG03434.1	Exoenzyme T
----	------------	-------------

519	AAG03907.1	Cytochrome c-551 precursor
1002	AAG04389.1	Quinolone signal response protein
1152	AAG04539.1	Pyocin S2
1177	AAG04564.1	NapD protein of periplasmic nitrate reductase
1179	AAG04566.1	Periplasmic nitratereductase protein NapE
1434	AAG04821.1	Auto inducer synthesis protein LasI
1698	AAG05085.1	Translocation protein in type III secretion
1700	AAG05087.1	Type III secretion outer membrane protein PopN precursor
1759	AAG05146.1	Homoserine kinase
2195	AAG05581.1	Hydrogen cyanide synthaseHcnA
2256	AAG05642.1	Pyoverdine biosynthesis protein PvcA
2257	AAG05643.1	Pyoverdine biosynthesis protein PvcB
3083	AAG06470.1	Glycine betainetransmethylase
3096	AAG06483.1	General secretion pathway protein M
3151	AAG06538.1	LPS biosynthesis proteinWbpG
3408	AAG06795.1	Heme acquisition protein HasAp
3477	AAG06864.1	Auto inducer synthesis protein RhII
3542	AAG06929.1	Alginate biosynthesis protein Alg8
3544	AAG06931.1	Alginate biosynthetic protein AlgK precursor
3866	AAG07253.1	Pyocin protein
4175	AAG07562.1	Pvds-regulatedendoprotease, lysyl class
4224	AAG07612.1	Pyochelin biosynthetic protein PchG
4419	AAG07807.1	Cell division protein FtsL
4550	AAG07938.1	Typ 4 fimbrial biogenesis protein FimU
4552	AAG07940.1	Typeimfbrial biogenesis protein PilW
5415	AAG08802.1	Sarcosine oxidase delta subunit

Table 12.4:

Gene id Accession No. Protein Name

7	AAG03397.1	hypothetical protein PA0007
12	AAG03402.1	hypothetical protein PA0012
14	AAG03404.1	hypothetical protein PA0014
26	AAG03416.1	phospholipase C, PlcB
27	AAG03417.1	hypothetical protein PA0027
28	AAG03418.1	hypothetical protein PA0028
33	AAG03423.1	hypothetical protein PA0033

39	AAG03429.1	hypothetical protein PA0039
42	AAG03432.1	hypothetical protein PA0042
48	AAG03438.1	probable transcriptional regulator
50	AAG03440.1	hypothetical protein PA0050
52	AAG03442.1	hypothetical protein PA0052
60	AAG03450.1	conserved hypothetical protein
61	AAG03451.1	hypothetical protein PA0061
62	AAG03452.1	hypothetical protein PA0062
63	AAG03453.1	hypothetical protein PA0063
64	AAG03454.1	hypothetical protein PA0064
68	AAG03458.1	hypothetical protein PA0068
70	AAG03460.1	hypothetical protein PA0070
72	AAG03462.1	hypothetical protein PA0072
76	AAG03466.1	hypothetical protein PA0076
94	AAG03484.1	hypothetical protein PA0094
96	AAG03486.1	hypothetical protein PA0096
97	AAG03487.1	hypothetical protein PA0097
98	AAG03488.1	hypothetical protein PA0098
99	AAG03489.1	hypothetical protein PA0099
100	AAG03490.1	hypothetical protein PA0100
101	AAG03491.1	hypothetical protein PA0101
104	AAG03494.1	hypothetical protein PA0104
109	AAG03499.1	hypothetical protein PA0109
118	AAG03508.1	hypothetical protein PA0118
122	AAG03512.1	conserved hypothetical protein
124	AAG03514.1	hypothetical protein PA0124
125	AAG03515.1	hypothetical protein PA0125
126	AAG03516.1	hypothetical protein PA0126
127	AAG03517.1	hypothetical protein PA0127
135	AAG03525.1	hypothetical protein PA0135
144	AAG03534.1	hypothetical protein PA0144
145	AAG03535.1	hypothetical protein PA0145
160	AAG03550.1	hypothetical protein PA0160
161	AAG03551.1	hypothetical protein PA0161
170	AAG03560.1	hypothetical protein PA0170
171	AAG03561.1	hypothetical protein PA0171
187	AAG03577.1	hypothetical protein PA0187
201	AAG03589.1	hypothetical protein PA0200
235	AAG03623.1	hypothetical protein PA0234

252	AAG03640.1	hypothetical protein PA0251
253	AAG03641.1	hypothetical protein PA0252
254	AAG03642.1	probable transcriptional regulator
257	AAG03645.1	hypothetical proteinPA0256
259	AAG03647.1	hypothetical proteinPA0258
260	AAG03648.1	hypothetical protein PA0259
261	AAG03649.1	hypothetical protein PA0260
262	AAG03650.1	hypothetical protein PA0261
265	AAG03653.1	hypothetical protein PA0264
272	AAG03660.1	hypothetical protein PA0271
275	AAG03663.1	hypothetical protein PA0274
285	AAG03673.1	hypothetical protein PA0284
295	AAG03683.1	transcriptional regulatorAguR
308	AAG03696.1	hypothetical protein PA0307
310	AAG03698.1	hypothetical protein PA0309
312	AAG03700.1	hypothetical protein PA0311
316	AAG03704.1	hypothetical protein PA0315
333	AAG03721.1	hypothetical protein PA0332
347	AAG03735.1	hypothetical protein PA0346
349	AAG03737.1	hypothetical protein PA0348
350	AAG03738.1	hypothetical proteinPA0349
360	AAG03748.1	hypothetical proteinPA0359
366	AAG03754.1	hypothetical protein PA0365
370	AAG03758.1	hypothetical proteinPA0369
378	AAG03766.1	hypothetical protein PA0377
385	AAG03773.1	hypothetical protein PA0384
392	AAG03780.1	hypothetical proteinPA0391
399	AAG03787.1	hypothetical protein PA0398
419	AAG03807.1	hypothetical protein PA0418
425	AAG03813.1	multidrug resistance operon repressor MexR
430	AAG03818.1	hypothetical protein PA0429
434	AAG03822.1	hypothetical protein PA0433
443	AAG03831.1	hypothetical protein PA0442
454	AAG03842.1	hypothetical protein PA0453
467	AAG03855.1	hypothetical protein PA0466
470	AAG03858.1	hypothetical proteinPA0469
475	AAG03863.1	hypothetical protein PA0474
482	AAG03870.1	hypothetical proteinPA0481
506	AAG03894.1	hypothetical protein PA0505

510	AAG03898.1	probable c-type cytochrome
513	AAG03901.1	conserved hypothetical protein
514	AAG03902.1	probable transcriptional regulator
515	AAG03903.1	heme d1 biosynthesis protein NirL
516	AAG03904.1	probable transcriptional regulator
517	AAG03905.1	heme d1 biosynthesis protein NirF
518	AAG03906.1	probable c-type cytochrome precursor
520	AAG03908.1	nitrite reductaseprecursor
523	AAG03911.1	hypothetical protein PA0522
524	AAG03912.1	nitric-oxide reductase subunit C
527	AAG03915.1	hypothetical protein PA0526
533	AAG03921.1	hypothetical protein PA0532
550	AAG03938.1	hypothetical protein PA0549
554	AAG03942.1	hypothetical protein PA0553
555	AAG03943.1	hypothetical protein PA0554
557	AAG03945.1	hypothetical protein PA0556
562	AAG03950.1	hypothetical protein PA0561
567	AAG03955.1	hypothetical proteinPA0566
569	AAG03957.1	hypothetical protein PA0568
570	AAG03958.1	hypothetical protein PA0569
572	AAG03960.1	hypothetical protein PA0571
574	AAG03962.1	hypothetical protein PA0573
586	AAG03974.1	hypothetical proteinPA0585
600	AAG03988.1	hypothetical protein PA0599
611	AAG03999.1	transcriptional regulatorPrtN
613	AAG04001.1	hypothetical protein PA0612
614	AAG04002.1	hypothetical protein PA0613
615	AAG04003.1	hypothetical proteinPA0614
622	AAG04010.1	conserved hypothetical protein
625	AAG04013.1	hypothetical protein PA0624
628	AAG04016.1	conserved hypothetical protein
631	AAG04019.1	hypothetical proteinPA0630
632	AAG04020.1	hypothetical proteinPA0631
633	AAG04021.1	hypothetical proteinPA0632
635	AAG04023.1	hypothetical protein PA0634
636	AAG04024.1	hypothetical protein PA0635
643	AAG04031.1	hypothetical proteinPA0642
644	AAG04032.1	hypothetical proteinPA0643
645	AAG04033.1	hypothetical proteinPA0644

646	AAG04034.1	hypothetical proteinPA0645
647	AAG04035.1	hypothetical protein PA0646
648	AAG04036.1	hypothetical protein PA0647
649	AAG04037.1	hypothetical protein PA0648
674	AAG04062.1	hypothetical protein PA0673
675	AAG04063.1	hypothetical protein PA0674
676	AAG04064.1	probable sigma-70 factor, ECF subfamily
678	AAG04066.1	HxcW putative pseudopilin
679	AAG04067.1	HxcU putative pseudopilin
680	AAG04068.1	hypothetical protein PA0679
681	AAG04069.1	HxcV putative pseudopodia
684	AAG04072.1	probable type II secretion system protein
685	AAG04073.1	probable type II secretion system protein
692	AAG04080.1	hypothetical protein PA0691
696	AAG04084.1	hypothetical protein PA0695
697	AAG04085.1	hypothetical protein PA0696
698	AAG04086.1	hypothetical protein PA0697
699	AAG04087.1	hypothetical protein PA0698
701	AAG04089.1	hypothetical proteinPA0700
708	AAG04096.1	transcriptional regulatorToxR
710	AAG04098.1	hypothetical protein PA0709
714	AAG04102.1	hypothetical protein PA0713
715	AAG04103.1	hypothetical protein PA0714
717	AAG04105.1	hypothetical protein PA0716
718	AAG04106.1	hypothetical protein of bacteriophage Pf1
719	AAG04107.1	hypothetical protein of bacteriophage Pf1
720	AAG04108.1	hypothetical protein of bacteriophage Pf1
721	AAG04109.1	helix destabilizing protein of bacteriophage Pf1
722	AAG04110.1	hypothetical protein of bacteriophage Pf1
723	AAG04111.1	hypothetical protein of bacteriophage Pf1
724	AAG04112.1	coat protein B of bacteriophage Pf1)
726	AAG04114.1	hypothetical protein of bacteriophage Pf1
728	AAG04116.1	hypothetical protein from bacteriophage Pf1
730	AAG04118.1	hypothetical protein PA0729
731	AAG04119.1	probable transferase
735	AAG04123.1	hypothetical protein PA0734
736	AAG04124.1	hypothetical protein PA0735
737	AAG04125.1	hypothetical proteinPA0736
738	AAG04126.1	hypothetical protein PA0737

739	AAG04127.1	conserved hypothetical protein
743	AAG04131.1	hypothetical protein PA0742
764	AAG04152.1	anti-sigma factor MucA
770	AAG04158.1	hypothetical protein PA0769
777	AAG04165.1	hypothetical protein PA0776
779	AAG04167.1	inhibitor of cysteine peptidase
803	AAG04191.1	hypothetical protein PA0802
806	AAG04194.1	hypothetical protein PA0805
809	AAG04197.1	hypothetical protein PA0808
813	AAG04201.1	hypothetical protein PA0812
819	AAG04207.1	hypothetical protein PA0818
820	AAG04208.1	hypothetical protein PA0819
822	AAG04210.1	hypothetical protein PA0821
823	AAG04211.1	hypothetical protein PA0822
824	AAG04212.1	hypothetical protein PA0823
825	AAG04213.1	hypothetical protein PA0824
826	AAG04214.1	hypothetical protein PA0825
827	AAG04215.1	hypothetical protein PA0826
828	AAO42616.1	translated portion of mRNA genesrA
829	AAG04216.1	hypothetical protein PA0827
830	AAG04217.1	probable transcriptional regulator
845	AAG04232.1	phospholipase accessory protein PlcR precursor
852	AAG04239.1	hypothetical protein PA0850
876	AAG04263.1	hypothetical protein PA0874
896	AAG04283.1	hypothetical protein PA0894
902	AAG04289.1	hypothetical protein PA0900
909	AAG04296.1	hypothetical protein PA0907
912	AAG04299.1	hypothetical protein PA0910
913	AAG04300.1	hypothetical protein PA0911
914	AAG04301.1	hypothetical protein PA0912
916	AAG04303.1	hypothetical protein PA0914
923	AAG04310.1	hypothetical protein PA0921
924	AAG04311.1	hypothetical protein PA0922
928	AAG04315.1	hypothetical protein PA0926
940	AAG04327.1	hypothetical protein PA0938
941	AAG04328.1	hypothetical protein PA0939
942	AAG04329.1	hypothetical protein PA0940
943	AAG04330.1	hypothetical protein PA0941
944	AAG04331.1	probable transcriptional regulator

950	AAG04337.1	hypothetical protein PA0948
954	AAG04341.1	hypothetical protein PA0952
957	AAG04344.1	hypothetical protein PA0955
962	AAG04349.1	hypothetical protein PA0960
979	AAG04366.1	hypothetical protein PA0977
982	AAG04369.1	hypothetical protein PA0980
983	AAG04370.1	hypothetical protein PA0981
984	AAG04371.1	hypothetical protein PA0982
986	AAG04373.1	colicin immunity protein
987	AAG04374.1	pyocin S5
990	AAG04377.1	hypothetical protein PA0988
991	AAG04378.1	hypothetical protein PA0989
999	AAG04386.1	Homologous to beta-keto-acyl-acyl-carrier protein synthase
1028	AAG04415.1	hypothetical protein PA1026
1036	AAG04423.1	hypothetical protein PA1034
1037	AAG04424.1	hypothetical protein PA1035
1038	AAG04425.1	hypothetical protein PA1036
1042	AAG04429.1	hypothetical protein PA1040
1044	AAG04431.1	conserved hypothetical protein
1064	AAG04451.1	hypothetical protein PA1062
1065	AAG04452.1	hypothetical protein PA1063
1077	AAG04464.1	hypothetical protein PA1075
1078	AAG04465.1	hypothetical protein PA1076
1090	AAG04477.1	hypothetical protein PA1088
1092	AAG04479.1	hypothetical protein PA1090
1098	AAG04485.1	hypothetical protein PA1096
1113	AAG04500.1	hypothetical protein PA1111
1119	AAG04506.1	hypothetical protein PA1117
1120	AAG04507.1	hypothetical protein PA1118
1125	AAG04512.1	hypothetical protein PA1123
1128	AAG04515.1	hypothetical protein PA1126
1134	AAG04521.1	hypothetical protein PA1132
1135	AAG04522.1	hypothetical protein PA1133
1150	AAG04537.1	exotoxin A precursor
1151	AAG04538.1	hypothetical protein PA1149
1154	AAG04541.1	hypothetical protein PA1152
1155	AAG04542.1	hypothetical protein PA1153
1162	AAG04549.1	hypothetical protein PA1160
1169	AAG04556.1	hypothetical protein PA1167

1170	AAG04557.1	hypothetical protein PA1168
1171	AAG04558.1	probable lipxygenase
1180	AAG04567.1	PhoP/Q and low Mg ²⁺ -inducible outer membrane protein H1 precursor
1208	AAG04595.1	hypothetical protein PA1206
1213	AAG04600.1	hypothetical protein PA1211
1218	AAG04605.1	hypothetical protein PA1216
1221	AAG04608.1	hypothetical protein PA1219
1222	AAG04609.1	hypothetical protein PA1220
1229	AAG04616.1	hypothetical protein PA1227
1230	AAG04617.1	hypothetical protein PA1228
1232	AAG04619.1	hypothetical protein PA1230
1235	AAG04622.1	hypothetical protein PA1233
1236	AAG04623.1	hypothetical protein PA1234
1244	AAG04631.1	hypothetical protein PA1242
1246	AAG04633.1	hypothetical protein PA1244
1247	AAG04634.1	hypothetical protein PA1245
1252	AAG04639.1	alkaline proteinase inhibitor AprI
1291	AAG04678.1	hypothetical protein PA1289
1295	AAG04682.1	hypothetical protein PA1293
1307	AAG04694.1	hypothetical protein PA1305
1310	AAG04697.1	hypothetical protein PA1308
1326	AAG04713.1	hypothetical protein PA1324
1334	AAG04721.1	hypothetical protein PA1332
1335	AAG04722.1	hypothetical protein PA1333
1345	AAG04732.1	hypothetical protein PA1343
1349	AAG04736.1	probable transcriptional regulator
1350	AAG04737.1	hypothetical protein PA1348
1357	AAG04744.1	hypothetical protein PA1355
1358	AAG04745.1	hypothetical protein PA1356
1364	AAG04751.1	hypothetical protein PA1362
1371	AAG04758.1	hypothetical protein PA1369
1372	AAG04759.1	hypothetical protein PA1370
1373	AAG04760.1	hypothetical protein PA1371
1380	AAG04767.1	hypothetical protein PA1378
1383	AAG04770.1	hypothetical protein PA1381
1385	AAG04772.1	hypothetical protein PA1383
1389	AAG04776.1	hypothetical protein PA1387
1390	AAG04777.1	hypothetical protein PA1388

1396	AAG04783.1	hypothetical protein PA1394
1397	AAG04784.1	hypothetical protein PA1395
1400	AAG04787.1	hypothetical protein PA1398
1406	AAG04793.1	hypothetical protein PA1404
1416	AAG04803.1	hypothetical protein PA1414
1422	AAG04809.1	hypothetical protein PA1420
1428	AAG04815.1	hypothetical protein PA1426
1429	AAG04816.1	hypothetical protein PA1427
1433	AAG04820.1	regulatory protein RsaL
1470	AAG04857.1	hypothetical protein PA1468
1473	AAG04860.1	hypothetical protein PA1471
1476	AAG04863.1	hypothetical protein PA1474
1480	AAG04867.1	hypothetical protein PA1478
1494	AAG04881.1	hypothetical protein PA1492
1496	AAG04883.1	conserved hypothetical protein
1510	AAG04897.1	hypothetical protein PA1508
1511	AAG04898.1	hypothetical protein PA1509
1512	AAG04899.1	hypothetical protein PA1510
1532	AAG04919.1	hypothetical protein PA1530
1533	AAG04920.1	hypothetical protein PA1531
1547	AAG04934.1	hypothetical protein PA1545
1573	AAG04960.1	hypothetical protein PA1571
1577	AAG04964.1	hypothetical protein PA1575
1580	AAG04967.1	hypothetical protein PA1578
1581	AAG04968.1	hypothetical protein PA1579
1594	AAG04981.1	hypothetical protein PA1592
1597	AAG04984.1	hypothetical protein PA1595
1607	AAG04994.1	hypothetical protein PA1605
1608	AAG04995.1	hypothetical protein PA1606
1614	AAG05001.1	hypothetical protein PA1612
1617	AAG05004.1	probable lipase
1626	AAG05013.1	hypothetical protein PA1624
1634	AAG05021.1	KdpF protein
1641	AAG05028.1	hypothetical protein PA1639
1643	AAG05030.1	hypothetical protein PA1641
1647	AAG05034.1	hypothetical protein PA1645
1654	AAG05041.1	hypothetical protein PA1652
1666	AAG05053.1	hypothetical protein PA1664
1678	AAG05065.1	hypothetical protein PA1676

1681	AAG05068.1	hypothetical protein PA1679
1690	AAG05077.1	hypothetical protein PA1688
1701	AAG05088.1	conserved hypothetical protein in type III secretion
1702	AAG05089.1	conserved hypothetical protein in type III secretion
1703	AAG05090.1	conserved hypothetical protein in type III secretion
1704	AAG05091.1	conserved hypothetical protein in type III secretion
1706	AAG05093.1	transcriptional regulator protein PcrR
1707	AAG05094.1	regulator in type III secretion
1708	AAG05095.1	type III secretion protein PcrV
1710	AAG05097.1	Translocator protein PopB
1711	AAG05098.1	Translocator outer membrane protein PopD precursor
1712	AAG05099.1	ExsC, exoenzyme S synthesis protein C precursor.
1713	AAG05100.1	ExsE
1714	AAG05101.1	exoenzyme S synthesis protein B
1716	AAG05103.1	ExsD
1717	AAG05104.1	type III export apparatus protein
1720	AAG05107.1	type III export protein PscE
1721	AAG05108.1	type III export protein PscF
1722	AAG05109.1	type III export protein PscG
1723	AAG05110.1	type III export protein PscH
1724	AAG05111.1	type III export protein PscI
1726	AAG05113.1	type III export protein PscK
1730	AAG05117.1	hypothetical protein PA1728
1736	AAG05123.1	hypothetical protein PA1734
1745	AAG05132.1	hypothetical protein PA1743
1746	AAG05133.1	hypothetical protein PA1744
1747	AAG05134.1	hypothetical protein PA1745
1748	AAG05135.1	hypothetical protein PA1746
1749	AAG05136.1	hypothetical protein PA1747
1753	AAG05140.1	hypothetical protein PA1751
1755	AAG05142.1	conserved hypothetical protein
1757	AAG05144.1	hypothetical protein PA1755
1763	AAG05150.1	hypothetical protein PA1761
1767	AAG05154.1	hypothetical protein PA1765
1776	AAG05163.1	CfrX protein
1786	AAG05173.1	hypothetical protein PA1784
1790	AAG05177.1	hypothetical protein PA1788
1793	AAG05180.1	hypothetical protein PA1791
1832	AAG05219.1	hypothetical protein PA1830

1836	AAG05223.1	hypothetical proteinPA1834
1842	AAG05229.1	hypothetical protein PA1840
1846	AAG05233.1	hypothetical protein PA1844
1847	AAG05234.1	hypothetical protein PA1845
1848	AAG05235.1	cis/trans isomerase
1854	AAG05241.1	hypothetical protein PA1852
1857	AAG05244.1	hypothetical protein PA1855
1867	AAG05254.1	hypothetical protein PA1865
1869	AAG05256.1	hypothetical protein PA1867
1872	AAG05259.1	hypothetical protein PA1870
1873	AAG05260.1	LasA protease precursor
1881	AAG05268.1	hypothetical protein PA1879
1889	AAG05276.1	hypothetical proteinPA1887
1891	AAG05278.1	hypothetical protein PA1889
1893	AAG05280.1	hypothetical protein PA1891
1894	AAG05281.1	hypothetical protein PA1892
1896	AAG05283.1	hypothetical protein PA1894
1897	AAG05284.1	hypothetical protein PA1895
1898	AAG05285.1	hypothetical protein PA1896
1901	AAG05288.1	probable phenazinebiosynthesis protein
1902	AAG05289.1	probable phenazinebiosynthesis protein
1908	AAG05294.1	hypothetical protein PA1906
1915	AAG05301.1	hypothetical protein PA1913
1916	AAG05302.1	conserved hypothetical protein
1931	AAG05317.1	hypothetical protein PA1929
1936	AAG05322.1	hypothetical protein PA1934
1937	AAG05323.1	hypothetical protein PA1935
1938	AAG05324.1	hypothetical protein PA1936
1943	AAG05329.1	hypothetical protein PA1941
1944	AAG05330.1	hypothetical protein PA1942
1954	AAG05340.1	hypothetical protein PA1952
1956	AAG05342.1	hypothetical protein PA1954
1957	AAG05343.1	hypothetical protein PA1955
1958	AAG05344.1	hypothetical protein PA1956
1965	AAG05351.1	hypothetical protein PA1963
1967	AAG05353.1	hypothetical protein PA1965
1968	AAG05354.1	hypothetical protein PA1966
1969	AAG05355.1	hypothetical protein PA1967
1970	AAG05356.1	hypothetical protein PA1968

1972	AAG05358.1	hypothetical protein PA1970
1976	AAG05362.1	hypothetical protein PA1974
1977	AAG05363.1	hypothetical protein PA1975
1985	AAG05371.1	cytochrome c550
1987	AAG05373.1	pyrroloquinoline quinonebiosynthesis protein A
1990	AAG05376.1	pyrroloquinoline quinonebiosynthesis protein D
1997	AAG05383.1	hypothetical protein PA1995
2023	AAG05409.1	hypothetical protein PA2021
2026	AAG05412.1	probable ring-cleavingdioxxygenase
2029	AAG05415.1	hypothetical protein PA2027
2032	AAG05418.1	hypothetical protein PA2030
2033	AAG05419.1	hypothetical protein PA2031
2038	AAG05424.1	hypothetical protein PA2036
2039	AAG05425.1	hypothetical protein PA2037
2050	AAG05436.1	hypothetical protein PA2048
2079	AAG05465.1	hypothetical protein PA2077
2080	AAG05466.1	hypothetical protein PA2078
2089	AAG05475.1	hypothetical protein PA2087
2090	AAG05476.1	hypothetical protein PA2088
2097	AAG05483.1	hypothetical protein PA2095
2104	AAG05490.1	hypothetical proteinPA2102
2109	AAG05495.1	hypothetical protein PA2107
2111	AAG05497.1	hypothetical protein PA2109
2119	AAG05505.1	hypothetical protein PA2117
2124	AAG05510.1	hypothetical protein PA2122
2136	AAG05522.1	hypothetical protein PA2134
2138	AAG05524.1	hypothetical protein PA2136
2139	AAG05525.1	hypothetical protein PA2137
2141	AAG05527.1	hypothetical protein PA2139
2142	AAG05528.1	probable metallothionein
2145	AAG05531.1	hypothetical proteinPA2143
2147	AAG05533.1	hypothetical proteinPA2145
2151	AAG05537.1	hypothetical protein PA2149
2168	AAG05554.1	hypothetical protein PA2166
2171	AAG05557.1	hypothetical protein PA2169
2172	AAG05558.1	hypothetical protein PA2170
2175	AAG05561.1	hypothetical protein PA2173
2176	AAG05562.1	hypothetical protein PA2174
2177	AAG05563.1	hypothetical proteinPA2175

2178	AAG05564.1	hypothetical proteinPA2176
2184	AAG05570.1	hypothetical protein PA2182
2188	AAG05574.1	hypothetical protein PA2186
2191	AAG05577.1	hypothetical proteinPA2189
2203	AAG05589.1	hypothetical proteinPA2201
2207	AAG05593.1	hypothetical protein PA2205
2210	AAG05596.1	hypothetical protein PA2208
2223	AAG05609.1	conserved hypothetical protein
2224	AAG05610.1	hypothetical protein PA2222
2225	AAG05611.1	hypothetical protein PA2223
2226	AAG05612.1	hypothetical protein PA2224
2227	AAG05613.1	hypothetical protein PA2225
2228	AAG05614.1	hypothetical protein PA2226
2239	AAG05625.1	probable glycosyl hydrolase
2242	AAG05628.1	hypothetical proteinPA2240
2247	AAG05633.1	hypothetical protein PA2245
2253	AAG05639.1	hypothetical protein PA2251
2262	AAG05648.1	hypothetical protein PA2260
2271	AAG05657.1	conserved hypothetical protein
2276	AAG05662.1	hypothetical protein PA2274
2285	AAG05671.1	hypothetical protein PA2283
2286	AAG05672.1	hypothetical protein PA2284
2287	AAG05673.1	hypothetical protein PA2285
2288	AAG05674.1	hypothetical protein PA2286
2289	AAG05675.1	hypothetical protein PA2287
2290	AAG05676.1	hypothetical protein PA2288
2294	AAG05680.1	hypothetical protein PA2292
2303	AAG05689.1	hypothetical protein PA2301
2305	AAG05691.1	hypothetical protein PA2303
2306	AAG05692.1	hypothetical protein PA2304
2313	AAG05699.1	hypothetical protein PA2311
2315	AAG05701.1	hypothetical protein PA2313
2360	AAG05746.1	hypothetical protein PA2358
2364	AAG05750.1	hypothetical protein PA2362
2365	AAG05751.1	hypothetical protein PA2363
2366	AAG05752.1	hypothetical protein PA2364
2370	AAG05756.1	hypothetical protein PA2368
2374	AAG05760.1	hypothetical protein PA2372
2377	AAG05763.1	hypothetical protein PA2375

2382	AAG05768.1	hypothetical protein PA2380
2383	AAG05769.1	hypothetical protein PA2381
2386	AAG05772.1	hypothetical protein PA2384
2394	AAG05780.1	PvdP
2398	AAG05784.1	pyoverdine synthetase F
2405	AAG05792.1	hypothetical protein PA2404
2406	AAG05793.1	hypothetical protein PA2405
2407	AAG05794.1	hypothetical protein PA2406
2411	AAG05798.1	hypothetical protein PA2410
2416	AAG05803.1	hypothetical protein PA2415
2423	AAG05810.1	hypothetical protein PA2422
2424	AAG05811.1	hypothetical protein PA2423
2428	AAG05815.1	hypothetical protein PA2427
2430	AAG05817.1	hypothetical protein PA2429
2434	AAG05821.1	hypothetical protein PA2433
2435	AAG05822.1	hypothetical protein PA2434
2437	AAG05824.1	hypothetical protein PA2436
2441	AAG05828.1	hypothetical protein PA2440
2442	AAG05829.1	hypothetical protein PA2441
2451	AAG05838.1	hypothetical protein PA2450
2453	AAG05840.1	hypothetical protein PA2452
2457	AAG05844.1	hypothetical protein PA2456
2460	AAG05847.1	hypothetical protein PA2459
2461	AAG05848.1	hypothetical protein PA2460
2462	AAG05849.1	hypothetical protein PA2461
2465	AAG05852.1	hypothetical protein PA2464
2486	AAG05873.1	hypothetical protein PA2485
2487	AAG05874.1	hypothetical protein PA2486
2488	AAG05875.1	hypothetical protein PA2487
2497	AAG05884.1	hypothetical protein PA2496
2502	AAG05889.1	hypothetical protein PA2501
2504	AAG05891.1	hypothetical protein PA2503
2507	AAG05894.1	hypothetical protein PA2506
2560	AAG05947.1	hypothetical protein PA2559
2561	AAG05948.1	hypothetical protein PA2560
2563	AAG05950.1	hypothetical protein PA2562
2566	AAG05953.1	hypothetical protein PA2565
2569	AAG05956.1	hypothetical protein PA2568
2570	AAG05957.1	hypothetical protein PA2569

2571	AAG05958.1	PA-I galactophilic lectin
2582	AAG05969.1	hypothetical protein PA2581
2583	AAG05970.1	hypothetical protein PA2582
2594	AAG05981.1	hypothetical protein PA2593
2608	AAG05995.1	conserved hypothetical protein
2633	AAG06020.1	hypothetical proteinPA2632
2634	AAG06021.1	hypothetical proteinPA2633
2656	AAG06043.1	hypothetical protein PA2655
2660	AAG06047.1	hypothetical proteinPA2659
2661	AAG06048.1	hypothetical protein PA2660
2662	AAG06049.1	hypothetical protein PA2661
2664	AAG06051.1	hypothetical protein PA2663
2668	AAG06055.1	conserved hypothetical protein
2669	AAG06056.1	hypothetical protein PA2668
2670	AAG06057.1	hypothetical proteinPA2669
2671	AAG06058.1	hypothetical proteinPA2670
2672	AAG06059.1	hypothetical proteinPA2671
2673	AAG06060.1	probable type II secretion system protein
2674	AAG06061.1	probable type II secretion system protein
2675	AAG06062.1	probable type II secretion system protein
2703	AAG06090.1	hypothetical protein PA2702
2704	AAG06091.1	hypothetical protein PA2703
2709	AAG06096.1	hypothetical protein PA2708
2721	AAG06108.1	hypothetical protein PA2720
2724	AAG06111.1	hypothetical protein PA2723
2725	AAG06112.1	hypothetical protein PA2724
2730	AAG06117.1	hypothetical proteinPA2729
2732	AAG06119.1	hypothetical protein PA2731
2737	AAG06124.1	hypothetical protein PA2736
2747	AAG06134.1	hypothetical proteinPA2746
2748	AAG06135.1	hypothetical protein PA2747
2751	AAG06138.1	hypothetical protein PA2750
2754	AAG06141.1	hypothetical protein PA2753
2760	AAG06147.1	hypothetical protein PA2759
2762	AAG06149.1	hypothetical protein PA2761
2764	AAG06151.1	hypothetical protein PA2763
2769	AAG06156.1	hypothetical proteinPA2768
2773	AAG06160.1	hypothetical protein PA2772
2774	AAG06161.1	hypothetical protein PA2773

2775	AAG06162.1	hypothetical protein PA2774
2776	AAG06163.1	hypothetical protein PA2775
2779	AAG06166.1	hypothetical protein PA2778
2780	AAG06167.1	hypothetical protein PA2779
2782	AAG06169.1	hypothetical protein PA2781
2783	AAG06170.1	hypothetical protein PA2782
2784	AAG06171.1	hypothetical protein PA2783
2785	AAG06172.1	hypothetical protein PA2784
2791	AAG06178.1	hypothetical protein PA2790
2792	AAG06179.1	hypothetical protein PA2791
2793	AAG06180.1	hypothetical protein PA2792
2794	AAG06181.1	hypothetical protein PA2793
2795	AAG06182.1	hypothetical protein PA2794
2800	AAG06187.1	hypothetical protein PA2799
2806	AAG06193.1	hypothetical protein PA2805
2809	AAG06196.1	hypothetical protein PA2808
2815	AAG06202.1	hypothetical protein PA2814
2817	AAG06204.1	hypothetical protein PA2816
2818	AAG06205.1	hypothetical protein PA2817
2820	AAG06207.1	hypothetical protein PA2819
2821	AAG06208.1	hypothetical protein PA2820
2830	AAG06217.1	hypothetical protein PA2829
2846	AAG06233.1	hypothetical protein PA2845
2854	AAG06241.1	Outer membrane lipoprotein OprI precursor
2856	AAG06243.1	hypothetical protein PA2855
2861	AAG06248.1	hypothetical protein PA2860
2869	AAG06256.1	hypothetical protein PA2868
2873	AAG06260.1	hypothetical protein PA2872
2884	AAG06271.1	hypothetical protein PA2883
2895	AAG06282.1	hypothetical protein PA2894
2896	AAG06283.1	hypothetical protein PA2895
2899	AAG06286.1	hypothetical protein PA2898
2902	AAG06289.1	hypothetical protein PA2901
2903	AAG06290.1	hypothetical protein PA2902
2920	AAG06307.1	hypothetical protein PA2919
2928	AAG06315.1	hypothetical protein PA2927
2929	AAG06316.1	hypothetical protein PA2928
2936	AAG06323.1	hypothetical protein PA2935
2938	AAG06325.1	hypothetical protein PA2937

2948	AAG06335.1	hypothetical protein PA2947
2956	AAG06343.1	hypothetical protein PA2955
2990	AAG06377.1	hypothetical protein PA2989
2993	AAG06380.1	hypothetical protein PA2992
3004	AAG06391.1	hypothetical protein PA3003
3009	AAG06396.1	hypothetical protein PA3008
3010	AAG06397.1	hypothetical protein PA3009
3011	AAG06398.1	hypothetical protein PA3010
3016	AAG06403.1	hypothetical protein PA3015
3017	AAG06404.1	hypothetical protein PA3016
3019	AAG06406.1	hypothetical protein PA3018
3033	AAG06420.1	cytochrome c Snr1
3034	AAG06421.1	hypothetical protein PA3033
3042	AAG06429.1	hypothetical protein PA3041
3043	AAG06430.1	hypothetical protein PA3042
3052	AAG06439.1	hypothetical protein PA3051
3053	AAG06440.1	hypothetical protein PA3052
3056	AAG06443.1	hypothetical protein PA3055
3057	AAG06444.1	hypothetical protein PA3056
3058	AAG06445.1	hypothetical protein PA3057
3059	AAG06446.1	hypothetical protein PA3058
3061	AAG06448.1	hypothetical protein PA3060
3062	AAG06449.1	hypothetical protein PA3061
3063	AAG06450.1	hypothetical protein PA3062
3064	AAG06451.1	hypothetical protein PA3063
3066	AAG06453.1	hypothetical protein PA3065
3073	AAG06460.1	hypothetical protein PA3072
3077	AAG06464.1	hypothetical protein PA3076
3091	AAG06478.1	hypothetical protein PA3090
3092	AAG06479.1	hypothetical protein PA3091
3105	AAG06492.1	secretion protein XcpP
3144	AAG06531.1	hypothetical protein PA3143
3155	AAG06542.1	B-band O-antigen polymerase
3179	AAG06566.1	hypothetical protein PA3178
3197	AAG06584.1	hypothetical protein PA3196
3204	AAG06591.1	hypothetical protein PA3203
3206	AAG06593.1	hypothetical protein PA3205
3208	AAG06595.1	hypothetical protein PA3207
3219	AAG06606.1	hypothetical protein PA3218

3225	AAG06612.1	hypothetical protein PA3224
3230	AAG06617.1	hypothetical protein PA3229
3239	AAG06626.1	hypothetical proteinPA3238
3260	AAG06647.1	hypothetical protein PA3259
3261	AAG06648.1	probable transcriptional regulator
3262	AAG06649.1	hypothetical protein PA3261
3274	AAG06661.1	hypothetical proteinPA3273
3275	AAG06662.1	hypothetical proteinPA3274
3277	AAG06664.1	hypothetical protein PA3276
3279	AAG06666.1	hypothetical protein PA3278
3291	AAG06678.1	hypothetical protein PA3290
3292	AAG06679.1	hypothetical protein PA3291
3293	AAG06680.1	hypothetical protein PA3292
3294	AAG06681.1	hypothetical protein PA3293
3299	AAG06686.1	hypothetical protein PA3298
3308	AAG06695.1	hypothetical protein PA3307
3319	AAG06706.1	hypothetical protein PA3318
3324	AAG06711.1	conserved hypothetical protein
3335	AAG06722.1	probable acyl carrier protein
3336	AAG06723.1	hypothetical protein PA3335
3346	AAG06733.1	hypothetical protein PA3345
3348	AAG06735.1	hypothetical protein PA3347
3353	AAG06740.1	hypothetical protein PA3352
3354	AAG06741.1	hypothetical proteinPA3353
3355	AAG06742.1	hypothetical proteinPA3354
3362	AAG06749.1	fucose-binding lectinPA-IIL
3364	AAG06751.1	aliphatic amidase regulator
3368	AAG06755.1	hypothetical protein PA3367
3370	AAG06757.1	hypothetical protein PA3369
3371	AAG06758.1	hypothetical protein PA3370
3372	AAG06759.1	hypothetical protein PA3371
3386	AAG06773.1	alginate and motility regulator Z
3390	AAG06777.1	hypothetical protein PA3390
3391	AAG06778.1	probable ring-cleaving dioxygenase
3393	AAG06780.1	nitrous-oxide reductase precursor
3400	AAG06787.1	hypothetical protein PA3399
3412	AAG06799.1	hypothetical protein PA3411
3413	AAG06800.1	hypothetical protein PA3412
3415	AAG06802.1	hypothetical protein PA3414

3424	AAG06811.1	probable transcriptional regulator
3429	AAG06816.1	hypothetical protein PA3428
3442	AAG06829.1	Probable molybdopterin-binding protein
3452	AAG06839.1	hypothetical protein PA3451
3464	AAG06851.1	conserved hypothetical protein
3465	AAG06852.1	hypothetical protein PA3464
3480	AAG06867.1	Rhamnosyl transferase chain A
3485	AAG06872.1	hypothetical protein PA3484
3486	AAG06873.1	hypothetical protein PA3485
3488	AAG06875.1	Phospholipase D
3489	AAG06876.1	hypothetical protein PA3488
3497	AAG06884.1	hypothetical protein PA3496
3502	AAG06889.1	hypothetical protein PA3501
3503	AAG06890.1	hypothetical protein PA3502
3519	AAG06906.1	hypothetical protein PA3518
3520	AAG06907.1	hypothetical protein PA3519
3521	AAG06908.1	hypothetical protein PA3520
3537	AAG06924.1	hypothetical protein PA3536
3543	AAG06930.1	alginate biosynthesis protein Alg44
3545	AAG06932.1	Alginate production outer membrane protein AlgE precursor
3546	AAG06933.1	lginate-c5-mannuronan-epimerase AlgG
3547	AAG06934.1	alginate biosynthesis protein AlgX
3548	AAG06935.1	poly (beta-d-mannuronate) lyase precursor AlgL
3550	AAG06937.1	alginate o-acetyltransferase AlgJ
3551	AAG06938.1	Alginate o-acetyltransferase AlgF
3573	AAG06960.1	hypothetical protein PA3572
3577	AAG06964.1	hypothetical protein PA3576
3578	AAG06965.1	hypothetical protein PA3577
3612	AAG06999.1	hypothetical protein PA3611
3620	AAG07007.1	hypothetical protein PA3619
3662	AAG07049.1	hypothetical protein PA3661
3663	AAG07050.1	hypothetical protein PA3662
3664	AAG07051.1	hypothetical protein PA3663
3670	AAG07057.1	hypothetical protein PA3669
3671	AAG07058.1	hypothetical protein PA3670
3676	AAG07063.1	hypothetical protein PA3675
3685	AAG07072.1	hypothetical protein PA3684
3689	AAG07076.1	hypothetical protein PA3688

3692	AAG07079.1	hypothetical protein PA3691
3695	AAG07082.1	hypothetical protein PA3694
3699	AAG07086.1	hypothetical protein PA3698
3705	AAG07092.1	hypothetical proteinPA3705
3707	AAG07094.1	hypothetical proteinPA3707
3713	AAG07100.1	hypothetical protein PA3713
3719	AAG07106.1	hypothetical protein PA3719
3720	AAG07107.1	hypothetical protein PA3720
3722	AAG07109.1	hypothetical protein PA3722
3728	AAG07115.1	hypothetical proteinPA3728
3730	AAG07117.1	hypothetical protein PA3730
3734	AAG07121.1	hypothetical protein PA3734
3740	AAG07127.1	hypothetical protein PA3740
3752	AAG07139.1	hypothetical protein PA3752
3771	AAG07158.1	probable transcriptional regulator
3772	AAG07159.1	hypothetical protein PA3772
3784	AAG07171.1	hypothetical protein PA3784
3786	AAG07173.1	hypothetical protein PA3786
3788	AAG07175.1	hypothetical proteinPA3788
3791	AAG07178.1	hypothetical protein PA3791
3793	AAG07180.1	hypothetical protein PA3793
3794	AAG07181.1	hypothetical protein PA3794
3833	AAG07220.1	hypothetical protein PA3833
3835	AAG07222.1	hypothetical proteinPA3835
3841	AAG07228.1	exoenzyme S
3842	AAG07229.1	probable chaperone
3843	AAG07230.1	hypothetical protein PA3843
3850	AAG07237.1	hypothetical protein PA3850
3852	AAG07239.1	hypothetical protein PA3852
3855	AAG07242.1	hypothetical proteinPA3855
3856	AAG07243.1	hypothetical proteinPA3856
3868	AAG07255.1	hypothetical protein PA3868
3869	AAG07256.1	hypothetical protein PA3869
3880	AAG07267.1	conserved hypothetical protein
3881	AAG07268.1	hypothetical protein PA3881
3884	AAG07271.1	hypothetical protein PA3884
3904	AAG07291.1	hypothetical protein PA3904
3905	AAG07292.1	hypothetical protein PA3905
3906	AAG07293.1	hypothetical protein PA3906

3907	AAG07294.1	hypothetical protein PA3907
3908	AAG07295.1	hypothetical protein PA3908
3928	AAG07315.1	hypothetical protein PA3928
3939	AAG07326.1	hypothetical protein PA3939
3959	AAG07346.1	hypothetical protein PA3959
3960	AAG07347.1	hypothetical protein PA3960
3962	AAG07349.1	hypothetical protein PA3962
3964	AAG07351.1	hypothetical protein PA3964
3967	AAG07354.1	hypothetical protein PA3967
3986	AAG07373.1	hypothetical protein PA3986
3990	AAG07377.1	conserved hypothetical protein
3991	AAG07378.1	hypothetical protein PA3991
4018	AAG07405.1	hypothetical protein PA4018
4028	AAG07415.1	hypothetical protein PA4028
4033	AAG07420.1	hypothetical protein PA4033
4039	AAG07426.1	hypothetical protein PA4039
4040	AAG07427.1	hypothetical protein PA4040
4046	AAG07433.1	hypothetical protein PA4046
4058	AAG07445.1	hypothetical protein PA4058
4062	AAG07449.1	hypothetical protein PA4062
4071	AAG07458.1	hypothetical protein PA4071
4076	AAG07463.1	hypothetical protein PA4076
4081	AAG07468.1	fimbrial subunit CupB6
4087	AAG07474.1	conserved hypothetical protein
4090	AAG07477.1	hypothetical protein PA4090
4103	AAG07490.1	hypothetical protein PA4103
4111	AAG07498.1	hypothetical protein PA4111
4116	AAG07503.1	hypothetical protein PA4116
4134	AAG07521.1	hypothetical protein PA4134
4139	AAG07526.1	hypothetical protein PA4139
4140	AAG07527.1	hypothetical protein PA4140
4141	AAG07528.1	hypothetical protein PA4141
4146	AAG07533.1	hypothetical protein PA4146
4149	AAG07536.1	conserved hypothetical protein
4183	AAG07570.1	hypothetical protein PA4183
4205	AAG07592.1	hypothetical protein PA4205
4210	AAG07597.1	probable phenazinebiosynthesis protein
4211	AAG07598.1	probable phenazinebiosynthesis protein
4220	AAG07608.1	hypothetical protein PA4220

4230	AAG07618.1	salicylate biosynthesis protein PchB
4278	AAG07666.1	hypothetical protein PA4278
4286	AAG07674.1	hypothetical protein PA4286
4291	AAG07679.1	hypothetical protein PA4291
4294	AAG07682.1	hypothetical protein PA4294
4295	AAG07683.1	hypothetical protein PA4295
4298	AAG07686.1	hypothetical protein PA4298
4306	AAG07694.1	hypothetical protein PA4306
4315	AAG07703.1	transcriptional regulator MvaT, P16 subunit
4320	AAG07708.1	hypothetical protein PA4320
4321	AAG07709.1	hypothetical protein PA4321
4324	AAG07712.1	hypothetical protein PA4324
4325	AAG07713.1	hypothetical protein PA4325
4326	AAG07714.1	hypothetical protein PA4326
4327	AAG07715.1	hypothetical protein PA4327
4346	AAG07734.1	hypothetical protein PA4346
4360	AAG07748.1	hypothetical protein PA4360
4368	AAG07756.1	hypothetical protein PA4368
4377	AAG07765.1	hypothetical protein PA4377
4379	AAG07767.1	conserved hypothetical protein
4405	AAG07793.1	hypothetical protein PA4405
4437	AAG07825.1	hypothetical protein PA4437
4441	AAG07829.1	hypothetical protein PA4441
4452	AAG07840.1	conserved hypothetical protein
4469	AAG07857.1	hypothetical protein PA4469
4471	AAG07859.1	hypothetical protein PA4471
4531	AAG07919.1	hypothetical protein PA4531
4535	AAG07923.1	hypothetical protein PA4535
4537	AAG07925.1	hypothetical protein PA4537
4551	AAG07939.1	type 4 fimbrial biogenesis protein PilV
4553	AAG07941.1	type 4 fimbrial biogenesis protein PilX
4555	AAG07943.1	type 4 fimbrial biogenesis protein PilY2
4570	AAG07958.1	hypothetical protein PA4570
4573	AAG07961.1	hypothetical protein PA4573
4575	AAG07963.1	hypothetical protein PA4575
4578	AAG07966.1	hypothetical protein PA4578
4586	AAG07974.1	hypothetical protein PA4586
4590	AAG07978.1	protein activator
4596	AAG07984.1	probable transcriptional regulator

4600	AAG07988.1	transcriptional regulator NfxB
4603	AAG07991.1	hypothetical protein PA4603
4607	AAG07995.1	hypothetical protein PA4607
4610	AAG07998.1	hypothetical protein PA4610
4611	AAG07999.1	hypothetical protein PA4611
4623	AAG08011.1	hypothetical protein PA4623
4630	AAG08018.1	hypothetical protein PA4630
4637	AAG08025.1	hypothetical protein PA4637
4638	AAG08026.1	hypothetical protein PA4638
4639	AAG08027.1	hypothetical protein PA4639
4641	AAG08029.1	hypothetical protein PA4642
4642	AAG08030.1	hypothetical protein PA4643
4660	AAG08048.1	Lipid A 3-O-deacylase
4676	AAG08064.1	hypothetical protein PA4677
4678	AAG08066.1	hypothetical protein PA4679
4679	AAG08067.1	hypothetical protein PA4680
4680	AAG08068.1	hypothetical protein PA4681
4681	AAG08069.1	hypothetical protein PA4682
4682	AAG08070.1	hypothetical protein PA4683
4683	AAG08071.1	hypothetical protein PA4684
4684	AAG08072.1	hypothetical protein PA4685
4695	AAG08083.1	hypothetical protein PA4697
4697	AAG08085.1	hypothetical protein PA4699
4700	AAG08088.1	hypothetical protein PA4702
4710	AAG08098.1	hypothetical protein PA4712
4711	AAG08099.1	hypothetical protein PA4713
4734	AAG08122.1	hypothetical protein PA4736
4735	AAG08123.1	hypothetical protein PA4737
4752	AAG08140.1	hypothetical protein PA4754
4773	AAG08161.1	hypothetical protein PA4775
4780	AAG08168.1	hypothetical protein PA4782
4789	AAG08177.1	hypothetical protein PA4791
4792	AAG08180.1	hypothetical protein PA4794
4799	AAG08187.1	hypothetical protein PA4801
4801	AAG08189.1	hypothetical protein PA4803
4814	AAG08201.1	hypothetical protein PA4816
4818	AAG08205.1	hypothetical protein PA4820
4821	AAG08208.1	hypothetical protein PA4823
4822	AAG08209.1	hypothetical protein PA4824

4824	AAG08211.1	hypothetical protein PA4826
4826	AAG08213.1	conserved hypothetical protein
4847	AAG08234.1	hypothetical protein PA4849
4867	AAG08254.1	hypothetical protein PA4869
4869	AAG08256.1	hypothetical protein PA4871
4879	AAG08266.1	hypothetical protein PA4881
4881	AAG08268.1	hypothetical protein PA4883
4882	AAG08269.1	hypothetical protein PA4884
4915	AAG08302.1	hypothetical protein PA4917
4931	AAG08318.1	hypothetical protein PA4933
4938	AAG08325.1	conserved hypothetical protein
4953	AAG08340.1	hypothetical protein PA4955
4959	AAG08346.1	hypothetical protein PA4961
4961	AAG08348.1	hypothetical protein PA4963
4964	AAG08351.1	hypothetical protein PA4966
4989	AAG08376.1	hypothetical protein PA4991
4991	AAG08378.1	hypothetical protein PA4993
4996	AAG08383.1	conserved hypothetical protein
4997	AAG08384.1	O-antigen ligase, WaaL
4999	AAG08386.1	hypothetical protein PA5001
5000	AAG08387.1	hypothetical protein PA5002
5004	AAG08391.1	hypothetical protein PA5006
5006	AAG08393.1	hypothetical protein PA5008
5025	AAG08412.1	hypothetical protein PA5027
5031	AAG08418.1	hypothetical protein PA5033
5071	AAG08458.1	hypothetical protein PA5073
5079	AAG08466.1	hypothetical protein PA5081
5084	AAG08471.1	hypothetical protein PA5086
5085	AAG08472.1	hypothetical protein PA5087
5099	AAG08486.1	hypothetical protein PA5101
5100	AAG08487.1	hypothetical protein PA5102
5106	AAG08493.1	hypothetical protein PA5108
5118	AAG08505.1	hypothetical protein PA5120
5124	AAG08511.1	hypothetical protein PA5126
5134	AAG08521.1	hypothetical protein PA5136
5142	AAG08529.1	hypothetical protein PA5144
5189	AAG08576.1	hypothetical protein PA5191
5218	AAG08605.1	hypothetical protein PA5220
5262	AAG08649.1	hypothetical protein PA5264

5263	AAG08650.1	hypothetical protein PA5265
5267	AAG08654.1	hypothetical protein PA5269
5271	AAG08658.1	hypothetical protein PA5273
5274	AAG08661.1	Lipopeptide LppL precursor
5283	AAG08670.1	hypothetical protein PA5285
5290	AAG08677.1	phosphoryl choline phosphatase
5316	AAG08703.1	hypothetical protein PA5318
5324	AAG08711.1	hypothetical protein PA5326
5328	AAG08715.1	hypothetical protein PA5330
5338	AAG08725.1	hypothetical protein PA5340
5345	AAG08732.1	hypothetical protein PA5347
5379	AAG08766.1	hypothetical protein PA5381
5395	AAG08782.1	hypothetical protein PA5397
5400	AAG08787.1	hypothetical protein PA5402
5402	AAG08789.1	hypothetical protein PA5404
5403	AAG08790.1	hypothetical protein PA5405
5404	AAG08791.1	hypothetical protein PA5406
5405	AAG08792.1	hypothetical protein PA5407
5406	AAG08793.1	hypothetical protein PA5408
5412	AAG08799.1	hypothetical protein PA5414
5417	AAG08804.1	sarcosine oxidase gamma subunit
5444	AAG08831.1	hypothetical protein PA5446
5453	AAG08840.1	hypothetical protein PA5455
5454	AAG08841.1	hypothetical protein PA5456
5458	AAG08845.1	hypothetical protein PA5460
5470	AAG08857.1	hypothetical protein PA5472
5475	AAG08862.1	hypothetical protein PA5477
5478	AAG08865.1	hypothetical protein PA5480
5492	AAG08879.1	hypothetical protein PA5494
5494	AAG08881.1	hypothetical protein PA5496
5500	AAG08887.1	hypothetical protein PA5502
5513	AAG08900.1	hypothetical protein PA5515
5524	AAG08911.1	hypothetical protein PA5526
5531	AAG08918.1	hypothetical protein PA5533
5532	AAG08919.1	hypothetical protein PA5534
5535	AAG08922.1	hypothetical protein PA5537
5564	AAG08951.1	hypothetical protein PA5566

Table 13.5:

Type of similar P. profile	NO. of hits for specific protein	Accession No.	Name of Protein
48	73	AAG03463.1	Probable ATP-binding component of ABC transporter
48	136	AAG03526.1	Probable ATP-binding component of ABC transporter
48	140	AAG03530.1	Alkyl hydroperoxide reductase subunit F
48	184	AAG03574.1	Probable ATP-binding component of ABC transporter
48	207	AAG03595.1	Probable ATP-binding component of ABC transporter
48	281	AAG03669.1	Sulfate transport protein CysA
48	303	AAG03691.1	Polyamine transport protein PotG
48	327	AAG03715.1	Probable ATP-binding component of ABC transporter
48	375	AAG03763.1	Cell division protein FtsE
48	604	AAG03992.1	Probable ATP-binding component of ABC transporter
48	851	AAG04238.1	Thioredoxin reductase 2
48	862	AAG04249.1	Probable ATP-binding/permease fusion ABC transporter
48	894	AAG04281.1	Arginine/ornithine transport protein AotP
48	1072	AAG04459.1	Branched-chain amino acid transport protein BraG
48	1073	AAG04460.1	Branched-chain amino acid transport protein BraF
48	1115	AAG04502.1	Probable ATP-binding/permease fusion ABC transporter
48	1248	AAG04635.1	Alkaline protease secretion protein AprD
48	1258	AAG04645.1	Probable ATP-binding component of ABC transporter
48	1341	AAG04728.1	Probable ATP-binding component of ABC transporter
48	1477	AAG04864.1	heme exporter protein CcmA
48	1809	AAG05196.1	Probable ATP-binding component of ABC transporter
48	1863	AAG05250.1	Molybdenum transport protein ModC
48	1878	AAG05265.1	Probable ATP-binding/permease fusion ABC transporter
48	1949	AAG05335.1	ribose transport protein RbsA
48	1966	AAG05352.1	Probable ATP-binding component of ABC transporter
48	2063	AAG05449.1	Probable ATP-binding component of ABC transporter
48	2296	AAG05682.1	probable ATP-binding component of ABC transporter
48	2310	AAG05696.1	Probable ATP-binding component of ABC transporter
48	2331	AAG05717.1	Probable ATP-binding component of ABC transporter
48	2343	AAG05729.1	probable ATP-binding component of ABC maltose/mannitol transporter
48	2352	AAG05738.1	Probable ATP-binding component of ABC transporter
48	2392	AAG05778.1	Probable ATP-binding/permease fusion ABC transporter
48	2399	AAG05785.1	Pyoverdine biosynthesis protein PvdE
48	2409	AAG05796.1	Probable ATP-binding component of ABC transporter
48	2813	AAG06200.1	Probable ATP-binding component of ABC transporter
48	2858	AAG06245.1	Probable ATP-binding component of ABC transporter
48	2913	AAG06300.1	Probable ATP-binding component of ABC transporter
48	2927	AAG06314.1	Histidine transport protein HisP
48	2988	AAG06375.1	Probable ATP-binding component of ABC transporter
48	3012	AAG06399.1	
48	3020	AAG06407.1	Probable ATP-binding component of ABC transporter
48	3188	AAG06575.1	Probable ATP-binding component of ABC transporter
48	3213	AAG06600.1	Probable ATP-binding component of ABC transporter

48	3229	AAG06616.1	ProbableATP-binding/permease fusion ABC transporter
48	3255	AAG06642.1	probable ATP-bindingcomponent of ABC transporter
48	3315	AAG06702.1	probable ATP-bindingcomponent of ABC transporter
48	3376	AAG06763.1	probable ATP-bindingcomponent of ABC transporter
48	3377	AAG06764.1	probable ATP-bindingcomponent of ABC transporter
48	3385	AAG06772.1	ATP-binding component ofABC phosphonate transporter
48	3395	AAG06782.1	NosF protein
48	3407	AAG06794.1	Transport protein HasD
48	3443	AAG06830.1	probable ATP-bindingcomponent of ABC transporter
48	3448	AAG06835.1	probable ATP-bindingcomponent of ABC transporter
48	3515	AAG06902.1	probable ATP-bindingcomponent of ABC transporter
48	3539	AAG06926.1	probable ATP-bindingcomponent of ABC transporter
48	3608	AAG06995.1	polyamine transport proteinPotA
48	3673	AAG07060.1	probable ATP-bindingcomponent of ABC transporter
48	3838	AAG07225.1	probable ATP-bindingcomponent of ABC transporter
48	3891	AAG07278.1	probable ATP-bindingcomponent of ABC transporter
48	3937	AAG07324.1	probable ATP-bindingcomponent of ABC taurine transporter
48	4037	AAG07424.1	probable ATP-bindingcomponent of ABC transporter
48	4064	AAG07451.1	probable ATP-bindingcomponent of ABC transporter
48	4143	AAG07530.1	probable toxin transporter
48	4158	AAG07545.1	Ferric enterobactintransport protein FepC
48	4192	AAG07579.1	probable ATP-bindingcomponent of ABC transporter
48	4222	AAG07610.1	probable ATP-bindingcomponent of ABC transporter
48	4223	AAG07611.1	probable ATP-bindingcomponent of ABC transporter
48	4456	AAG07844.1	probable ATP-bindingcomponent of ABC transporter
48	4461	AAG07849.1	probable ATP-bindingcomponent of ABC transporter
48	4505	AAG07893.1	probable ATP-bindingcomponent of ABC transporter
48	4506	AAG07894.1	probable ATP-bindingcomponent of ABC dipeptide transporter
48	4594	AAG07982.1	probable ATP-bindingcomponent of ABC transporter
48	4595	AAG07983.1	probable ATP-bindingcomponent of ABC transporter
48	4704	AAG08092.1	probable ATP-bindingcomponent of ABC transporter
48	4859	AAG08246.1	probable ATP-bindingcomponent of ABC transporter
48	4860	AAG08247.1	probable ATP-bindingcomponent of ABC transporter
48	4907	AAG08294.1	probable ATP-bindingcomponent of ABC transporter
48	4908	AAG08295.1	probable ATP-bindingcomponent of ABC transporter
48	4995	AAG08382.1	Transport protein MsbA
48	5072	AAG08459.1	probable ATP-bindingcomponent of ABC transporter
48	5092	AAG08479.1	probable ATP-bindingcomponent of ABC transporter
48	5150	AAG08537.1	probable ATP-bindingcomponent of ABC transporter
48	5229	AAG08616.1	probableATP-binding/permease fusion ABC transporter
48	5364	AAG08751.1	ATP-binding component ofABC phosphate transporter
48	5374	AAG08761.1	probable ATP-bindingcomponent of ABC transporter
48	5498	AAG08885.1	zinc transport protein ZnuC
48	5501	AAG08888.1	probable ATP-bindingcomponent of ABC transporter
47	374	AAG03762.1	Signal recognition particle receptor FtsY
47	593	AAG03981.1	rRNA(adenine-N6,N6)-dimethyltransferase
47	669	AAG04057.1	tyrosyl-tRNA synthetase 2

47	768	AAG04156.1	GTP-binding protein LepA	
47	772	AAG04160.1	GTP-binding protein Era	
47	905	AAG04292.1	alanyl-tRNA synthetase	
47	965	AAG04352.1	aspartyl-tRNA synthetase	
47	1106	AAG04493.1	flagellum-specific ATP synthase FliI	
47	1388	AAG04775.1	probable ATP-binding component of ABC transporter	
47	1699	AAG05086.1	ATP synthase in type III secretion system	
47	2073	AAG05459.1	elongation factor G	
47	2617	AAG06004.1	thioredoxin reductase 1	
47	2741	AAG06128.1	phenylalanyl-tRNA synthetase, alpha-subunit	
47	2749	AAG06136.1	probable methionine aminopeptidase	
47	3483	AAG06870.1	methionyl-tRNA synthetase	
47	3658	AAG07045.1	methionine aminopeptidase	
47	3701	AAG07088.1	lysyl-tRNA synthetase	
47	3746	AAG07133.1	signal recognition particle protein Ffh	
47	3802	AAG07189.1	histidyl-tRNA synthetase	
47	3834	AAG07221.1	valyl-tRNA synthetase	
47	3903	AAG07290.1	peptide chain release factor 3	
47	3987	AAG07374.1	leucyl-tRNA synthetase	
47	4240	AAG07628.1	30S ribosomal protein S11	
47	4241	AAG07629.1	30S ribosomal protein S13	
47	4243	AAG07631.1	secretion protein SecY	
47	4251	AAG07639.1	50S ribosomal protein L5	
47	4253	AAG07641.1	50S ribosomal protein L14	
47	4257	AAG07645.1	30S ribosomal protein S3	
47	4260	AAG07648.1	50S ribosomal protein L2	
47	4265	AAG07653.1	elongation factor Tu	
47	4266	AAG07654.1	elongation factor G	
47	4269	AAG07657.1	DNA-directed RNA polymerase beta* chain	
47	4270	AAG07658.1	DNA-directed RNA polymerase beta chain	
47	4273	AAG07661.1	50S ribosomal protein L1	
47	4274	AAG07662.1	50S ribosomal protein L11	
47	4277	AAG07665.1	elongation factor Tu	
47	4385	AAG07773.1	GroEL protein	
47	4442	AAG07830.1	ATP sulfurylase GTP-binding subunit/APS kinase	
47	4560	AAG07948.1	isoleucyl-tRNA synthetase	
47	4566	AAG07954.1	GTP-binding protein Obg	
47	4672	AAG08060.1	conserved hypothetical protein	
47	4742	AAG08130.1	translation initiation factor IF-2	
47	4805	AAG08193.1	selenocysteine-specific elongation factor	
47	5115	AAG08502.1	regulatory protein TypA	
47	5250	AAG08637.1	probable ATP-binding component of ABC transporter	
47	5448	AAG08835.1	ABC subunit of A-band LP Sefflux transporter	
47	5552	AAG08939.1	ATP synthase beta chain	
47	5554	AAG08941.1	ATP synthase alpha chain	
46	2	AAG03392.1	DNA polymerase III, beta chain	
46	4	AAG03394.1	DNA gyrase subunit B	

46	581	AAG03969.1	O-sialoglycoprotein endopeptidase		
46	658	AAG04046.1	probable ATPase		
46	670	AAG04058.1	probable DNA polymerase alpha chain		
46	1534	AAG04921.1	DNA polymerase subunit gamma and tau		
46	1797	AAG05184.1	cysteinyl-tRNA synthetase		
46	2064	AAG05450.1	probable pyridoxal-phosphate dependent enzyme		
46	2445	AAG05832.1	serine hydroxymethyltransferase		
46	2613	AAG06000.1	seryl-tRNA synthetase		
46	2745	AAG06132.1	threonyl-tRNA synthetase		
46	2960	AAG06347.1	conserved hypothetical protein		
46	3169	AAG06556.1	DNA gyrase subunit A		
46	3565	AAG06952.1	conserved hypothetical protein		
46	3636	AAG07023.1	enolase		
46	3641	AAG07028.1	DNA polymerase III, alpha chain		
46	3653	AAG07040.1	undecaprenyl pyrophosphate synthetase		
46	3657	AAG07044.1	30S ribosomal protein S2		
46	3668	AAG07055.1	probable pyridoxal-phosphate dependent enzyme		
46	3687	AAG07074.1	adenylate kinase		
46	3799	AAG07186.1	conserved hypothetical protein		
46	3814	AAG07201.1	L-cysteine desulfurase (pyridoxal phosphate-dependent)		
46	4061	AAG07448.1	probable thioredoxin		
46	4234	AAG07622.1	excinuclease ABC subunit A		
46	4246	AAG07634.1	30S ribosomal protein S5		
46	4248	AAG07636.1	50S ribosomal protein L6		
46	4249	AAG07637.1	30S ribosomal protein S8		
46	4263	AAG07651.1	50S ribosomal protein L3		
46	4267	AAG07655.1	30S ribosomal protein S7		
46	4407	AAG07795.1	cell division protein FtsZ		
46	4432	AAG07820.1	30S ribosomal protein S9		
46	4433	AAG07821.1	50S ribosomal protein L13		
46	4602	AAG07990.1	serine hydroxymethyltransferase		
46	4749	AAG08137.1	cell division protein FtsH		
46	4934	AAG08321.1	probable rRNA methylase		
46	4962	AAG08349.1	topoisomerase IV subunit A		
46	4965	AAG08352.1	topoisomerase IV subunit B		
46	5238	AAG08625.1	thioredoxin		
46	5413	AAG08800.1	serine hydroxymethyltransferase		
45	1	AAG03391.1	chromosomal replication initiator protein DnaA		
45	18	AAG03408.1	methionyl-tRNA formyltransferase		
45	90	AAG03480.1	probable ClpA/B-type chaperone		
45	460	AAG03848.1	probable ClpA/B protease ATP binding subunit		
45	1158	AAG04545.1	ribonucleoside reductase, large chain		
45	1531	AAG04918.1	DNA ligase		
45	1589	AAG04976.1	lipoamide dehydrogenase-glc		
45	1664	AAG05051.1	probable ClpA/B-type protease		
45	1772	AAG05159.1	phosphoenolpyruvate synthase		
45	1798	AAG05185.1	10-methylene-tetrahydrofolate dehydrogenase		

45	2027	AAG05413.1	glutathione reductase		
45	2252	AAG05638.1	lipoamide dehydrogenase-Val		
45	2373	AAG05759.1	probable ClpA/B-type protease		
45	2586	AAG05973.1	excinuclease ABC subunit C		
45	2621	AAG06008.1	ATP-binding protease component ClpA		
45	2726	AAG06113.1	probable chaperone		
45	2740	AAG06127.1	phenylalanyl-tRNA synthetase, beta subunit		
45	2852	AAG06239.1	translation elongation factor P		
45	2867	AAG06254.1	secretion protein MttC		
45	3135	AAG06522.1	glutamyl-tRNA synthetase		
45	3139	AAG06526.1	excinuclease ABC subunit B		
45	3366	AAG06753.1	probable chaperone		
45	3618	AAG07005.1	RecA protein		
45	3654	AAG07041.1	ribosome recycling factor		
45	3655	AAG07042.1	uridylate kinase		
45	3656	AAG07043.1	elongation factor Ts		
45	3742	AAG07129.1	50S ribosomal protein L19		
45	3743	AAG07130.1	tRNA(guanine-N1)-methyltransferase		
45	3810	AAG07197.1	heat shock protein HscA		
45	4138	AAG07525.1	tyrosyl-tRNA synthetase		
45	4238	AAG07626.1	DNA-directed RNA polymerase alpha chain		
45	4239	AAG07627.1	30S ribosomal protein S4		
45	4256	AAG07644.1	50S ribosomal protein L16		
45	4258	AAG07646.1	50S ribosomal protein L22		
45	4262	AAG07650.1	50S ribosomal protein L4		
45	4264	AAG07652.1	30S ribosomal protein S10		
45	4268	AAG07656.1	30S ribosomal protein S12		
45	4420	AAG07808.1	conserved hypothetical protein		
45	4422	AAG07810.1	conserved hypothetical protein		
45	4439	AAG07827.1	tryptophanyl-tRNA synthetase		
45	4542	AAG07930.1	ClpB protein		
45	4567	AAG07955.1	50S ribosomal protein L27		
45	4568	AAG07956.1	50S ribosomal protein L21		
45	4569	AAG07957.1	octaprenyl-diphosphate synthase		
45	4664	AAG08052.1	peptide chain release factor 1		
45	4669	AAG08057.1	ribose-phosphate pyrophosphokinase		
45	4671	AAG08059.1	peptidyl-tRNA hydrolase		
45	4722	AAG08110.1	probable aminoacyl-transfer RNA synthetase (class I)		
45	4743	AAG08131.1	N utilization substance protein A		
45	4758	AAG08146.1	DnaJ protein		
45	4759	AAG08147.1	DnaK protein		
45	4929	AAG08316.1	replicative DNA helicase		
45	5294	AAG08681.1	ATP-dependent DNA helicase Rep		
45	5441	AAG08828.1	DNA helicase II		
45	5491	AAG08878.1	DNA polymerase I		
45	5565	AAG08952.1	conserved hypothetical protein		
44	19	AAG03409.1	polypeptide deformylase		

44	325	AAG03713.1	probable permease of ABC transporter		
44	377	AAG03765.1	sigma factor RpoH		
44	429	AAG03817.1	probable ATP-dependent RNA helicase		
44	456	AAG03844.1	RNA helicase DbpA		
44	553	AAG03941.1	phosphoglycerate kinase		
44	577	AAG03965.1	sigma factor RpoD		
44	578	AAG03966.1	DNA primase		
44	607	AAG03995.1	probable permease of ABC transporter		
44	608	AAG03996.1	ribulose-phosphate 3-epimerase		
44	958	AAG04345.1	prolyl-tRNA synthetase		
44	1427	AAG04814.1	probable ATP-binding component of ABC transporter		
44	1431	AAG04818.1	probable cation-transporting P-type ATPase		
44	1551	AAG04938.1	probable cation-transporting P-type ATPase		
44	1616	AAG05003.1	glycerol-3-phosphate dehydrogenase, biosynthetic		
44	1636	AAG05023.1	potassium-transporting ATPase, B chain		
44	1802	AAG05189.1	trigger factor		
44	2436	AAG05823.1	probable cation-transporting P-type ATPase		
44	2620	AAG06007.1	initiation factor		
44	2742	AAG06129.1	50S ribosomal protein L20		
44	2744	AAG06131.1	translation initiation factor IF-3		
44	2799	AAG06186.1	probable two-component response regulator		
44	2841	AAG06228.1	probable ATP-dependent RNA helicase		
44	2992	AAG06379.1	soluble pyridine nucleotide transhydrogenase		
44	3003	AAG06390.1	transcription-repair coupling protein Mfd		
44	3078	AAG06465.1	probable two-component response regulator		
44	3205	AAG06592.1	probable two-component response regulator		
44	3467	AAG06854.1	probable ATP-dependent RNA helicase		
44	3496	AAG06883.1	endonuclease III		
44	3555	AAG06942.1	conserved hypothetical protein		
44	3623	AAG07010.1	sigma factor RpoS		
44	3691	AAG07078.1	probable metal-transporting P-type ATPase		
44	3861	AAG07248.1	ATP-dependent RNA helicase RhlB		
44	3920	AAG07307.1	probable metal transporting P-type ATPase		
44	3950	AAG07337.1	probable ATP-dependent RNA helicase		
44	4043	AAG07430.1	geranyltranstransferase		
44	4068	AAG07455.1	probable epimerase		
44	4237	AAG07625.1	50S ribosomal protein L17		
44	4244	AAG07632.1	50S ribosomal protein L15		
44	4247	AAG07635.1	50S ribosomal protein L18		
44	4254	AAG07642.1	30S ribosomal protein S17		
44	4259	AAG07647.1	30S ribosomal protein S19		
44	4271	AAG07659.1	50S ribosomal protein L7 /L12		
44	4403	AAG07791.1	secretion protein SecA		
44	4415	AAG07803.1	phospho-N-acetylmuramoyl-pentapeptide-transferase		
44	4498	AAG07886.1	probable metalloproteinase		
44	4738	AAG08126.1	polyribonucleotide nucleotidyltransferase		
44	4739	AAG08127.1	30S ribosomal protein S15		

44	4746	AAG08134.1	triosephosphate isomerase	
44	4747	AAG08135.1	phosphoglucosamine mutase	
44	4766	AAG08154.1	SmpB protein	
44	4823	AAG08210.1	Mg(2+) transport ATPase, P-type 2	
44	4871	AAG08258.1	probable heat-shock protein	
44	5049	AAG08436.1	arginyl-tRNA synthetase	
44	5198	AAG08585.1	two-component response regulator OmpR	
44	5358	AAG08745.1	two-component response regulator PhoB	
44	5362	AAG08749.1	probable two-component response regulator	
44	5468	AAG08855.1	probable peptide chain release factor	
44	5562	AAG08949.1	glucose inhibited division protein B	
43	182	AAG03572.1	probable short-chain dehydrogenase	
43	342	AAG03730.1	prolipoprotein diacylglycerol transferase	
43	388	AAG03776.1	conserved hypothetical protein	
43	410	AAG03798.1	twitching motility protein PilH	
43	495	AAG03883.1	probable acyl-CoA carboxylase subunit	
43	547	AAG03935.1	methionine adenosyltransferase	
43	549	AAG03937.1	transketolase	
43	552	AAG03940.1	D-erythrose 4-phosphate dehydrogenase	
43	659	AAG04047.1	probable short-chain dehydrogenase	
43	757	AAG04145.1	probable two-component response regulator	
43	767	AAG04155.1	serine protease MucD precursor	
43	771	AAG04159.1	ribonuclease III	
43	966	AAG04353.1	conserved hypothetical protein	
43	969	AAG04356.1	Holliday junction DNA helicase RuvB	
43	1025	AAG04412.1	probable short-chain dehydrogenase	
43	1100	AAG04487.1	two-component sensor	
43	1101	AAG04488.1	two-component response regulator	
43	1181	AAG04568.1	two-component response regulator PhoP	
43	1346	AAG04733.1	probable short-chain dehydrogenase	
43	1381	AAG04768.1	probable short-chain dehydrogenase	
43	1386	AAG04773.1	UDP-glucose 4-epimerase	
43	1402	AAG04789.1	probable pyruvate carboxylase	
43	1439	AAG04826.1	probable two-component response regulator	
43	1472	AAG04859.1	probable short-chain dehydrogenase	
43	1539	AAG04926.1	probable short-chain dehydrogenase	
43	1639	AAG05026.1	two-component response regulator KdpE	
43	1651	AAG05038.1	probable short-chain dehydrogenase	
43	1803	AAG05190.1	ATP-dependent Clp protease proteolytic subunit	
43	1804	AAG05191.1	ATP-dependent Clp protease ATP-binding subunit ClpX	
43	1829	AAG05216.1	probable short-chain dehydrogenase	
43	1830	AAG05217.1	probable short-chain dehydrogenase	
43	2005	AAG05391.1	3-hydroxybutyrate dehydrogenase	
43	2014	AAG05400.1	alpha subunit of geranyl-CoA carboxylase, GnyA	
43	2144	AAG05530.1	probable short-chain dehydrogenase	
43	2274	AAG05660.1	penicillin-binding protein 3A	
43	2480	AAG05867.1	probable two-component response regulator	

43	2516	AAG05903.1	cis-1,2-dihydroxycyclohexa-3,4-diene carboxylate dehydrogenase
43	2524	AAG05911.1	probable two-component response regulator
43	2555	AAG05942.1	probable short-chain dehydrogenase
43	2558	AAG05945.1	probable AMP-binding enzyme
43	2627	AAG06014.1	tRNA methyltransferase
43	2658	AAG06045.1	probable two-component response regulator
43	2829	AAG06216.1	probable aminotransferase
43	2888	AAG06275.1	probable short-chain dehydrogenase
43	2892	AAG06279.1	probable biotin carboxylase/biotin carboxyl carrier protein
43	2962	AAG06349.1	DNA polymerase III, delta prime subunit
43	2968	AAG06355.1	3-oxoacyl-[acyl-carrier-protein] reductase
43	3002	AAG06389.1	probable glyceraldehyde-3-phosphate dehydrogenase
43	3107	AAG06494.1	probable short-chain dehydrogenase
43	3109	AAG06496.1	midophosphoribosyltransferase
43	3115	AAG06502.1	tRNA-pseudouridine synthase I
43	3129	AAG06516.1	probable short-chain dehydrogenase
43	3163	AAG06550.1	30S ribosomal protein S1
43	3174	AAG06561.1	probable short-chain dehydrogenase
43	3196	AAG06583.1	glyceraldehyde 3-phosphate dehydrogenase
43	3278	AAG06665.1	probable short-chain dehydrogenase
43	3300	AAG06687.1	long-chain-fatty-acid--CoA ligase
43	3301	AAG06688.1	long-chain-fatty-acid--CoA ligase
43	3327	AAG06714.1	probable Clp-family ATP-dependent protease
43	3331	AAG06718.1	probable short chain dehydrogenase
43	3388	AAG06775.1	beta-ketoacyl reductase
43	3428	AAG06815.1	probable short-chain dehydrogenases
43	3508	AAG06895.1	probable short-chain dehydrogenase
43	3512	AAG06899.1	probable short-chain dehydrogenase
43	3538	AAG06925.1	ornithine carbamoyltransferase, anabolic
43	3563	AAG06950.1	probable phosphotransferase system enzyme I
43	3610	AAG06997.1	polyamine transport protein PotC
43	3638	AAG07025.1	CTP synthase
43	3643	AAG07030.1	ribonuclease HII
43	3660	AAG07047.1	probable aminotransferase
43	3798	AAG07185.1	probable aminotransferase
43	3807	AAG07194.1	nucleoside diphosphate kinase
43	3860	AAG07247.1	probable AMP-binding enzyme
43	3883	AAG07270.1	probable short-chain dehydrogenase
43	3924	AAG07311.1	probable medium-chain acyl-CoA ligase
43	3957	AAG07344.1	probable short-chain dehydrogenase
43	4002	AAG07389.1	rod shape-determining protein
43	4003	AAG07390.1	penicillin-binding protein 2
43	4020	AAG07407.1	UDP-N-acetylmuramate:L-alanyl-gamma-D-glutamyl-meso-diaminopimelate
43	4044	AAG07431.1	1-deoxyxylulose-5-phosphate synthase
43	4089	AAG07476.1	probable short-chain dehydrogenase
43	4098	AAG07485.1	probable short-chain dehydrogenase
43	4148	AAG07535.1	probable short-chain dehydrogenase

43	4162	AAG07549.1	probable short-chain dehydrogenase
43	4201	AAG07588.1	D-alanine-D-alanine ligaseA
43	4272	AAG07660.1	50S ribosomal protein L10
43	4275	AAG07663.1	transcription antitermination protein NusG
43	4280	AAG07668.1	BirA bifunctional protein
43	4381	AAG07769.1	probable two-component response regulator
43	4389	AAG07777.1	arginine decarboxylase(ADC)
43	4411	AAG07799.1	UDP-N-acetylmuramate--alanine ligase
43	4412	AAG07800.1	UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide)pyrophosphor
43	4413	AAG07801.1	cell division protein FtsW
43	4414	AAG07802.1	UDP-N-acetylmuramoylalanine--D-glutamate ligase
43	4417	AAG07805.1	UDP-N-acetylmuramoylalanine-D-glutamate-2, 6-diaminopimelate ligase
43	4418	AAG07806.1	penicillin-binding protein3
43	4446	AAG07834.1	AlgW protein
43	4450	AAG07838.1	UDP-N-acetylglucosamine1-carboxyvinyltransferase
43	4547	AAG07935.1	two-component response regulator PilR
43	4713	AAG08101.1	probable aminotransferase
43	4720	AAG08108.1	probable aminotransferase
43	4724	AAG08112.1	two-component response regulator CbrB
43	4725	AAG08113.1	poly(A) polymerase
43	4753	AAG08141.1	transcription elongation factor GreA
43	4779	AAG08167.1	probable two-component response regulator
43	4830	AAG08217.1	probable short-chain dehydrogenase
43	4846	AAG08233.1	biotin carboxylase
43	4850	AAG08237.1	conserved hypothetical protein
43	4905	AAG08292.1	probable short-chain dehydrogenase
43	4930	AAG08317.1	50S ribosomal protein L9
43	4943	AAG08330.1	delta2-isopentenylpyrophosphate transferase
43	4974	AAG08361.1	aspartate transaminase
43	5048	AAG08435.1	primosomal protein N'
43	5052	AAG08439.1	heat shock protein HslU
43	5145	AAG08532.1	A / G specific adenine glycosylase
43	5148	AAG08535.1	probable short-chain dehydrogenase
43	5159	AAG08546.1	dTDP-D-glucose4,6-dehydratase
43	5170	AAG08557.1	ornithine carbamoyltransferase, catabolic
43	5320	AAG08707.1	phosphomannomutase AlgC
43	5343	AAG08730.1	ATP-dependent DNA helicase RecG
43	5434	AAG08821.1	probable biotin carboxylase subunit of a transcarboxylase
43	5481	AAG08868.1	two-component response regulator AlgB
43	5482	AAG08869.1	probable two-component sensor
43	5495	AAG08882.1	hypothetical protein PA5497
43	5509	AAG08896.1	probable two-component response regulator
43	5519	AAG08906.1	probable short-chain dehydrogenase
43	5522	AAG08909.1	probable short-chain dehydrogenase
43	5566	AAG08953.1	conserved hypothetical protein
42	25	AAG03415.1	shikimate dehydrogenase
42	179	AAG03569.1	probable two-component response regulator

42	205	AAG03593.1	probable permease of ABC transporter
42	224	AAG03612.1	probable dihydrodipicolinate synthetase
42	283	AAG03671.1	hypothetical protein PA0244
42	332	AAG03720.1	probable major facilitator superfamily (MFS) transporter
42	338	AAG03726.1	sulfate transport protein CysW
42	379	AAG03767.1	sulfate transport protein CysT
42	383	AAG03771.1	probable glutamine synthetase
42	403	AAG03791.1	polyamine transport protein PotH
42	409	AAG03797.1	D-3-phosphoglycerate dehydrogenase
42	414	AAG03802.1	threonine dehydratase, biosynthetic
42	930	AAG04317.1	phosphoenolpyruvate-protein phosphotransferase PtsP
42	936	AAG04323.1	phosphopantetheine adenylyltransferase
42	1012	AAG04399.1	probable transglycosylase
42	1256	AAG04643.1	DNA mismatch repair protein MicA
42	1332	AAG04719.1	probable oxidase
42	1529	AAG04916.1	aspartate carbamoyltransferase
42	1536	AAG04923.1	twitching motility protein PilG
42	1613	AAG05000.1	still frameshift probable component of chemotactic signal transduction system
42	1783	AAG05170.1	two-component response regulator CreB
42	1800	AAG05187.1	tRNA nucleotidyltransferase
42	2025	AAG05411.1	anthranilate synthase component II
42	2537	AAG05924.1	L-aspartate oxidase
42	2616	AAG06003.1	aspartate kinase alpha and beta chain
42	2687	AAG06074.1	D-lactate dehydrogenase (fermentative)
42	2810	AAG06197.1	sensor/response regulator hybrid
42	2811	AAG06198.1	two-component response regulator
42	2825	AAG06212.1	GTP pyrophosphokinase
42	2893	AAG06280.1	Holliday junction DNA helicase RuvA
42	2976	AAG06363.1	probable coenzyme A ligase
42	3130	AAG06517.1	dihydrodipicolinate synthase
42	3164	AAG06551.1	probable two-component response regulator
42	3165	AAG06552.1	probable dihydrodipicolinate synthetase
42	3247	AAG06634.1	probable short-chain dehydrogenase
42	3347	AAG06734.1	conserved hypothetical protein
42	3463	AAG06850.1	3-oxoacyl-acyl carrier protein synthase II
42	3650	AAG07037.1	erythronate-4-phosphate dehydrogenase
42	3652	AAG07039.1	conserved hypothetical protein
42	3704	AAG07091.1	recombination protein RecR
42	3748	AAG07135.1	oxygen-independent coproporphyrinogen III oxidase
42	3760	AAG07147.1	succinate dehydrogenase (A subunit)
42	3946	AAG07333.1	beta-ketoacyl-ACP synthase I
42	3983	AAG07370.1	probable sensor/response regulator hybrid
42	4032	AAG07419.1	probable DNA methylase
42	4101	AAG07488.1	chorismate synthase
42	4112	AAG07499.1	assimilatory nitrite reductase large subunit
42	4188	AAG07575.1	probable two-component sensor
42	4252	AAG07640.1	probable two-component response regulator

42	4410	AAG07798.1	probable two-componentsensor
42	4416	AAG07804.1	probable two-componentsensor
42	4544	AAG07932.1	UTP--glucose-1-phosphateuridylyltransferase
42	4561	AAG07949.1	probable glutaminesynthetase
42	4663	AAG08051.1	probable short-chaindehydrogenase
42	4698	AAG08086.1	probable 2-hydroxyaciddehydrogenase
42	4723	AAG08111.1	pyoverdine synthetase D
42	4754	AAG08142.1	PvdL
42	4774	AAG08162.1	probable phosphatidatecytidylyltransferase
42	4784	AAG08172.1	DP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase
42	4827	AAG08214.1	cell division protein FtsK
42	4883	AAG08270.1	two-component responseregulator PfeR
42	4928	AAG08315.1	conserved hypotheticalprotein
42	4932	AAG08319.1	probable two-componentresponse regulator
42	5011	AAG08398.1	probable two-componentsensor
42	5043	AAG08430.1	probable sensor/responseregulator hybrid
42	5123	AAG08510.1	probable short-chaindehydrogenase
42	5164	AAG08551.1	probable short-chaindehydrogenase
42	5300	AAG08687.1	conserved hypotheticalprotein
42	5318	AAG08705.1	beta-ketoacyl-acyl carrierprotein synthase II
42	5334	AAG08721.1	malonyl-CoA-[acyl-carrier-protein] transacylase
42	5336	AAG08723.1	ribosomal large subunitpseudouridine synthase C
42	5359	AAG08746.1	probable two-componentsensor
42	5360	AAG08747.1	conserved hypotheticalprotein
42	5452	AAG08839.1	folylpolyglutamatesynthetase
42	5546	AAG08933.1	conserved hypotheticalprotein
42	5547	AAG08934.1	cytidylate kinase
41	245	AAG03633.1	still frameshift3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE
41	247	AAG03635.1	two-component responseregulator GltR
41	282	AAG03670.1	pseudouridine synthase RluA
41	299	AAG03687.1	probable two-componentsensor
41	304	AAG03692.1	probable short-chaindehydrogenase
41	317	AAG03705.1	probable two-componentresponse regulator
41	364	AAG03752.1	probable sensor/responseregulator hybrid
41	387	AAG03775.1	conserved hypotheticalprotein
41	464	AAG03852.1	conserved hypotheticalprotein
41	585	AAG03973.1	phosphatidatecytidylyltransferase
41	650	AAG04038.1	probable chemotaxissensor/effector fusion protein
41	762	AAG04150.1	integrase/recombinase XerD
41	906	AAG04293.1	conserved hypotheticalprotein
41	929	AAG04316.1	probable aminotransferase
41	931	AAG04318.1	probablephosphotransferase protein
41	968	AAG04355.1	GMP synthase
41	998	AAG04385.1	secretion protein SecF
41	1159	AAG04546.1	secretion protein SecD
41	1333	AAG04720.1	probable 2-hydroxyaciddehydrogenase
41	1375	AAG04762.1	probable two-componentsensor

41	1377	AAG04764.1	Lost Adherence Sensor, LadS
41	1548	AAG04935.1	conserved hypotheticalprotein
41	1585	AAG04972.1	probable two-componentresponse regulator
41	1611	AAG04998.1	probable two-componentsensor
41	1680	AAG05067.1	probable two-componentresponse regulator
41	1683	AAG05070.1	probable sensor/responseregulator hybrid
41	1801	AAG05188.1	conserved hypotheticalprotein
41	1978	AAG05364.1	pyochelin biosynthesisprotein PchD
41	1994	AAG05380.1	50S ribosomal protein L24
41	2042	AAG05428.1	superoxide dismutase
41	2101	AAG05487.1	GroES protein
41	2265	AAG05651.1	D-alanine--D-alanine ligase
41	2401	AAG05787.1	UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopimelate--D-alanyl-D
41	2425	AAG05812.1	superoxide dismutase
41	2585	AAG05972.1	rod shape-determiningprotein MreB
41	2796	AAG06183.1	pseudouridine synthase
41	2919	AAG06306.1	two-component sensor PilS
41	2964	AAG06351.1	riboflavin kinase/FADsynthase
41	2966	AAG06353.1	DNA repair protein RadA
41	2969	AAG06356.1	glycerate dehydrogenase
41	3045	AAG06432.1	probable methyl transferase
41	3089	AAG06476.1	penicillin-binding protein1B
41	3112	AAG06499.1	two-component sensor CbrA
41	3193	AAG06580.1	tRNA pseudouridine 55synthase
41	3272	AAG06659.1	carbamoylphosphatesynthetase large subunit
41	3325	AAG06712.1	PmrA: two-componentregulator system response regulator PmrA
41	3639	AAG07026.1	probable short-chaindehydrogenase
41	3738	AAG07125.1	dihydrolipoamidedehydrogenase 3
41	3759	AAG07146.1	probable two-componentresponse regulator
41	3769	AAG07156.1	hosphoribosylamine--glycine ligase
41	3820	AAG07207.1	two-component responseregulator
41	3821	AAG07208.1	biosynthetic alanineracemase
41	3896	AAG07283.1	30S ribosomal protein S18
41	3974	AAG07361.1	probable two-componentsensor
41	4036	AAG07423.1	probable two-componentresponse regulator
41	4228	AAG07616.1	branched-chain amino acidtransferase
41	4366	AAG07754.1	penicillin-binding protein1A
41	4386	AAG07774.1	two-component responseregulator NtrC
41	4468	AAG07856.1	glucose-1-phosphatethymidyltransferase
41	4481	AAG07869.1	probable two-componentresponse regulator
41	4546	AAG07934.1	probable beta-ketoacylsynthase
41	4609	AAG07997.1	aminopeptidase P
41	4626	AAG08014.1	catabolic alanine racemase
41	4740	AAG08128.1	Phosphopantothenoylcysteinesynthase/(R)-4'-phospho-N-pantothenoylcys
41	4841	AAG08228.1	guanylate kinase
41	4853	AAG08240.1	uanosine-3',5'-bis(diphosphate) 3'-pyrophosphohydrolase
41	4980	AAG08367.1	two-component sensor PhoR

41	4981	AAG08368.1	conserved hypotheticalprotein	
41	5161	AAG08548.1	oxidoreductase Rmd	
41	5172	AAG08559.1	probable major facilitatorsuperfamily (MFS) transporter	
41	5222	AAG08609.1	lucosamine--fructose-6-phosphate aminotransferase	

12.5: List of microbes, which are used for creation of phylogenetic profile database -

Micro-organism name	4-letter code
<i>Acinetobacter baumannii</i>	abau
<i>Agrobacterium tumefaciens.</i>	atum
<i>Aquifex aeolicus</i>	aaeo
<i>Archeoglobus fulgidus</i>	aful
<i>Bacillus anthracis</i>	bant
<i>Bacillus cereus</i>	bcer
<i>Bacillus halodurans</i>	bhal
<i>Bacillus subtilis</i>	bsub
<i>Bacteroides fragilis</i>	bfra
<i>Bordetella pertussis</i>	bper
<i>Borrelia burgdorferi</i>	bbur
<i>Campylobacter jejuni</i>	cjej
<i>Chlamydia trachomatis</i>	ctra
<i>Clostridium perfringens</i>	cper
<i>Clostridium tetani</i>	ctet
<i>Corynebacterium diphtheriae</i>	cdip
<i>Corynebacterium glutamicum</i>	cglu
<i>Deinococcus radiodurans</i>	drad
<i>Enterococcus faecalis</i>	efae
<i>E. coli.</i>	ecolik12
<i>E. coli : w3110</i>	ecoliw
<i>Haemophilus influenzae</i>	hinf
<i>Helicobacter pylori</i>	hpyl
<i>Listeria monocytogenes</i>	lmon
<i>Methanococcus jannaschii</i>	mjan
<i>Mycobacterium avium</i>	mavi
<i>Mycobacterium bovis</i>	mbov
<i>Mycobacterium tuberculosis: cdc1551</i>	mtbcd
<i>Mycobacterium leprae</i>	mlep
<i>Mycobacterium smegmatis</i>	msme
<i>Mycoplasma genitalium.</i>	mgen
<i>Mycoplasma pneumoniae</i>	mpne
<i>Neisseria gonorrhoea</i>	ngon
<i>Neisseria meningitidis</i>	nmen
<i>Nocardia farcinica</i>	nfar
<i>Rickettsia prowazekii</i>	rpro
<i>Salmonella typhi</i>	styp
<i>Shewanella oneidensis</i>	sone
<i>Shigella flexneri</i>	sfle
<i>Staphylococcus aureus</i>	saur
<i>Streptococcus pneumoniae</i>	spne
<i>Streptococcus pyogenes</i>	spyo

<i>Streptomyces coelicolor</i>	scoe
<i>Thermos thermophilus</i>	tthe
<i>Treponema pallidum.</i>	tpal
<i>Vibrio cholerae</i>	vcho
<i>Xanthomonas campestris</i>	xcam
<i>Xylella fastidiosa</i>	xfas
<i>Yersinia pestis</i>	ypes

12.6. Perl Scripts :

1) **Formatdb.pl-**

```
#!/usr/perl/bin
use strict;
my $file="bacteria_list.txt";
open(FILE,$file) or die "Cannot open $file\n";
my @data=<FILE>;
print"hi\n";
#acinetobacter baumanii      abau
foreach my $line(@data)
{
    my $file1="";
    my $db="";
    my $dbname="";
    if($line=~/(.*)\t(.*)/)
    {
        $file1=$1;
        $db=$2;
        $dbname=$db;
        my $cmd1="formatdb -i C:/amitabh/inputfiles/$file1 -p T -n $dbname";
        system($cmd1);
        my $cmd2="blastall -p blastp -d $dbname -i
C:/amitabh/inputfiles/pseudomonas_aeruginosa.txt -o $db.out -F F";
        system($cmd2);
    }
}
```

2) **Hd_parser.pl-**

```
#!/usr/perl/bin
use strict;

my $file="bacteria_list.txt";
open(FILE,$file) or die "Cannot open $file\n";
my @data=<FILE>;
foreach my $line(@data)
{
    if($line=~/(.*)\.txt\t(\w+)/i)
    {
        my $openfile="$2.out";
```

```

        my $outfile="p_$.txt";
        my $rfh=open_to_read($openfile);
        open(OUTFILE,">$outfile");

my $counter=0;
my $flag=0;
my $qry="";
my $i=0;
my @rec="";
my $qryflag=0;
my $giflag=0;
my $eval=0;
my $E=0;
#my $p=0;
my $e=0;
my $hit="";

while(my $rec=get_next_record($rfh))
{

    @rec=split/\n/,$rec;
    $counter=0;
    foreach my $line(@rec)
    {
        if($line=~/(.*) (Query=)/)
        {
            $qryflag=1;
        }
        if($qryflag==1)
        {
            if($line=~/(Database:)/)
            {
                #print "$qry\n\n";
                print OUTFILE "$qry\n\n";
                $qry="";
                $qryflag=0;
            }
            else
            {
                $qry.=$line;
            }
        }
        if($line=~/\>gi/)
        {
            $giflag=1;

```

```

}
if($giflag==1)
{
    if($line=~Expect(\s*)(\s*)(\s*))
    {
        $eval='$';
        $hit=~s/\s{3,}/ /g;
        #print "$hit\n\n";
        if($eval=~/(d*)(e)(-)(d*)/)
        {
            $E=($1)* (10**(-$4));
            if($E<=0.0000000001)
            {
                print OUTFILE "$hit \n\n ";
            }
        }
        elseif($eval=~/(s+)(0)(.)(0)(s*)$/)
        {
            $E=($2)* (10**(-0));
            if($E==0)
            {
                print OUTFILE "$hit \n\n ";
            }
        }
        elseif($eval=~/(s*)(d+\.d+)(s*)$/)
        {
            $E=($2);
            if($E<=0.0000000001)
            {
                print OUTFILE "$hit \n\n ";
            }
        }
        $giflag=0;
        $hit="";
        $eval=0;
        $E=0;
        $e=0;
        last;
    }
}
else

```

```

        {
            $hit.=$line;
        }
    }

}
}
}
#####Subroutines#####

sub open_to_read
{
    my($file)=@_;
    my $rfh;

    unless(open ($rfh, "$file"))
    {
        print "Cannot open $file\n\n";
        exit;
    }
    return ($rfh);
}

sub get_next_record
{
    my($fh) = @_;

    my($record) = "";
    my($save_input_separator) = $/;

    $/ = "BLASTP";

    $record = <$fh>;

    $/ = $save_input_separator;
    return $record;
}

```

3) Hits.pl-

```
use strict;
my $file="bacteria_list.txt";
open(FILE,$file) or die "Cannot open $file\n";
my @data=<FILE>;
my $resultfile="result_files.txt";
open(RESULTFILES,">$resultfile");

foreach my $line(@data)
{
    if($line=~/(.*\.\txt)\t(\w+)/i)
    {
        my $openfile="p_$2.txt";
        my $outfile="result_$2.txt";
        my $rfh=open_to_read($openfile);
        open(OUTFILE,">$outfile");
        print RESULTFILES "$outfile\n";

my $flag=0;
my $gene="";
my $pflag=0;
# M. tuberculosis H37Rv|Rv0001|DnaA: 507 aa - CHROMOSOMAL REPLICATION
INITIATOR PROTEIN DNAA      (507 letters)
#      Query=      gi|9945823|gb|AAG03395.1|AE004440_5      lysophosphatidic
acidacyltransferase, LptA [Pseudomonas aeruginosa PAO1]      (257 letters)

while(my $rec=get_next_record($rfh))
{
    my @rec=split/\n/,$rec;
    $pflag=0;
    foreach my $line(@rec)
    {
        if($line=~/^>/)
        {
            if($flag==1)
            {

                if($line=~/^>/)
                {
                    print OUTFILE "$gene      1\n";
                    $flag=0;
                    $pflag=1;
                }
            }
        }
    }
}
```



```

        else
        {
            if($line=~/\gi\\d+\\|w+\\(|w*d*\\.d*)\\|ig)
            {

                $gene=$1;
                $flag=1;

            }
        }

    }

    if($pflag==0)
    {
        print OUTFILE "$gene      0\n";
    }
}
}
}

#####Subroutines#####

sub open_to_read
{
    my($file)=@_;
    my $rfh;

    unless(open ($rfh, "$file"))
    {
        print "Cannot open file\n\n";
        exit;
    }
    return ($rfh);
}

sub get_next_record
{
    my($fh) = @_;

    my($record) = "";
    my($save_input_separator) = $/;

```

```

$/ = "Query=";

$record = <$fh>;

$/ = $save_input_separator;
return $record;

}

```

4) Db_insert.pl-

```

#!/usr/bin/perl
use strict;
#use DBI;
#use DBD::mysql;
my $file="result_files.txt";
open(FILE,$file) or die "Cannot open file $file";
my @all=<FILE>;
close FILE;
my @multi_arr="";
my @all_data="";
my @tot="";
my $outfile="db_insert_file.txt";
open(OUTFILE,">$outfile");
for(my $i=1;$i<=scalar(@all);$i++)
{
    $multi_arr[$i-1]=$all[$i-1];
    if($multi_arr[$i-1]=~/(\w+)\_(\w+)/)
    {
        $tot[$i-1]=$2;
    }
    print OUTFILE "\n$multi_arr[$i-1]\n";
    open(FILE1,$multi_arr[$i-1]);
    my @data1=<FILE1>;
    my @d="";
    my $a=0;
    my $x=0;
    my $rv="";
    my @gene="";
#AAC06612.1
    for(my $j=0; $j<=scalar(@data1);$j++)
    {
        if($data1[$j]=~/(\w*\d*\.\d*)\t(\d)/i)
        {
            $rv=$1;

```

```

        $gene[$a]=$1;
        $d[$x]=$2;
        print OUTFILE "$gene[$a]  $d[$x] ";
        $a++;
        $x++;
    }

}

print "\n";

}

close OUTFILE;
my $db_file="db_insert_file.txt";
open(DBFILE,$db_file) or die "Cannot open $db_file";
my $no_of_fields=scalar(@tot);
my @dbfile=<DBFILE>;

my $dsn = 'dbi:mysql:proteome';
my $dbh = DBI->connect($dsn) or die "Can't connect to the DB: $DBI::errstr\n";
my $createtable=$dbh->prepare("create table phylobact (id int(11) primary key
AUTO_INCREMENT, gene varchar(30))");
$createtable->execute();

for(my $i=0;$i<$no_of_fields;$i++)
{
    my $altertable=$dbh->prepare("alter table phylobact add $tot[$i] int(5)");
    $altertable->execute();
}

#>M. tuberculosis H37Rv|Rv0001|DnaA: 507 aa - CHROMOSOMAL REPLICATION
INITIATOR PROTEIN DNAA
#>gi|15595199|ref|NP_064721.1| chromosomal replication initiator protein DnaA
[Pseudomonas aeruginosa PAO1]
#>gi|13699918|dbj|BAB41217.1| chromosomal replication initiator protein
[Staphylococcus aureus subsp. aureus N315]
my $query_prtn_file="p_atum.txt";
open(QUERY,$query_prtn_file);
my @query=<QUERY>;
my @genes="";
my $k=0;

foreach my $line(@query)
{
    if($line=~/^s*Query|=s*gi\\d+\\w+\\(\\w*d*\\.d*)\\w{10,11}\\(.{0,90}\\)/[i)

```

```

    {
        $genes[$k]=$1;
        #print "hi\n";
        $k++;
    }

}

for(my $m=0;$m<scalar(@genes);$m++)
{

    my $val=0;
    my $gn="";
    my $var="";
    my @temp="";
    my $s=0;
    my $p=0;
    foreach my $line(@dbfile)
    {

        if($line=~/^(\s)*$/)
        {
            next;
        }
        elsif($line=~/_result_\w*/)
        {
            next;
        }
        elsif($line=~/$genes[$m]\t(\d)\t(.*)/i)
        {
            $gn=$1;
            $val=$1;
            $temp[$s]=$val;
            #print "$gn    $temp[$s]\n";
            $s++;
            if($gn==1)
            {
                $p++;
            }
            $val="";
            $gn="";
        }
        print "$p ,($m+1)\n";
    }

    $s=0;

```

```

        my $sql="insert into phylobact values(".$m+1).",'$genes[$m]',";

for(my $a=0;$a<scalar(@temp);$a++)
{
    if($a<=scalar(@temp)-2)
    {
        $sql=$temp[$a].",";
        #$var="$temp[$a],";
    }
    elsif($a==scalar(@temp)-1)
    {
        $sql=$temp[$a];
        #$var="$temp[$a],";
    }
}

$sql.=")";
#print "hi\n";
#print "$sql\n";
#my $sql="insert into phylobact values(($m+1),'$genes[$m]','$var)";
my $sth = $dbh->prepare($sql);
print "$sql\n";
$sth->execute();
$sql="";

}

```