Comparative Genome Analysis of *Pseudomonas aeruginosa* N002 and *Enterobacter* sp. RC4 Using Bioinformatics Approaches and Their Prospectives in Soil Quality Improvement

Submitted By
Mr. Dhrubajyoti Das

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT OF THE DEGREE OF DOCTOR OF PHILOSOPHY IN BOTANY

DEPARTMENT OF BOTANY
NAGALAND UNIVERSITY, LUMAMI 798 627
NAGALAND, INDIA

2020

Contents

Chapters	Pages
Contents	2
Dedication	3
Declaration	4-5
Acknowledgement	6-7
Ph. D. Course Work Mark Sheet and Certificate	8
List of Tables	9
List of Figures	10-12
Chapter 1: Introduction	13-18
Chapter 2: Review of Literatures	19-40
Chapter 3: Materials and Methods	41-50
Chapter 4: Results	51-176
Chapter 5: Discussion	177-192
Chapter 6: Summary and Conclusions	193-196
References	197-221
List of Publications	222-222
Paper Presented in Seminars/Conferences	222-222

Dedication

I would like to dedicate this Thesis to my loving Parents, All my family members, friends, and well-wishers.

CSIR-North East Institute of Science and Technology (CSIR-NEIST)
(Formerly Regional Research Laboratory)
Jorhat-785006, Assam, India
Phone: 0376-2370121 (O); 09706550664 (M)
Fax: 0376-2370011Gram: Research
Website: http://www.neist.res.in

सीएसआईआर-
उत्तरपूर्वविज्ञानतथाप्रौद्योगिकीसंस्थान
जोरहाट-785006, असम, भारत
फ़ोन: 0376-2370121 (O); 09706550664(M)
फैक्स: 0376-2370011
वेबसाइट: http://www.neist.res.in

Dr. H. P.DekaBoruah
Sr. Principal Scientist & HoD
Biological Science & Technology Division
Email: dekabaruahhp@neist.res.in; hpdekaboruah@yahoo.com

Certificate

This is to certify that the thesis entitled "**Comparative genome analysis of *Pseudomonas aeruginosa* N002 and *Enterobacter* sp. RC4 using bioinformatics approaches and their perspectives in soil quality improvement"** is the result of research work of **Mr. Dhrubajyoti Das** carried under my Co-Supervision at Biological Science and Technology Division, CSIR-North East Institute of Science & Technology, Jorhat-785006, submitted to the Department of Botany, Nagaland University for the award of the degree of Doctor of Philosophy in Botany.

This thesis conforms to the standard of Ph. D. thesis under Nagaland University. The information collected from other sources has been duly acknowledged.

Date: 10/02/2020

(H P DekaBoruah)
Co-Supervisor

DR. H.P.DEKA BORUAH
प्रधान, बी एस टी डा
HEAD, BSTD
सी. एस. आई. आर-उत्तर पूर्व विज्ञान तथा प्रौद्योगिकी संस्थान
CSIR-North East Institute of Science & Tech
जोरहाट-785006 (असम) Jorhat-785

Nagaland University

(A Central University Estd. By the Act of Parliament No. 35 of 1989)
Lumami 798627, Nagaland, India

12th **February, 2020**

DECLARATION

I, Mr. Dhrubajyoti Das bearing Ph. D. Registration No. 686/2015 dated May 27, 2015 hereby declare that the subject matter of my thesis entitled 'Comparative Genome Analysis of *Pseudomonas aeruginosa* N002 and *Enterobacter* sp. RC4 Using Bioinformatics Approaches and Their Prospectives in Soil Quality Improvement' is the record of original work done by me, and that the contents of this thesis did not form the basis for award of any degree to me or to anybody else to the best of my knowledge. This thesis has not been submitted by me for any Research Degree in any other University/Institute.

This is further certified that the Ph. D. thesis is submitted in compliance with the UGC Regulation 2016 dated May 05, 2016 (Minimum Standard and Procedure for Award of M. Phil./Ph. D. Degree). This thesis is being submitted to the Nagaland University for the degree of 'Doctor of Philosophy in Botany'.

Dhruba

(Dhrubajyoti Das)
Candidate

Ch. Ranjan Deb 12/02/2020

(Prof. Chitta Ranjan Deb)
Supervisor
Prof. Chitta Ranjan Deb
Department of Botany
Nagaland University
Lumami-798627, Nagaland, Ind

Ch. Ranjan Deb 12/02/2020

(Head, Department of Botany)
Head I/C
Department of Botany
Nagaland University
Hqs. Lumami.

5

Acknowledgment

There are a number of people without whom this thesis would not have been a reality. I am using this opportunity to express my deepest gratitude to everyone who has supported me throughout the course of my research work by their aspiring guidance, constructive criticism and friendly advice.

It is my immense pleasure to acknowledge my research Supervisor Prof.Chitta Ranjan Deb, Department of Botany, Nagaland University, Lumami, Nagaland and my Co-Supervisor Dr. Hari Prasanna Deka Boruah, Sr. Principal Scientist, Biotechnology Group, Biological Science and Technology Division, CSIR-North East Institute of Science & Technology, Jorhat, Assam. Their constant inspiration, cordial demeanour, guidance and constructive comments helped me a lot to focus my views in proper directions. I am grateful to them for giving me the liberty to carry out my research work independently throughout the course of the Ph.D. programme. Their enthusiastic attitude, innovative ideas and deep scientific knowledge have inspired me deeply. I shall be remaining ever grateful in my life to them.

I would like to thank Dr. D. Ramaiah, Director, CSIR-NEIST, Jorhat, Assam for his permission to carry out my research work and providing me all infrastructural facilities required for my research work. I would like to express my sincere thanks to Dr. Anil Kumar Singh, Scientist, BSTD-Biotechnology Group, CSIR-NEIST, Jorhat for his constructive encouragement, valuable advices, constant support and cooperation during my research work. I am also expressing my gratefulness to Dr. Channakeshavaiah Chikkaputtaiah, Dr. Ratul Saikia for their constructive encouragement, valuable suggestion, advices and constant support throughout my research period. I would like to express my sincere thanks to all the Scientists of Biological Science and Technology

Division as well as other divisions of the Institute for their help in various stage of my Ph. D. work. I would also like to extend my heartiest appreciation to all the technical and non-technical staff members of BSTD, CSIR-NEIST, Jorhat, for their constant help and kind cooperation.Thanks are also to all research scholars of BSTD as well as other divisions of this institute, for their suggestion and cooperation.

I am also grateful to the Head, Department of Botany, Nagaland University for all the facilities for my research work. I am also thankful to faculty members and research scholars of the Department for their kind cooperation during my research tenure.

I am highly grateful to CSIR, New Delhi for financial support to smoothly carry out my Ph. D. work.

It gives me immense pleasure to thank my seniors, colleagues, and friends for their helpful hand and making the lab feel like a family. Without their support it would not have been possible to complete this task.

I sincerely acknowledge and express my deep respect and gratitude to all my teachers and well-wishers for their immense contribution to my life for whom what I am today.

Finally, I extend my gratitude and love to my loving mother, father, wife, brother, sisters and all my family members for their ever presence as a constant source of love, concern, support and strength which make me competent enough to carry out my Ph. D. work.

Finally, I must say that it would not have been possible to achieve this goal without the Grace of ALMIGHTY.

Date: 10/02/2020

Dhruba

(Dhrubajyoti Das)

Ph. D. Coursework Mark sheet and Certificate

Sl No. 159

NAGALAND UNIVERSITY

HEAD QUARTERS : LUMAMI

Department of ..BOTANY.....
Ph. D Course Work Marks 2015.

The following are the marks secured by Mr/Ms ..... DHRUBAJYOTI DAS

Roll No. ..... 04/14 .....*of Ph. D Course Work Examination held in 20*..15.....

Subject(s)/Paper(s)	Max. Marks	Minimum Qualifying Marks	Marks Scored
Course No..... Research Methodology B. Ph. D-I	100	35	47
Course No..... Elective/Optional B. Ph. D-II	100	35	52
Course No..... Review of Literature Reports/Seminar B. Ph. D-III	100	35	78
Total Aggregate marks			177
Average Pass Mark - 55%			

Result	Rank/Class/Division	Percentage
PASSED	3RD	59%

Marks Entered by
Dated .....

Marks Compared by
Dated .....

Asstt./Deputy Register (Examination)
Dated .....

NAGALAND UNIVERSITY

HEAD QUARTERS : LUMAMI

Ph. D COURSE WORK EXAMINATION

This is to certify that Mr/Ms..... DHRUBAJYOTI DAS .....

of Nagaland University bearing Roll No..... 04/14 .....*is qualified in the Ph.D Course Work Examination*

in the Department of..... BOTANY .....*Nagaland University held in the Year 20*..15.....

Head of Department
Head
Department of Botany
Nagaland University
Hqs. Lumami.

Dean
School of Sciences
Nagaland University
Hqs. Lumami Nagaland

List of Tables

Table No.	Table Legend	Page No.
4.1	Genomics features of *Pseudomonas aeruginosa* N002 and *Enterobacter* sp. RC4 genome	53
4.2	Comparison of COG categories between *P. aeruginosa* N002 and *Enterobacter* sp. RC4 with other genomes of hydrocarbon degrading bacteria in IMG database	56-57
4.3	Comparison of COG categories between *P. aeruginosa* N002 and *Enterobacter* sp. RC4 within the same group in available genome in IMG database	57-58
4.4	Insertion sequences predicted in *P. aeruginosa* N002 and *Enterobacter* sp. RC4	59-60
4.5	Genomic Island (GI) analysis of *P. aeruginosa* N002 and *Enterobacter* sp. RC4	60-62
4.6	Two Component system (TCS) genes of *P. aeruginosa* N002 and *Enterobacter* sp. RC4 genome	64-74
4.7	Adaptive genes found in *P. aeruginosa* N002 and *Enterobacter* sp. RC4	90
4.8	Xenobiotic degradation genes of *P. aeruginosa* N002 and *Enterobacter* sp. RC4.	92-99
4.9	*P. aeruginosa* N002 and *Enterobacter sp.* RC4 genomes cell motility and chemotaxis genes	115-120
4.10	Signal transduction genes of *P. aeruginosa* N002 and *Enterobacter* sp. RC4 genome	122-146
4.11	Transcription genes of *P. aeruginosa* N002 and *Enterobacter sp.* RC4	146-173
4.12	Validation of crude oil degrading genes of *P. aeruginosa* N002 and *Enterobacter* sp. RC4. through qRT-PCR	174
4.13	Soil improvement and functional characteristics of *P. aeruginosa* N002 and *Enterobacter* sp. RC4	175-176

List of Figures

Figure	Figure Legend	Page No.
3.1	Crude oil contaminated soil sites of Duliajan and Gelakey, Assam in North East (NE) India	42
3.2.A	Culture plate of *Pseudomonas aeruginosa* N002	46
3.2.B	*Enterobacter* sp. RC4. bacteria isolated from Duliajan and Gelakey, Assam of North East India	46
4.1.A	Phylogenetic tree of *Pseudomonasaeruginosa* N002	52
4.1:B	Phylogenetic tree of *Enterobacter* sp. RC4	52
4.2:A	Circular chromosome map of *P. aeruginosa* N002	54
4.2:B	Circular chromosome map of *Enterobacter* sp. RC4	55
4.3.A	Genomic Island of *P. aeruginosa* N002 prediction by different methods	63
4.3.B	Genomic Island of *Enterobacter* sp. RC4 prediction by different methods	63
4.4.A	*P. aeruginosa* N002 CitB family two component system pathways	74
4.4.B	*Enterobacter* sp. RC4 CitB family two component system pathways	75
4.4.C	*P. aeruginosa* N002 Chemotaxis family two component system pathways	76
4.4.D	*Enterobacter* sp. RC4 Chemotaxis family two component system pathways	77
4.4.E	*P. aeruginosa* N002 Lux, LuxR, and, Cell cycle family two component system pathways	78
4.4.F	*Enterobacter* sp. RC4 Lux, LuxR, and, Cell cycle family two component system pathways	79
4.4.G	*P. aeruginosa* N002 LytT family two component system pathways	79
4.4.H	*Enterobacter* sp. RC4 LytT family two component system pathways	80
4.4.I	*P. aeruginosa* N002 NarL family two component system	81

	pathways	
4.4.J	***Enterobacter* sp. RC4 NarL family two component system pathways**	**82**
4.4.K	***P. aeruginosa* N002 NtrC family two component system pathways**	**83**
4.4.L	***Enterobacter* sp. RC4 NtrC family two component system pathways**	**84**
4.4.M	***P. aeruginosa* N002 OmpR family two component system pathways**	**85**
4.4.N	***Enterobacter* sp. RC4 OmpR family two component system pathways**	**86**
4.4.O	***P. aeruginosa* N002 Other family two component system pathways**	**87**
4.4.P	***Enterobacter* sp. RC4 Other family two component system pathways**	**88**
4.4.Q	***P. aeruginosa* N002 Sporilation family two component system pathways**	**89**
4.4.R	***Enterobacter* sp. RC4 Sporilation family two component system pathways**	**89**
4.5	**Environmental adaptation pathway of *P. aeruginosa* N002 and *Enterobacter* sp. RC4**	**91**
4.6.A	***P. aeruginosa* N002 genome chloroalkane and chloroalkene degrading pathway**	**101**
4.6.B	***Enterobacter* sp. RC4 genome chloroalkane and chloroalkene degrading pathway**	**102**
4.6.C	***P aeruginosa* N002 genome polycyclic aromatic hydrocarbon degradation pathway**	**103**
4.6.D	***Enterobacter* sp. RC4 genome polycyclic aromatic hydrocarbon degrading pathway**	**104**
4.6.E	***P. aeruginosa* N002 genome Chlorocyclohexane and Chelrobenzene degrading pathway**	**105**
4.6.F	***Enterobacter* sp. RC4 genome Chlorocyclohexane and Chelrobenzene degradation pathway**	**106**

4.6.G	***P. aeruginosa* N002 Benzoate degradation pathway**	**107**
4.6.H	***Enterobacter* ps. RC4 genome benzoate degradation pathway**	**108**
4.6.I	***P. aeruginosa* N002 genome fluorobenzoate degradatioon pathway**	**109**
4.6.J	***P. aeruginosa* N002 genome xylene degradation pathway**	**110**
4.6.K	***P. aeruginosa* N002 genome Styrene degradation pathway**	**111**
4.6.L	***Enterobacter* sp. RC4 genome aminobenzoate degradation pathway**	**112**
4.6.M	***Enterobacter* sp. RC4 genome caprolactam degradation pathway**	**113**
4.6.N	***Enterobacter* sp. RC4 genome toluene degradation pathway**	**114**
4.7.A	***P. aeruginosa* N002 genomes bacterial chemotaxis pathway**	**121**
4.7.B	***Enterobacter* sp. RC4genomes bacterial chemotaxis pathway**	**121**

CHAPTER – 1

Introduction

Comparative genomics is a broad field of bioinformatics which compares two or more genomes (Gresham et al., 2008). Comparative study covers global comparative hybridization studies using nonsequencing technologies such as microarrays (Willenbrock et al., 2007), target gene studies that focus on only specific gene or noncoding region studies of specific pathways and whole-genome sequence alignments (Wang et al., 2009; Wei et al., 2002). *Pseudomonas* and *Enterobacter* are two bacterial species highly adapted to various environments and have been exploited for various economic uses (Sanders et al., 1997; Stover et al., 2000; Xu et al., 2018). The main carrier of genetic information, the genome plays a critical role in transmitting hereditary characteristics from one generation to the next. Considering the same comparative genomic analysis of bacterial species for various activities is studied in detail (Al-Dous et al., 2011). However these need further analysis for efficient application.

1.1. Next generation sequencing for comparative genomics study

The traditional DNA sequencing methods including the Maxam-Gilbert method (Maxam et al., 1977) and the Chain-termination method (Sanger et al., 1975) along with the other new sequencing methods like next generation sequencing (NGS) have been developed to unravel the genetic information. Researchers have predicted that many genes of the bacterial genome have converged from different organisms in the course of evolution. Important mechanism during this course of evolution may include horizontal gene transfer, insertion sequences (IS) and transposons playing functional rule (Gilbert et

al., 2013). These mechanisms contribute to the variability of bacterial genomes by enabling bacteria to acquire and incorporate genetic material into their genome, where it may form genomic islands (Dobrindt et al., 2004). Such genetic material may not always be advantageous to the host and is therefore a genetic and metabolic burden to bacteria. In such cases bacterial genomes tend to lose their substantial information (Moran, 2002; Ahmed et al., 2008). Genetic material coding for pathogenicity or fitness tors confers a selective advantage to the host. In the case of pathogenic bacteria, this horizontally acquired genetic material may contribute to the colonization and invasion of host tissue. Increased bacterial fitness or pathogenicity promotes the stabilization of the corresponding determinants in the recipient's genome, and the stable integration of horizontally acquired DNA is most frequently connected to a distinct biological function (Ochman et al., 2000). As such the investigation of genomic island (GI) is an interesting subject in the modern era of genomic analysis. It has been reported that bacteria not only work with single component. But the integration of two component system (TCS) is essential for bacterial. TCS are ubiquitous in bacteria and lower eukaryotes (Wuichet et al., 2010).They are the main signal transduction pathways in these species and how an organism can coordinate the activity of so many highly related signaling systems is the focus of much attention (Laub et al., 2007). TCS generally consist of a sensor histidine protein kinase (HPK) and a response regulator (RR). Classically HPKs contain a periplasmic sensory region, a transmembrane region and a cytoplasmic signaling region. The cytoplasmic part of the kinase consists of a HAMP linker domain (present in Histidine kinases, Adenylcyclases, Methyl-accepting proteins and Phosphatases), a dimerisation and histidinephospho accepting domain (DHp) and a catatlytic domain (CA). The RR usually consists of a phosphyl receiver (REC) domain and a regulatory DNA binding domain. In a canonical TCS the histidine kinase senses environmental

changes which trigger autophosphorylation and phospho transfer to the response regulator and it leads to a conformational change in the response regulator allowing binding to a specific region of DNA and regulation of transcription (Mattison et al., 2002; Szurmant et al., 2010). Therefore, analysis of TCS in bacterial genome is important to predict the bacterial functions. Comparative genome analysis of related species should provide such general approach for identifying functional elements without prior knowledge of function. Evolution insistently plays with genome sequence and tests the results by natural selection. Mutations in non-functional nucleotides are tolerated and accumulate over evolutionary time. However, mutations in functional nucleotides are deleterious to the organism that carry them, and become sparse or extinct. Hence, functional elements should stand out by virtue of having a greater degree of conservation across the genomes of related species. Recent studies have demonstrated the potential power of comparative genomic comparison. Cross-species conservation has previously been used to identify putative genes or regulatory elements in small genomic regions (Oeltjen et al., 1997; McGuire et al., 2000; Pennacchio et al., 2001). Light sampling of whole-genome sequence has been used as a way to improve genome annotation (Blandin et al., 2000; Cliften et al., 2001).

1.2. Important of Microbes in Degraded Environment

Complete bacterial genomes have been compared to identify evolutionary *adaptivity,* pathogenicity of genes and wide comparison has been used to estimate the proportion of the selected bacterial genome (McClelland et al., 2000; Carlton et al., 2002; Perrin et al., 2002). The utility of adapted microbial consortium and nutrients has been used in bioremediation of environmental pollution (Onwurah et al., 2007). However, fluorescent pseudomonads (gram-negative, non-spore forming, motile, rod shaped and versatile bacteria) have been reported to promote plant growth in rhizosphere either

directly by producing plant growth regulators and by increasing the uptake of some micro and macro elements from rhizosphere (Lugtenberg et al., 1991; Deka Boruah et al., 2003; Glick, 2012) or indirectly through biological control of pathogens or induction of host defense mechanisms (Van Loon et al., 1998; Deka Boruah et al., 2002). It is well known that plants support several species of bacteria including those that can affect the growth and formation of root system and their mode of action (Whiteman et al., 1987). In one approach plants are used in cleaning up of contaminated environments, and also promise effective, inexpensive, and less intrusive cleaning up and restoration of contaminated environments (Stomp et al., 1993; Glick, 2003; Suresh and Ravishankar, 2004).

Soil is considered to be the most diverse form of natural environment on earth (Vogel et al., 2009). Lowering and losing of soil functions degradation causes serious environmental problems in the world. It is estimated that 22% of total cropland, pasture, forest, and woodland in the world have been degraded (Jie et al., 2002) and more than 15% of this degradations are due to erosion, nutrient status change, salinisation, chemical deterioration, compaction and pollution (Bridges et al., 1999). Among these, soil pollution has been a recent topic of discussion of great concern for the world. Soil pollution is mainly responsible for the changes in physical and chemical properties of soils which result in adverse effects on ecosystems, and to tackle such situation microbes assisted augmentation is an economical and versatile alternative approach to the physico-chemical treatment methods (Margesin et al., 2001; Al-Mailem et al., 2010).

1.3. Study Area and Scope of the Present Study

FromNorth East (NE) India mainly Duliajan and Geleky, Assam houses large reserves of crude oil. Due to exploration, production, maintenance, transportation, storage and abandonment of crude oil from those areas soil are contaminated (Yenn et al., 2014; Das et al., 2015). The non-degradability and persistent nature of hydrocarbon pollutants is

a significant environmental problem (Hu et al., 2013; Oudot et al., 1998). Having saturated and aromatic hydrocarbons, polar compounds, resins and asphaltene as constituents, crude oil possesses diverse toxic effects on plants, animalsandhuman health (Atlas, 1981). The manifold effects of carcinogenic, mutagenic and immuno-toxic properties of crude oil components pose major threat to environment (Vasudevan and Rajaram, 2001; Lim et al., 2016;).Various studies have been reported that biological ways mainly bacterial degradation or decontaminating ofthat hydrocarbon contaminated sites is very important. From the above discussions, it is of importance to study on genome analysis of beneficial microbes like *Pseudomonas, Enterobacter* etc. that might help to identify beneficial genes and help in restorations of degraded environment for long term sustainability. Therefore, this study aimed to analyze the bacterial genome and the genes associated with crude oil bioremediation, soil quality restoration and adaptation in hazard oil contaminated environment.

1.4. Objectives of the Present Study

The work was done by the following objectives:

1. Whole genome sequencing of *Pseudomonas aeruginosa* N002 and *Enterobacter*spp.RC4.
2. Prediction of the genomic island of *P. aeruginosa* N002 and *Enterobacter*spp.RC4.
3. Determination of the two component systems of *P. aeruginosa* N002 and *Enterobacter*spp. RC4 on environmental adaptation.
4. Determination of the genes and traits responsible for environmental soil stress tolerant (moisture, metal, degraded forest environmental biological activities and some polycyclic aromatic hydrocarbon pollutants).
5. Validation of some of the functional genes ameliorates environmental soil stress of *P. aeruginosa* and *Enterobacter* spp. RC4.

6. Validation of improvement of soil functional characteristics (biological activities) under small field condition.

CHAPTER - 2

Review of Literatures

2.1: Comparative Genomics

Comparative genomic analyses lend insight into structural features such as variations related to genomic rearrangements, changes in the gene repertory, identification of horizontal gene transfer elements and prophage related sequences, and hence expose particularities on the evolution in this species (Land et al., 2015). It analyses have a conserved “core” genome shared among nearly all members of the species interspersed with “accessory” genomic elements that are present in some but absent in other strains (Tettelin et al., 2008). The genome sequencing of strains belonging to the same species offers the possibility of defining their pan-genome, which comprises the core-genome and the accessory genome compartment. This strain-specific accessory genome may also be involved in bioremediation or clinical activities of pathogenicity, soil quality restoration, drug resistance, and stress responses in different environment (Margesin et al., 2001; Syvanen et al., 2012;). While these factors may increase the adaptability of strain to the particular niches they inhabit, such as oil contaminate site, forest soil, clinically isolate etc. they are not imperative to the survival of the organism. Moreover, some of these genes can be acquired by horizontal gene transfer and have also been shown to be over represented in genomic islands or insertion sequence (Syvanen et al., 2012). Sequencing of DNA and protein started in the 1970s when the virus Lambda (50,000 nucleotides) was sequenced by Sanger et al (Sanger et al., 1997). Around this

time DNA sequencing was carried out for small genomes such as viruses and organelles, and complete sequencing of a bacterium, was not feasible because of economic and technical limitations. However, later on, sequencing of the human genome, and improvements in sequencing technologies facilitated whole genome sequencing of bacteria. The first bacterium to be sequenced was *Haemophilus influenza* **(**Fleischmann et al., 1995**)**, and this was done by the shotgun method developed by Sanger (Sanger et al., 1997) in briefly, the shot gun method of sequencing consists of randomly sampling and determining 500–700 nucleotide reads and then assembling them to reconstruct the sampled sequence (Fraser et al., 1997). Because the assembly process is based on finding regions that overlap, more than 1 million bases must be sequenced in order to sequence a 1 Mb genome. The mean value of the number of times each base is sequenced in a genome project is called genome coverage and is usually between 6 and 8 **(**Fraser et al., 1997). The method of sequencing developed by Sanger is considered the gold standard, and over the years, whole genome sequencing of many bacteria has been carried out using this method.

2.1.1: Important of Comparative Genomics Study

The important of comparative genomic analyses are not only for distantly related genomes, but also for closely related genomes, because of their applications in health and industry. Therefore, whole genome comparative analysis could have numerous advantages in narrowing down the valuable genomic information and identifying candidate regions in genomes (Drukewitz et al., 2019). For comparison strategies, there is no standard criterion for how many genomes (gene and protein sequences) shall be initially compared, one can start from two to an unlimited number of genomes. It also gives high levels of similarity between closely related organisms, similarity or dissimilarity between two or more organisms, such as same group of bacteria or different

group of bacteria, extreme diversity of the gene composition in different evolutionary lineages (Sivashankari et al., 2007). Moreover, the comparative studies may be performed on intra or inter-species level, using bacteria with similar or different lifestyles, depending on the study objective. Taking into account the importance of the comparative genomic studies for understanding the inter and intra-species genomic variations, conserved core and species pan-genome, protein-protein interaction and regulatory mechanisms, virulence factors and candidate genes, proteins, and its application in designing bioremediation, vaccines, diagnostics, drug development etc (Sivashankari et al., 2007; Drukewitz et al., 2019).

2.2: Next-Generation Sequencing

Next generation sequencing technologies have emerged, which are high throughput and able to generate three to four orders of magnitude more sequences and are also relatively less expensive (Bonetta et al., 2006). Next generation sequencing methods employ a wide spectrum of technologies such as sequencing by synthesis (Mardis et al., 2008; Su et al., 2011), sequencing by ligation, single molecule DNA sequencing (Su et ai., 2011) and polony sequencing (Shendure et al., 2005). In recent times, the sequencing industry seems to be dominated by Illumina, who have introduced three next generation sequencing platform including GAIIx, Hiseq-2000 and Miseq (Quail et al., 2012). These sequencing platforms employ a sequencing-by-synthesis approach Mardis et al., 2008; Su et al., 2011). In this method, DNA molecules and primers are attached on a slide and amplified with DNA polymerase resulting in the formation of clonal DNA colonies (DNA clusters). To evaluate the DNA sequence, four types of fluorescently labeled reversible-terminator nucleotides are added and the incorporated nucleotides are imaged. The fluorescent dye with the terminal 3' blocker is then chemically eliminated from the DNA, allowing for the next cycle to start.

2.2.1: Next-Generation Sequencing in Bioremediation

The complete genome sequence of microorganism having potential in bioremediation studies are not yet accelerated in quick way (Tiedje, 2002; Golyshin et al., 2003; Rabus et al., 2005; Seshadri et al., 2005). Using the complete genome sequence researchers can analyze the expression of all genes in each genome under different environmental condition with the help of DNA microarray technique (Muffler et al., 2002; Schut et al., 2003; Gao et al., 2004). Such type of genome expression analysis provides important information for identification of regulatory pathway in the microorganisms (Muffler et al., 2002; Lovley, 2003; Rabus et al., 2005). Next-Generation sequencing (NGS) technology also lead to start real revolution in environmental biotechnology and bioremediation, it extends its novelty with the other edge of science such as genomics, proteomics, metagenomics and transcriptomics (Ma et al., 2012). Bioremediation and biodegradation is the now developed with standard molecular technologies and play important role in degradation process. NGS technology is capable to produced new monoclonal and digital DNA data in huge amount with negligible prices. Because of this, it may be more convenient tools and techniques for researcher working in the field of environmental science (Eiler et al., 2012). NGS approach is helpful for the identification and quantification of microbes which are uncultivable in laboratory condition and very important in the bioremediation studies. This type of long read sequence contains two hyper variable region which enables specific taxonomic classification of bacteria, Achaea, fungi, protozoa and algae (Sims, 2013). NGS technology enhanced the functional genomics for cultivated microorganism laboratory and with the help of daily routing molecular techniques such as fragmentation of DNA, restriction digestion and sequencing it leads to standardize the organisms and this will be helpful in development of bioremediation area (Bihari, 2013). NGS technique

is strong, straightforward and cost effective technology with high accuracy but it required proper instrumental and infrastructure which may help to investigate the novel catabolic pathways, mutations, peculiar genetic arrangements in chromosomes or in cryptic plasmids.

2.2.2: Different type of Platforms and Tools Used in Next Generation Sequencing

Sequencing platforms that employ next generation sequencing technologies are being produced at a fast rate, with two major sequencing platforms introduced in 2011, namely Ion Torrent Personal Genome Machine (ITPGM) and the Pacific Biosciences (PacBio) (Rothberg et al., 2011; Eid et al., 2009). PacBio sequences single molecules in real time without amplification (Eid et al., 2009). In this method, a conjugate of DNA polymerase and DNA template are attached to 50 nm-wide wells. Using nucleotide fluorescently labeled with γ-phosphate, the DNA polymerase carry out second strand DNA synthesis. Incorporation of bases during DNA synthesis is detected by means of a distinct pulse of fluorescence. ITPGM employs technological advances in semi-conductor science and non-sensitive transistors to sequence DNA (Rothberg et al., 2011). This method differs from other next generation sequencing methods as polymerisation events are detected by *p*H changes rather than light. DNA fragments carrying specific adapter sequences are linked to a bead and then clonally amplified by emulsion PCR. The template beads are loaded onto a chip which has proton sensing wells that are fabricated on a silicon wafer, and sequencing is primed from a predetermined location in the adapter sequence. As bases are incorporated during the sequencing process, protons are released and a signal is detected proportional to the number of bases incorporated. Further advances in genome sequencing are expected in the near future as the so called third generation technologies are being developed to further increase throughput, decrease cost, and reduce the time to obtaining results. One interesting area of such sequencing methods

involves microscopy based techniques such as atomic force microscopy that are used to identify the locations of nucleotides within long DNA fragments. (Xu et al., 2009)

Many tool, software and pipeline are used for analysis of comparative genomics study. These are like BLAST Ring Image Generator *(*BRIG), Mauve, ACT, Microscope, IMG and EDGAR etc. (Len et al. 2003; Alikhan et al. 2011) are used for comparative genomics study.

2.3: Horizontal Gene Transfer (HGT) and Mobile Genetic Elements (MGEs)

Horizontal gene transfer generates rapid and extremely dynamic genomes, rather than evolution through the modification of existing genetic information (Ochman et al., 2000). There are three fundamentally distinct mechanisms by which HGT can occur: transformation, conjugation and transduction. Transformation refers to the process when a cell takes up isolated DNA from the environment and has the potential to transfer DNA between distantly related organisms. Many bacterial species are naturally competent to uptake DNA, such as *Bacillus subtilis, Haemophilus influenzae* and *Neisseria* sp. A second mechanism is conjugation, which is defined as the direct transmission of DNA from one cell to another. In contrast to transformation, which is DNase sensitive and does not require cell-to-cell contact, conjugation is DNase-resistant and does require cell-to-cell contact. A third mechanism is transduction, which is thought to be phage mediated transfer of genetic materials.

Bacterial genomes generally consist of stable regions called the “core genome” and variable regions, obtained from the “flexible gene pool” (Hacker et al., 2003). The “core genome” of a certain species has a fairly homogeneous G+C content and codon usage, and often encodes housekeeping functions and carries gene clusters with relatively low mutational capacity (Hacker and Kaper, 2000). In contrast, the” flexible gene pool” represents the total amount of foreign DNAs available for recipient cells. If genes from

the “flexible gene pool” were obtained from a source organism with a different mutational bias, their G+C content and codon usage will be different from the rest of the genome. Most often, these foreign genes are carried by MGEs, which are able to move within or between genomes via HGT. The acquisition of sequences from distantly related organisms may confer very new phenotypic characteristics on the recipient cell over a short evolutionary timescale. In the past few years, there has been growing evidence that HGT has played a vital role in the evolution of bacterial genomes (Ochman et al., 2000). MGEs include plasmids, transposons, bacteriophages, integrons, insertion sequence elements and genomic islands. Their size ranges from a few hundred base pairs up to 100 kb (Dobrindt et al., 2004).

2.3.1: Insertion Sequence (IS) Elements

Insertion Sequence elements are small, genetically compact DNA sequences, normally less than 2.5 kb in length (Mahillon et al., 1998; Mahillon et al., 1999). The overall structure of most IS elements is very similar, and includes a central tansponsase (Tpase) gene flanked by inverted repeat (IR) sequences that exactly define the borders of the elements (Mahillon et al. 1998). Generally, IS elements encode no functions other than their own translocation, and the central Tpases function to transpose their own elements both within and between genomes (Mahillon el al., 1998; Mahillon et al., 1999). The IRs is short, between 10 and 40 bp, and often contains the promoter for the Tpase gene at one side (Mahillon et al., 1999). In addition, IS elements are also flanked by further short (between 2 and 14 bp), duplicated direct repeated sequences. However, these direct repeats (DR) does not belong to IS elements, but arise from duplication in the recipient DNA at the insertion site (Ou et al., 2006). Because of the presence of repeated sequences, IS elements can also be regarded as repetitive sequences that are randomly scattered throughout the bacterial genome. The presence of two copies of an IS element

can lead to homologous recombination between them, which results in inversion, deletion or replicon fusion of the region between two IS elements (Siguier et al., 2006b). For this reason, IS elements play an important role in promoting genome rearrangement. Over 1,600 different ISs have been identified to date (www-is.biotoul.fr) (Siguier et al., 2006b). These ISs have been classified into about 20 families, based on similarities and conservation in the sequence of their Tpases, genetic organisation and IRs (Mahillon et al., 1999; Siguier et al., 2006a). In addition to chromosomal DNA, IS elements are also commonly found in bacterial plasmids, and are an integral part of many naturally occurring bacterial plasmids (Mahillon et al., 1999). The ISs in plasmids function in plasmid transfer and integration into the host chromosome. Furthermore, many antibiotic resistance genes are often encoded on plasmids and spread within bacterial populations with the aid of ISs (Mahillon et al., 1999). Thus, although the basic function of ISs is simply translocation, they are involved in other activities to shape genomic variation. In the post-genomic era, complete genome sequences and genomic comparisons provide advantages in defining IS classification, understanding mechanisms of distribution and identifying their role in evolution. For example, when the first *Francisella* genome Schu S4 was published in 2005, the genome was found to be notable for its large complement of IS elements (Larsson et al., 2005). It is now known that about 1 to 5% of annotated *Francisella* genes are transposase genes (Titball et al., 2007). Five different types of IS elements (ISFtu1-ISFtu5) were found in the Schu S4 genome, and are conserved in other sequenced *Francisella* genomes such as OSU18 (Petrosino et al., 2006). More surprisingly, ISFtu1 belongs to the IS630 Tc-1 mariner family of transposons, of which about 50 copies were found in the Schu S4 genome and 58 copies in the OSU18 genome (Larsson et al., 2005; Petrosino et al., 2006). However, the presence of this kind of Tc-1 transposon is not usual in bacteria, but they are generally found in eukaryotes and have

been reported in a range of invertebrates such as nematodes and insects (Larsson et al., 2005). The high level of Francisella IS630 elements was suggested to be acquired originally from the infected insect vectors, which commonly mediate Francisella's transmission (Larsson et al., 2005).

The expansion of the transposable elements, particularly IS elements, in bacterial pathogens is a recent emerging common feature (Siguier et al., 2006a). This is because IS elements facilitate homologous recombination within a genome, a process that can provoke large-scale genome rearrangements. These genomic changes often disrupt the ancestral gene order, and bring about a high level of gene inactivation and gene loss. Horizontally acquired genes have been balanced by gene loss, so that the bacterial genome size will not continuously increase. For example, comparison among three members of the Bordetella family, *Bordetella bronchiseptica, Bordetella parapertussis* and *Bordetella pertussis*, revealed a smaller genome accompanied by more ISs (Parkhill et al., 2003). A similar scenario can be seen in the Yersinae, of which a *Yersinia pestis* CO92 strain shows amplification of at least four different ISs: 66 copies of IS1541, 44 copies of IS100 (a member of the IS21 family), 21 copies of IS285 (a member of the IS256 family), and 9 copies of IS1661 (a member of the IS3 family) (Parkhill et al., 2001). Overall numbers of IS copies in the CO92 strain are around ten-fold higher than those in another strain, *Yersinia pseudotuberculosis* (Parkhill et al., 2001). In addition, the genome displays anomalies in GC base-composition bias, indicating frequent intragenomic recombination through ISs (Parkhill et al., 2001).

2.3.2: Bacteriophages

Bacterophages (or phages for short) contain either DNA or RNA enclosed in a protein coat. They are simply viruses that infect bacteria. Phage DNA fulfils a number of criteria for being an ideal vehicle for horizontal gene transfer. The residual footprints of

prophages are different G+C contents and codon usage from the host's genome, the adjacent tRNA genes and the IS elements in the phage essential regions. However, the prophages in bacterial genome sequences are accurately recognized through the similarity of their genes to known phage genes. Although phage genomes encompass an enormous amount of sequence diversity, their genes appear to be highly conserved (Casjens, 2003).

2.3.3: Genomic Islands

Genomic islands are referred to as large chromosomal regions that contain a cluster of functionally related genes. They are often flanked by direct repeat sequences and are located near an integrase or transposase gene and also close to a tRNA gene (Dobrindt et al., 2004). Genomic islands encoding virulence factors of pathogenic bacteria have been designated "pathogenicity islands" (PAIs) (Hacker et al., 1997). Pathogenicity islands are characterized by five features: (i) they are large clusters (10-200 kb in size) present in the genomes of pathogenic strains but absent from those non-pathogenic strains of the related species; (ii) their G+C content differs from the rest of the host genome; (iii) they are often associated with tRNA genes; (iv) they are presumed to be generated by horizontal gene transfer; (v) they are recognized by conferring upon the host bacterium a complex and distinctive virulence phenotype in a single step (Hacker et al., 1997). For example could be taken from a well-characterized pathogenicity island known as the Locus for *Enterocyte Effacement* (LEE). LEE PAIs are found in the genomes of enteropathogenic *E. coli* (EPEC) and enterohaemorrhagic *E. coli* (EHEC) (Nataro et al., 1998), as well as the closely related mouse pathogen *Citrobacter rodentium*. The LEE gene cluster encodes a bacterial Type III secretion system, and can be present on a single horizontally transferable plasmid (Deng et al., 2001), or inserted into the EPEC or EHEC genome at *selC, pheV,* or *pheU*tRNA sites (Jores et al., 2004). The entire LEE is 35.6 kb in size, with a GC content of 38.36%, far below the *E. coli*

genomic average (50.8%). LEE-encoded proteins are responsible for the development of a characteristic histopathological feature known as “attaching and effacing” lesions (Frankel et al., 1998).

2.4: Pan-Genome and Core Genome

The main goal of pan-genome is the genomic comparison of different strains of the same species, or even genus (Snipen et al., 2009). Currently, the availability of a large number of genomes from different isolates of the same pathogen has opened the possibility of investigating several genomic characteristics that are intrinsic to one or more species (Tettelin et al., 2005). The first work that described the term pan-genome was conducted by Tettelin et al. (2005) who used eight different strains of *Streptococcus agalactiae*, a pathogenic species isolated from human. After this research, other studies were performed using pan-genomic analysis for different microorganisms, including *Bacillus cereus*, *Escherichia coli*, *Sulfolobusislandicus*, *Streptococcus pneumoniae*, *Methanobrevibactersmithii*, *Corynebacterium*, *Corynebacterium pseudotuberculosis*, and *Pantoeaananatis*, among others (De Maayer et al., 2014; Rasko et al., 2005). The idea of pan-genomic studies brings significant insights of the understanding of bacterial evolution, niche adaptation, population structure and host interaction as well as inferences in more applied issues, such as vaccine and drug design and the identification of virulence genes. The term “pan-genome” reflects the total number of non-redundant genes that are present in a given dataset (Tettelin et al., 2005; Snipen et al., 2009). It consists basically of three parts: i) core genome, formed by genes shared by all genomes and usually involved in essential cellular processes; ii) accessory or dispensable genome, composed of genes absent in some isolates; and iii) species-specific or strain-specific genes, which are those genes that are present in a single genome (Tettelin et al., 2005; Mira et al., 2010). Usually, genes that are present in accessory and species-specific or

strain-specific genes are involved in niche adaptation (Medini et al., 2005; Tettelin et al., 2005).

The core genome is the subset of genes that are present in all of the genomes, and it can be determined by comparing the different genomes (Muzzi et al., 2007). Lapierre and Gogarten (2009) said that over 250 gene families have been characterized as part of the bacterial core genome and these gene families constitute evidence that gene conservation highlights the conservative nature of evolution. Normally, genes that are present in the core genome are associated with the maintenance of the basic aspects of the organism's biology and are mainly related to replication, translation and maintenance of cellular homeostasis (Medini et al., 2005; Tettelin et al., 2005). Moreover, the core genome undergoes significant selective pressure in relation to its function, which inhibits the occurrence of drastic changes (Lapierre et al., 2009). The number of genes that compound the core genome could indicate the genetic diversity among the studied organisms; thus, the core genome becomes smaller when diversity increases among the organisms (Lawrence et al., 2005). On the other hand, phylogenetically related genomes tend to share more genes and consequently present a larger core genome (Lerat et al., 2003; Lawrence et al., 2005).

2.5: Crude Oil Contamination and its Effect in Environment

2.5.1: Crude Oil or Petroleum Hydrocarbons (PHC)

Crude oil is natural liquid which is formed when huge quantities of dead organisms and organic matter get buried underneath sedimentary rock and subjected to intense pressure and heat; it's found in geological formations beneath the Earth's surface (Balba et al., 1998). Crude oil is complex mixture of hydrocarbons, composed of paraffins (15- 60%), naphthenes (30-60%), aromatics (3-30%) and asphaltics (remainder) fractions along with nitrogen, oxygen and sulfur containing compounds. The aliphatic

and aromatic fraction includes linear or branched-chain alkanes, cycloalkanes and poly-nuclear cyclic aromatic hydrocarbons (PAH) containing alkyl side chains or fused cycloalkanes. The resins and asphaltenes contain more polar compounds, consisting of heterocyclic, oxygenated hydrocarbons and high molecular weight aggregates. Crude oil are fractionized the petrochemical factories to produce commercial fuels like petrol, diesel, kerosene etc., plastics, synthetic rubbers, and other chemicals (Harayama et al., 1999).

2.5.2: Pollution of Petroleum Hydrocarbons

Petroleum hydrocarbons are one of the major environmental pollutants (Okoh et al., 2006). In the time of natural deposits, leakage of storage tanks and pipelines, land disposal of petroleum wastes, transportation its damage the environment and recognized as the most significant contamination problem on the world (Balba et al., 1998; Snape et al., 2001). These oil spills include widespread, long-term, and serious damage to human health, natural resources, marine ecosystems and terrestrial life.

2.5.3: Effect of Crude Oil Pollution in Soils

Crude oil are absorbed on the surface of mineral and soil organic matter, it fixed within the soil pores which found in a continuous cover on the soil surface and also increase organic carbon content in the soil, soil humus also enriched in humic acids, whereas the degree of humification of soil organic matter decreases (Trofimov and Rozanova, 2003). As a result it's affects certain soil parameters such as the mineral and organic matter content, the cation exchange capacity (CEC), redox properties, pH, lower water holding capacity, moisture content and hydraulic conductivity as compared to unpolluted soils (Nwaoguikpe, 2011). It affects the biotic components of the soil ecosystem. Crude oil pollution leads to decrease growth, lower rate of seed germination, seedlings and productivity of the plant (Sharifi et al., 2007). Aromatic compound are

carcinogenic, mutagenic it also cause lower respiratory problems, long term mental health effects, genotoxic damage, hormonal imbalance, reproductive and developmental toxic effects, skin and lung cancer in human or other animal and bird (Aguilera et al.,2010). This crude oil contaminates soil effects microbial diversity, population and the structural and functional properties of microbial membrane (Van Hamme et al., 2003).

2.6: Method for Treatment of Contaminated Soil

The toxic affects of crude oil on different biotic components its demand for the remediation of the contaminated soils. Various physical, chemical and biological methods are available for the treatment of PHC contaminated soils. These methods can be employed to restore soil quality or microbial diversity.

2.6.1: Physical Methods

Commonly used methods are cap and contain, landfill and incineration. A cap and contain method the contaminated soil is treated on the site. The contaminated site is capped, as in landfill, and is monitored periodically for reduction in the contamination.In landfill is a carefully engineered pit that is dug in the ground where the contaminated soil from actual site is excavated and put, then covered with soil and spread evenly in layers. This method aims to isolate and contain the hazardous waste and avoid pollution of ground water and surface water.Incineration is a disposal method that involves combustion of hazardous material and converts this material into ash, gas, heat and steam.

2.6.2: Chemical Methods

The most commonly used chemical methods for PHC removal are Ozonation and surfactant washing. In this method molecular ozone or its decomposition products react with organic compounds to convert them into oxidized products, which are more water soluble, less toxic and more bio-available than parental compounds. The costs for the clean-up of the contaminated sites with conventional physical and chemical techniques

are enormous, due to lack of public acceptance and technological complexities, these methods have not been successfully applied (Vidali et al., 2001; Liu et al., 2010). Therefore, alternative methods such as biological methods which is environmental friendly, less expensive also to restore contaminated sites.

2.6.3: Biological Methods

2.6.3.1: Phytoremediation

Phytoremediation is the use of plants for in situ remediation of contaminated soil, sludge, sediment and groundwater through different mechanism. A number of plants which have extensive fibrous roots such as common grasses, corn, wheat, soyabean, peas, beans and several trees of salicaceae family have potential to rhizoremediation (Trapp et al., 2001; Glick, 2003). Many researchers also reported phytoremediation technology for the treatment of soil contaminated with crude oil contaminated area. Some of the plant species, alfalfa (*Medicago sativa*), switch grass (*Panicum virgatum*) and little bluestem grass (*Schizachyrium scoparium*) were found to remediate 72% of total PAH within 6 months (Pradhan et al., 1998). Some another researcher also studied phytoremediation of two to four ring alkylated in crude oil-contaminated soil using treatment systems involving combination of fescue (*Lolium arundinaceum*), ryegrass (*Lolium multiflorum*L.), bermuda grass (*Cynodon dactylon* L.) (White et al., 2006). After phytoremediation plants can then be subsequently harvested, processed and disposed. The plants usually influence rhizosphere microbial community which can play role in rhizoremediation and also numbers of plants are produce enzymes such as cytochrome P450 and peroxidase involved in the metabolism of n-alkanes (Vega-Jarquin et al., 2001; Lee et al., 2008). Phytoremediation is limited to the surface area and depth occupied by the roots. Secondly, the time required is also more as the plants involved grow slowly

(Kuiper et al., 2004). (Kuiper et al., 2004; Vega-Jarquin et al., 2001; Lee et al., 2008; White et al., 2006; Trapp et al., 2001; Glick, 2003; Pradhan et al., 1998)

2.6.3.2: Bioremediation Using Microorganisms

Microbial bioremediation refers to treatment processes that use microorganisms such as bacteria, fungi, yeast or their enzymes to break down hazardous substances into less toxic or nontoxic substances thereby restoring the contaminated site (Bhatnagar and Kumari, 2013). Land farming, composting, biopiling, slurry bioreactors, natural attenuation, bioventing, biosparging, biostimulation and bioaugmentation are different techniques which employ microorganisms for bioremediation (Kumar et al., 2011).

2.6.3.3: Different Microbial Bioremediation Techniques

Excavation involves removal of contaminated soils from sites with subsequent treatment either by land farming, composting, and biopiling or slurry bioreactors. Land farming involves the spreading of excavated contaminated soils in a thin layer on the ground surface of a treatment site and stimulating aerobic microbial activity within the soils through aeration and/or the addition of nutrients, minerals, and water/moisture (Paudyn et al., 2008).

2.7: Crude Oil Degrading Bacteria

The use of biological material like bacteria, fungi gained impetus in remediation of crude oil contaminated soil since its discovery in oil degradation and mineralization. But the screening and its use is varied and depends on the nature of the crude oil contamination (Roy et al., 2013; Yenn et al., 2014). Therefore, screening of microbes capable in degrading crude oil for bioremediation purpose is being continued. Studieshave reported different oil degrading bacteria belong to the genus *Bacillus, Bravibacillus, Achromobacter, Pseudomonas* etc. (Lal and Khanna, 1996; Roy et al., 2013). Assam is the potential source for different crude oil contamination environment

hence it is predicted that this environment could be good habitats for potential crude oil utilizing microbes.

Many heterotrophic bacteria are able to utilize hydrocarbons as a source of carbon and energy (Nichols et al., 1996). The most commonly used heterotrophic soil bacteria are *Pseudomonas* and *Arthrobacter*. They are capable of breaking down hydrocarbons through various metabolic pathways (Morelli et al., 2005) and breakdown of the hydrocarbon compounds depends on solubility and concentration of hydrocarbons. At low concentrations of hydrocarbon, all fractions are to be expected of being attacked by bacteria. However, at high concentrations, only those fractions most susceptible to degradation will be broken down. Also the concentration of contaminants will affect the number of organisms present.

Although numerous bacteria are able to metabolize organic contaminants, a single bacterium does not have the metabolic potential to degrade all or even most of the organic compounds in a polluted soil. The genetic information of more than one organism is necessary to degrade the complete mixtures of crude oil compounds present in the contaminated region, for this reason mixed microbial community have the most influential degradation potential than pure culture (Olajire et al., 2014). As reported by Bouchez et al. (1995) mixed culture of microorganisms may possibly enhance PAH utilization since the intermediatory biotransformation products of one microorganism may serve as substrate for catabolism and growth for other bacteria. As mentioned above degradation of crude oil seems to involve a consortium of microorganisms, including both eukaryotic and prokaryotic such as *Acinetobacter, Actinobacter, Arthrobacter, Bravibacillins, Berijerinckia, Cellulosimicrobium, Flavobacterium, Microbacterium, Mycobacterium, Enterobacter, Mycococcus, Nocardia, Pseudomonas* etc. Among the crude oil mineralizing organisms usually the gram negative bacteria, most of them belong

to the genus *Pseudomonas* are common in degradation of polycyclic aromatic hydrocarbon like Naphthalene. Besides these, studies have also been reported the bacteria from the genera *Corynbacterium, Aeromonas, Rhodococcus* and *Bacillus* (Neilson and Allard, 1998; Annweiler et al., 2000). On the other hand, a few organisms are capable of degrading fluoranthene which includes mainly bacteria from the genera *Mycobacterium* and *Alcaligenes*. Studies have also been conducted on strains from *Pseudomonas, Aeromonas, Arthrobacter, Sphingomonas, Mycobacterium* and *Nocardia* on phenanthrene degradation (Cerniglia et al., 1984; Pinyakong et al., 2000).

2.8: Bacterial Genome

The genome of an organism refers to its entire complement of genes contained in the DNA of its chromosomes. The bacterial genome is usually contained in a circular DNA molecule which is super coiled and localized within the nucleoid of the cell. There are exceptions, as some bacteria have two or more chromosomes and some chromosomes may be linear. Among medically important bacteria, *Vibrio, Burkholderia, Leptospira* and *Brucella* species are those with two or more chromosomes, while *Borrelia burgdorferi* has its genome in a linear chromosome (Guzman et al., 2008). Most bacterial genomes are less than 5 MB, although a few, such as *Bacillus megaterium*, may be as large as 30 MB (Allen et al., 2006). The major pattern in bacterial genome size is that, on average, free-living species have larger genomes than parasitic species which in turn have larger genomes than obligate pathogens. Bacterial genomes vary greatly between species in terms of nucleotide composition: The G+C content may vary locally within a genome, but it is relatively uniform within a bacterial genus or species, ranging from around 25% in *Mycoplasma* sp. to around 75% in some *Micrococcus* species (Guzman et al., 2008).

On the average, a typical bacterial genome has about 2,500 genes, which are maintained in a certain genomic architecture through selective compression, rather than

through a random succession of genes (Guzman et al., 2008; Allen et al., 2006). The genome of bacteria encodes all the biochemical functions that are necessary for survival. Additionally, pathogenic bacteria may carry genetic features required for virulence, while non-coding regions are also located in the bacterial genome. Characteristically, bacterial genes may be organized into operons, which refer to a group of genes located adjacent to one another, and are functionally related. An example of an operon is the lactose operon in *Escherichia coli*, which contains three genes involved in the conversion of lactose, a disaccharide into monosaccharide unit glucose and galactose (Jacob et al., 1961).

Bacterial genomes are dynamic, and are exposed to various genetic events, including, mutations, duplications, inversions, transpositions, recombination, insertion, and deletions. Gene acquisition through horizontal gene transfer is probably the mechanism having the greatest impact on the organism's lifestyle by conferring a novel metabolic capacity, such as acquisition of antibiotic resistance genes, adaptation of various environmental condition and virulence factors (Juhas et al., 2009).

In some bacterial cells, apart from the genome, there may be extra chromosomal DNA molecules referred to as plasmids. Sometimes, the distinction between a megaplasmid and a second chromosome may not be clear. Generally, plasmids are circular and double stranded, and replicate independently of the bacterial chromosome. *Plasmids facilitate* horizontal gene transfer within a microbial population of microbes and typically provide a selective advantage under a given unfavorable environmental condition (Barnett et al., 2001).

Bacterial genome sequencing was done long year back, from NCBI in 2014 reported there are more than 30,000 sequenced bacterial genomes currently publically available in database. During this period, the powerful combination of genome sequencing and bioinformatics driven analysis of sequence data has transformed our

understanding of how bacteria function, evolve and interact with each other, with their hosts, and with their surroundings, while also providing numerous avenues for translational impact (Loman et al., 2015). Sequence-based analyses have delivered unexpected insights into microbial diversity from strains to super-phyla and have allowed us to explore microbial communities. This bacterial genome sequencing encompasses three technological revolutions such as whole-genome shotgun sequencing, high-throughput sequencing and single-molecule long-read sequencing (Loman et al., 2015).

2.8.1: *Pseudomonas aeruginosa* Genome

In 19th Century, Dr. Migula, Professor at Karlsruhe Institute of Germany, first proposed the name Pseudomonas and it read as cells with polar organs of motility, with development of spores in a few categories (Palleroni, 2010). In 1926, Den Dooren de Jong stressed on microbes of soil and featured the extreme adaptability of *Pseudomonas* (Palleroni, 2010). It was represented by unicellular rods, with the long axis curvedor straight, motility by one or more polar flagella, Gram-negative, non-spore forming, sheaths, or stalks (Stanier et al., 1966). The respiration is the only process involved in energy-yielding metabolism and all species utilize oxygen as a terminal oxidant, whereas some species use denitrification as an anaerobic respiratory system. All Pseudomonas species are chemoorganotrophs, while few are facultative chemolithotrophs which use H_2 as energy source. *Pseudomonas* has three subgeneric groups: fluorescent group having species *Pseudomonas aeruginosa, P. fluorescens, P. putida*; acidovorans group, with *P. acidovorans, P. testosterone*; and, alcaligenes group representing *P. alcaligenes, P. pseudoalcaligene, P. mutivorans, P. stutzeri,* and *P. maltophilia* (Stanier et al., 1966). *Pseudomonas fluorescens* is represented by seven biotypes denoted by the letters A, B, C, D, E, F, and G (Stanier et al., 1966). Consequently, DNA/RNA hybridization confirmed the presence of five diverse rRNA groups (rRNA groups I–V) (Palleroni et al., 1972;

Palleroni, 1993; Kersters et al., 1996) Pseudomonas rRNA group I contained *P. aeruginosa*, all the fluorescent species (*P. fluorescens, P. putida, P. syringae*), and some non-fluorescent species (*P. stutzeri, P. alcaligenes, P. pseudoalcaligenes,* and *P. mendocina*) (Palleroni et al., 1973).

Pseudomonas aeruginosa is an opportunistic pathogen frequently responsible for nosocomial infections. It is the main pathogen associated with respiratory tract infection in cystic fibrosis patients, and it can cause a wide variety of infections among compromised hosts (Stover et al., 2000). Infections in healthy individuals can also occur as keratitis, otitis and others. *Pseudomonas aeruginosa* can be detected in human and animal faecal samples and has been identified as an animal pathogen responsible for ocular infections in dogs and as an occasional cause of bovine mastitis (Stanier et al., 1966; Palleroni, 2010). It is a metabolically versatile Gram-negative bacterium described as a ubiquitous micro-organism able to colonize a wide variety of environments. It has been frequently isolated from water sources including rivers, sea, water, bottled and tap waters and wastewaters. Its isolation from ornamental plants or vegetables as well as its detection in hydrocarbon-contaminated environments or agricultural lands has also been reported (Stanier et al., 1966; Palleroni et al., 1973; Diggle et al. 2020).

2.8.2: *Enterobacter* Genome

*Enterobacter*is rod-shaped and gram-negative bacteria that are classified as facultative anaerobes, which means that they are able to thrive in both aerobic and anaerobic environments (Sanders et al., 1997). These are ubiquitous in nature and their presence in the intestinal tracts of animals results in their wide distribution in soil, water, sewage and also found in plants. In humans, multiple *Enterobacter* species are known to act as opportunistic pathogens including *E. cloacae*, *E. aerogenes*, *E. gergoviae* and *E. agglomerans* (Sanders et al., 1997; Wu et al., 2018). Pathogenic *Enterobacter* can cause

any of a variety of conditions, including eye and skin infections, meningitis, bacterial blood infection, pneumonia, and urinary tract infections. In many instances, illness caused by *E. cloacae* or by *E. aerogenes* is associated with exposure to the organisms in nosocomial settings, such as hospitals or nursing homes (Wu et al., 2018). Free-living *Enterobacter* are capable of nitrogen fixation. Certain species, notably *E. cloacae*, are involved in symbiotic nitrogen fixation in plants and have been isolated from the root nodules of certain crops, such as wheat and sorghum, and rhizosphere of rice, some enterobacter sp are found in hydrocarbon-contaminated environments also (Swamy et al. 2016; Xu et al., 2018).

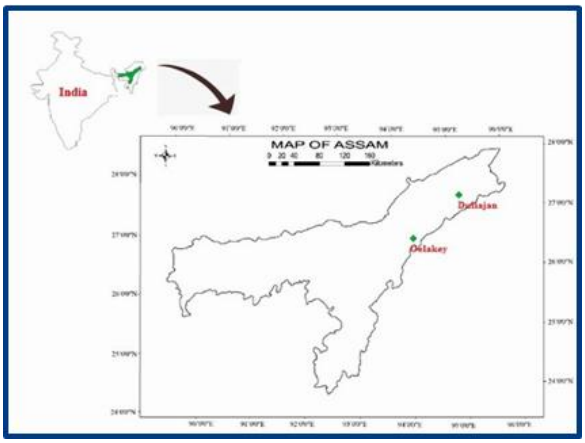
CHAPTER - 3

Materials and Methods

3.1: Section I: Soil sample analysis

3.1.1: Soil sampling and characterization

All the crude oil contaminated soil samples were collected from different areas of North East (NE) India mainly Duliajan and Gelakey, Assam, India (Figure 3.1) and immediately samples were stored at 4^{o}C. Before chemical and biological analysis, soil samples were sieved to separate large particles like plant parts (roots, stem, and leave), cobbles, pebbles etc. Soil *p*H was estimated in soil: water suspension (1:2.5) using *p*H meter (Eutech, Malaysia), while moisture content were determined by drying the soil samples at 70^{o}C until a constant weight was obtained. Soil conductivity was determined in soil suspension using digital conductivity meter (IKON, India). Total soil organic carbon (SOC) was determined according to Walkley-Black, total soil nitrogen (N) by Kjeldahl digestion, while phosphorus (P) using phospho molybdic acid and potassium (K) by Flame photometry respectively (Jackson, 1973).



Duliajan oil degrading area **Gelakey oil degrading area**

Figure 3.1: Crude oil contaminated soil sites of Duliajan and Gelakey, Assam in North East (NE) India

3.1.2: Soil Enzyme Activity

3.1.2.1: Dehydrogenase Activity

Dehydrogenase activity was estimated by using 2-3-5 Triphenyltetrazolium chloride (TTC) red technique of Casida (1977). Onegm fresh soil was taken in a test tube.

The soil was then mixed with 0.1 gm of $CaCO_3$ and 1 ml of 1% TTC solution. The tubes are incubated at 30^{o}C for 24 hrs in an incubator. The resulting slurry was transferred on Whatman no 1 and extracted with successive aliquots of concentrated methanol. The filtrate volume was made up to 50 ml by adding methanol. The absorbance of the filtrate was measured at 485 nm by using UV-Vis Spectrophotometer (220) setting concentrated methanol as a blank. The activity was represented in terms of concentration of Triphenyl Formazan (TPF) in methanol. Dehydrogenase activity of per gram soil was expressed in terms in terms of µg TPF per gm dry soil per 24 hrs.

$$\text{TTC (Water soluble colourles)} \xrightarrow{\text{Dehydrogenase}} \text{TPF (red, water soluble)}$$

3.1.2.2: Phosphatase Activity

Phosphatase activity was determined as per by Tabatabami and Bremnerj (1969). Onegm of air dried soil taken into a 50 ml conical flask. Then to it 4 ml of universal buffer (*p*H 7), 0.25 ml of toluene and 1 ml of 0.115 ml M p-nitro phenyl phosphate solution was added to the flask. The flask was swirled for few seconds and then incubated at 37^{o}C for 1 hrs in the incubator. After incubation, 1 ml of 0.5 M $CaCl_2$ and 0.5 M NaOH solution was added to the mixture. The soil suspension was filtered through Whatman filter no 1 filter paper and the optical density (OD) of the filtrate were measured at 430 nm in Specord 210 UV-Vis spectrophotometer. A blank was maintained similarly without soil sample. The Phosphatase activity was measured in terms of concentration of PNP in each sample from a standard curve of PNP in water and is expressed as mole of PNP released per gm dry soil per hour.

$$\text{P-NITROPHENYL PHOSPHATE} \xrightarrow{\text{Phosphatase}} \text{P-NITROPHENOL (YELLOW)} + \text{PHOSPHATE}$$

3.1.2.3: Urease Activity

Urease activity of soil samples was measured following by McGarity and Mayers, (1969) method. One g of soil was taken in a 100 ml volumetric flask, add 1 ml of toluene

and allowed to stand for 15 min. Thereafter 10 ml buffer (*p*H 7), add 5 ml of 10% urea solution and swirl the flask again for a few seconds. The flask was then placed in an incubator at 37°C for 3 hrs. A control was set with distilled water in place of urea solution. Post incubation, the contents was adjusted to 100 ml with distilled water, mixed thoroughly and filtered through Whatman no 5 filter paper. Now 0.5 ml of filtrate was taken in a 25 ml volumetric flask and 5 ml of distilled water was added to adjusted 30 ml. Then the mixture in the flask was treated with 2 ml phenolate solution and 1.5 ml sodium hypochlorite solution. The absorbance was read in a Specord 210 UV-Vis Spectrophotometer at 630 nm. The amount of $NH4^{+}$-N released was calculated by a reference calibrated curve and was expressed as NH^{+}-N mg per gm dry soil per 3 hrs.

$$\text{Urea} + H_20 \longrightarrow 2NH3+CO_2$$

$$NH_3+\text{Phenol}+\text{NaOCl} \longrightarrow \text{Indophenol Blue (630nm)}$$

3.1.3: Microcosm Study of Crude Oil Contaminated Soil

To test the efficacy of remediation of crude oil contaminated soil using the screened bacteria, simulated crude oil contaminated field plots as microcosm were created under green house condition. In brief, the soil collected from experimental garden of the institute was made to 8x3 m^2 microcosms with a depth of six inches. Each microcosm was mixed with Assam light crude oil-soil ratio of 3:1 (w/w) and left for 30 days. The zero days reading of the soil was recorded at the end of 30 days and biological treatment was performed. A total of 7 different bio-formulations, consisting nitrogen (N) as urea; potassium (K) as murate of potash; and phosphorous (P) as diammonium phosphate i.e., mixtures of nitrogen-phosphorus-potassium (NPK) was added in a ratio-60:40:40 ha^{-1} (~150,100,100 mg $microcosm^{-1}$); organic farmyard manure (OM), vermicompost (VC) at the rate of 80 kg ha^{-1} (~430 mg $microcosm^{-1}$). It is important to note that the cfu 10^{13} ml^{-1} observed in these bacteria is indeed very high which suggest

that these bacterial isolates can grow to a much higher density in growth medium. All the experiments were repeated 3 times with three replications in each treatment.

3.1.4: Sampling and Soil Analysis of Microcosms

Triplicate samples ~100 g were taken from each microcosm at an interval of 4 weeks from the day of treatments. The soil samples from each microcosm were then analyzed for TPH, *p*H, moisture content as described above and different soil enzyme activities (dehydrogenase, phosphatase and urease) as per standard protocol Bremner and Tabatabai (1969); Smith and Chalk (1980); Camina et al., (1998) was used. Soil respiration, evaporation rate, CO_2flux was analyzed by IRGA-CIRAS-2 (PP System, USA) attached with soil respiration chamber SRC-1. For measurement SRC-1was attached to main system CIRAS-2. The main system was then pre-set to area (cm^2) at 78.5, CO2 change at 60 ppm, the maximum amount of time from start of measurement to the end at 60 s. After successful completion of the set up the SRC-1 chamber was hold for ~15 s to flush out. After that the SRC-1 chamber was placed on the soil and allowed to measure. After completion of the measurement the data were recorded as soil-evaporation (gm^{L2}), respiration (gCO_2 m^{-2} h^{-1}) and temperature (C), respectively.

3.2: Section II: Extraction of Bacterial Strain

3.2.1: Isolation and Characterization of Bacterial Strain

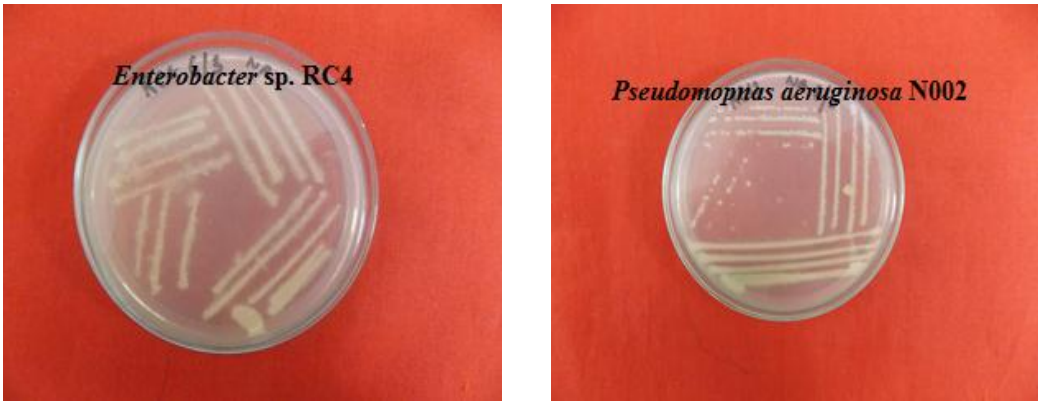
Pseudomonas aeruginosa N002 and *Enterobacter* sp. RC4 was isolated from crude oil contaminated soil of Duliajan and Geleky, Assam, India, using enrichment culture method in M1 mineral media (g L^{-1}) (4.0gm $NaNO_3$; 3.61gm Na_2HPO_4; 1.75 gm KH_2PO_4; 0.2gm $MgSO_4.7H_2O$; 0.01gm $FeSO_4$; 0.05gm $CaCl_2$, trace element solution 1 ml/L) with 2% crude oil as the sole carbon source. Details of isolation, identification, degradation ability of crude oil of *Pseudomonas aeruginosa* N002 was verified, and good

growth was observed to occur independently in all crude oil components, i.e., aliphatic fraction, aromatic fraction, asphaltene fractions.

3.2.2: Morphological Characterization of Bacterial Strain

3.2.2.1: Colony Appearance

Size, shape, color, texture, elevation of the colony, type of margin, consistency and translucent or opaqueness were determined for isolated strains (Figure 3.2 A& B).



A **B**

Figure 3.2: A. Culture plate of *Pseudomonas aeruginosa* N002 and B. *Enterobacter* sp. RC4. bacteria isolated from Duliajan and Gelakey, Assam of North East India

3.2.2.2: Gram Reaction

It is done to differentiate bacteria in to two principal group: gram-positive and gram- negative. For gram staining 24 h nutrient agar slant cultures of the isolated strains are taken. Using sterile technique, a thin smear of bacterial culture is prepared. The reagents used are Crystal violet, Gram's iodine, 95% ethyl alcohol and safranin.

3.2.3: Biochemical Characterization of Bacterial Strain

3.2.3.1: Nitrate Reduction Test

Bacteria were inoculated in Trypticase nitrate broth. Then each isolate was gently inoculated in the medium to distribute it throughout the tube. Then tubes were incubated for 24 to 48 h at 37º C. Following incubation of the organisms, an organism's ability to reduce nitrates to nitrites is determined by the addition of two reagents: Solution A,

which is sulfanilic acid, followed by solution B, which is di-methyl-alpha-naphthylamine. Following reduction, the addition of solutions A and B will produce an immediate cherry red color.

3.2.3.2: Indole Production Test

Pure bacterial culture must be grown in sterile tryptophan or peptone broth for 24-48 hrs before performing the test. Following incubation, add 5 drops of Kovac's (Isoamyl alcohol, para-Dimethylaminobenzaldehyde, concentrated hydrochloric acid) reagent to the culture broth. A positive result is shown by the presence of a red or red-violet color in the surface alcohol layer of the broth. A negative result appears yellow.

3.2.3.3: MR-VP (Methyl red-Voges- Proskauer Test)

Four ml of MR-VP medium was prepared by dissolving 0.1 g of methyl red in 300 ml of 95% ethyl alcohol which was then diluted to a total volume of 500 ml with distilled water. The agar tubes were inoculated with the test bacterial isolate and incubated at 37º C for 48 h. If the reagent turns red, which is due to accumulation of acidic product of fermentation of glucose, it indicates a positive test. For VP test, 0.5ml of 5% alpha napthyl certain was added to another half of the 2 ml culture. The tube was shaken thoroughly and allowed to stand for 5-10 minutes. Positive test was indicated by the appearance of a pink color which is due to formation of acetone.

3.2.3.4: Citrate Utilization Test

The Simmons agar media comprised of Ammonium dihydrogen phosphate 1.0 g, Dipotassium phosphate 1.0 g, Sodium chloride 5.0g, Sodium citrate 2g, Magnesium sulphate 0.2 g, Agar 15 g, Bromthymol blue 0.08 g. Simmons agar slants were streaked inoculated with the bacterial isolates and incubated at 37º C. After 7 days of incubation blue color appearance in the tube indicated the utilization of citrate as the sole carbon source by the test bacteria

3.2.3.5: Starch Hydrolysis

Nutrient agar plate supplemented with 0.2% soluble starch was spot inoculated in the culture plate with the test bacteria. After incubation for 2 days at 37º C the plate was flooded with iodine solution. A positive test was indicated with clear haloes around the colony.

3.2.4: Identification of 16S rDNA Sequencing

Bacterial genomic DNA was extracted from the isolates obtained using GenElute bacterial genomic DNA kit (Sigma Aldrich). PCR amplification of the genomic DNA was carried out using the universal 16S rDNA primers fD1 AGAGTTTGATCCTGGCTCAG and rP2 ACGGCTACCTTGTTACGACTT (Weisburg et al., 1991) with reaction conditions set at 94 °C for 5 min, followed by 35 cycles of 94 °C for 30 s, 52 °C for 45 s and 72 °C for 1.30 min and then final extension at 72 °C for 10 min. 16S rDNA amplicon from each strain was sequenced, and the data were searched using NCBI-BLAST (http://www.ncbi.nlm.nih.gov/blast) search tool for identification of the strain type.

3.3: Section III: Bioinformatics Analysis of Bacterial Genome

3.3.1: Whole Genome Sequencing

The whole genome sequencing of *P. aeruginosa* N002 and *Enterobacter* sp. RC4 was obtained by whole-genome shotgun sequencing using Ion Torrent method (Kube et al., 2013; Nie et al., 2012). Sequencing was carried out as per the Ion 316 chip sequencing protocol. The genome sequence was assembled by using GS *de novo* Assembler version 2.6. The total number of reads generated using reference based approach was 1,074,106, with a mean length of 123 bp. The mapping of Ion Torrent 2.0 high-quality reads on the reference genome was performed using TMAP v0.0.28, to get consensus sequence.

3.3.2: Genome Analysis

The genome of *Pseudomonas aeruginosa* N002 and *Enterobacter* sp. RC4 was submitted to Integrated Microbial Genomes (IMG) server (http://img.jgi.doe.gov) of Joint Genome Institute (JGI) and Rapid Annotation using Subsystem Technology (RAST) server (http://rast.nmpdr.org/) for deep analysis and genome comparisons (Mougous et al., 2006). The rRNA genes, tRNA genes, total ORFs, GC and circular map of the genomes was retrieved from the server. For phylogenetic relationship the sequence of the genome was compared with reported genome by NCBI-BLAST tool (http://blast.ncbi.nlm.nih.gov/ blast/treeview.cgi) (Jeukens et al., 2013). The similar sequence was downloaded and phylogenetic tree was prepared using Mega6 software. Frame shifts identified by the NCBI submission check tool, were manually checked. Insertion sequence (IS) elements and transposons were identified using the IS finder database (http://www-is. biotoul.fr/). Transporter and Genomic Islands were analyzed using SeqWord Sniffer (Jeukens et al., 2013) and Island Viewer (Head et al., 2003; Mougous et al., 2006; Maddocks et al., 2008), respectively. A one-sample t-test was used to evaluate the statistically significant differences of gene abundance in each COG category between. The taxonomic position of the genomes strain was performed one hand by comparing with already sequenced genomes of other same strains in IMG database.

3.3.3: Phylogenetic Analysis

To determine evolutionary relationship among the bacterial isolates as well as the cultured 16S rRNA sequences, the gene sequence were compared against those obtained from GenBank database using NCBI-Blast as mentioned above and SeqMatch tool (Cole et al., 2007). Multiple sequence alignment was carried out using ClustalW (http:// www.ebi.ac.uk/clustalw/) sequence alignment tool. Phylogenetic tree was then constructed using the neighbor joining method of MEGA6 program (Tamura et al., 2013).

Branching orders of the trees were ascertained and compared using distance-based neighbor-joining algorithms with Kimura two-parameter corrections to improve the reliability of internal branches (Kimura, 1980).

3.3.4: Statistical Analysis

All the statistical analysis was performed by using statistical analysis tools OriginPro8, SPSS 16, and Excel 2010.

3.3.5: Nucleotide sequence accession number

The whole genome shotgun project and 16s rRNA sequence has been deposited in DDBJ/EMBL/GenBank with accession number ALBV00000000.2 and, JX035794.1 for *Pseudomonas aeruginosa* N002 and PTLU00000000, KJ499800 for *Enterobacter* sp. RC4 genome sequence and 16s rRNA.

CHAPTER - 4

Results

In the present study whole genome sequencing and genome analysis was done for two bacteria i.e., *Pseudomonas aeruginosa* N002 and *Enterobacter* sp. RC4. Sequences were submitted to Genbank after suitable analysis. The whole genome shotgun project and 16s rRNA sequence has been deposited in DDBJ/EMBL/GenBank with accession number ALBV00000000.2 and, JX035794.1 for *Pseudomonas aeruginosa* N002 and PTLU00000000, KJ499800 for *Enterobacter* sp. RC4 genome sequence and 16s rRNA. Present works was stated with six defined objectives which are presented below. For convenience all the results are presented in tabulated form together for both the species for comparative understanding.

4.1: Identification of the strain

Gram straining of *P. aeruginosa* N002 and *Enterobacter* sp. RC4 configure as gram –ve , positive to nitrate reduction and citrate utilization, molecular identification of 16s rDNA sequences confirmed that the study bacteria were *Pseudomona aeruginosa* designate as N002 and *Enterobacter* sp. RC4. The phylogenetic trees of *P. aeruginosa* N002 and *Enterobacter* sp. RC4 are shown in figure 4.1.A and 4.1.B.

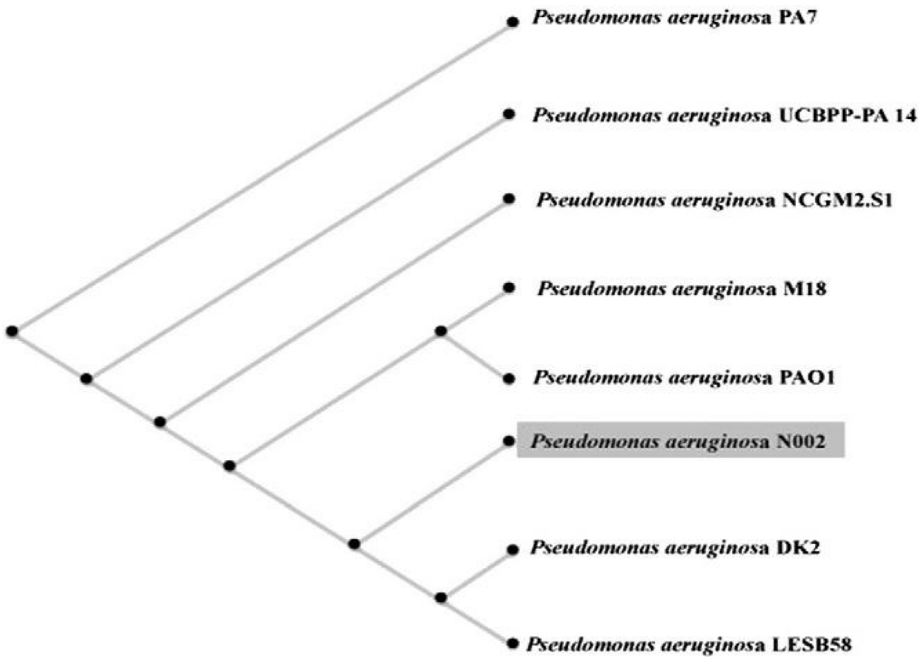


Figure 4.1.A: Phylogenetic tree of *Pseudomonasaeruginosa* N002

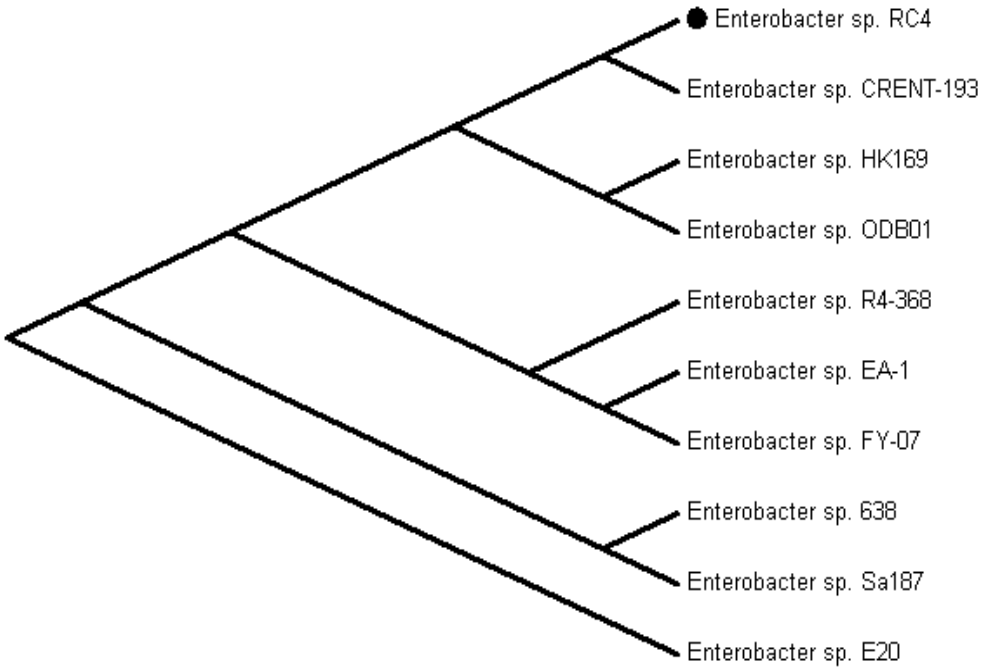


Figure 4.1.B: Phylogenetic tree of *Enterobacter* sp. RC4

4.2: Genome characterization

The genome features of both the bacteria are presented in Table 4.1. Genome sequence of *P. aeruginosa* N002 indicates presence of total 65,37,648 bp with 5629 CDS, 62.36 % GC, 135 RNA genes while the genome of *Enterobacter* sp. RC4 having a total of 5,029,294 bp with 4,898 CDS, 54.77 % GC, 186 RNA genes.

Table 4.1: Genomics features of *Pseudomonas aeruginosa* N002 and *Enterobacter* sp. RC4 genome

Features	*Pseudomonas aeruginosa* N002	*Enterobacter* sp. RC4
DNA, total number of bases	65,37,648	5,029,294
DNA coding number of bases	54,91,038	45,46,392
DNA G+C Content (%)	62.36	54.77
Genes total number	5,764	4,984
Protein coding genes	5,629	4,898
Pseudo Genes	0	202
RNA genes	135	186
rRNA genes	12	8
5S rRNA	4	6
16S rRNA	4	1
23S rRNA	4	1
tRNA genes	63	74
ORF number	11,038	10,021
CDS	5,629	4,898

4.3: Cluster of orthologous group (COG)

The circuler chromosome map was generated on the basis of COG category which generated from Integrated Microbial Genomes (IMG) database deposited of *P. aeruginosa* N002 (ALBV00000000.2) and *Enterobacter* sp. RC4 (PTLU00000000) are shown in figure In the circle from outside to the center first circle indicate genes on forward strand, second circle indicate genes on reverse strand, third circle indicate RNA genes (rRNA are Red, tRNA are Green, Other RNAs are Black), fourth circle indicate

GC content of the genome and fifth circle indicate GC skew of genome sequence (Figure 4.2.A and 4.2.B). Altogether twenty six distinct functional orthologous group of genes was identified witjh a variation of genes abundance of 0.06 to 11.62 % in P. aeruginosa while 0.02 to 9.52 %.

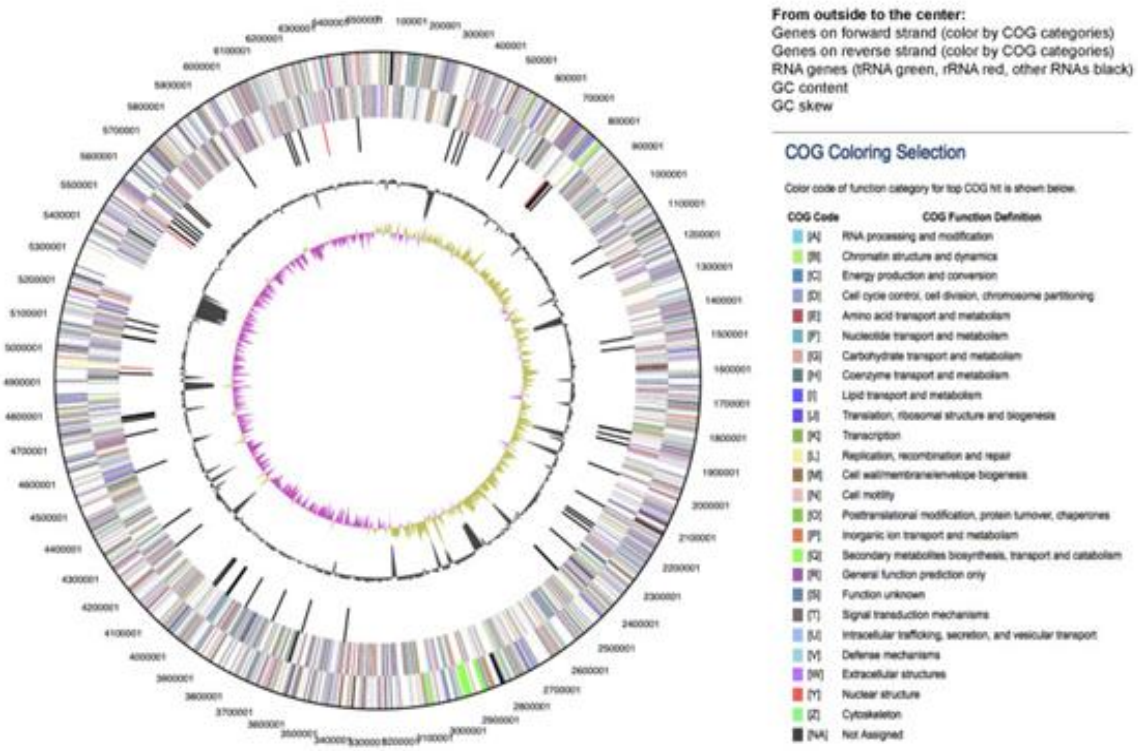


Figure 4.2.A: Circular chromosome map of *P. aeruginosa* N002

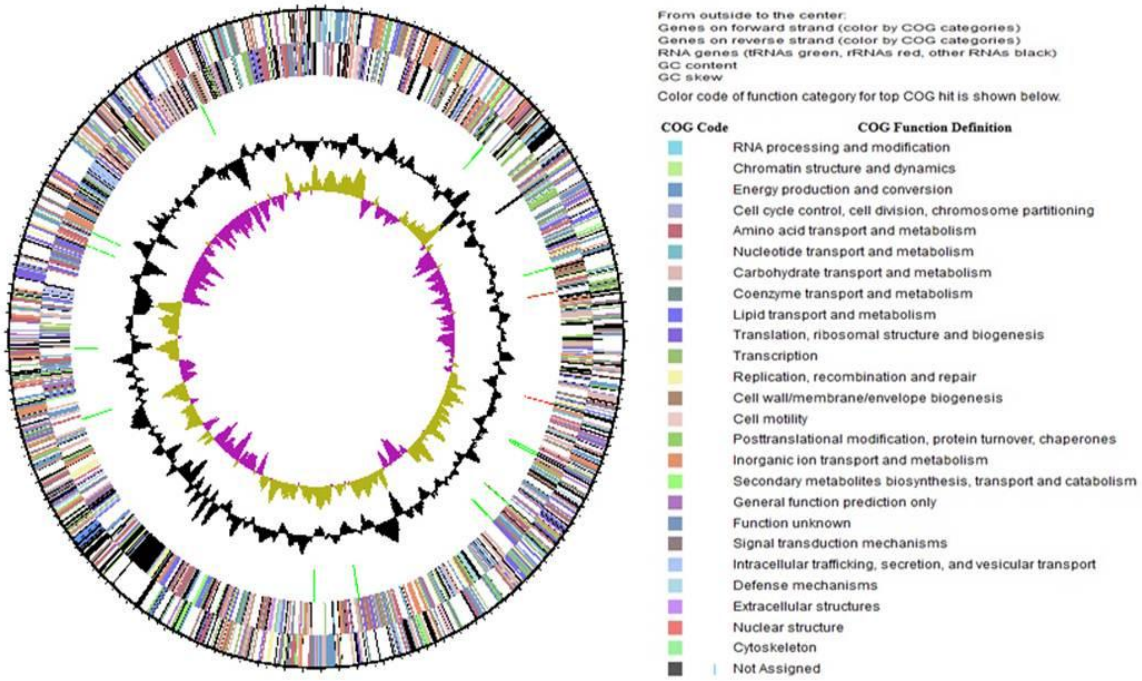


Figure 4.2.B: Circular chromosome map of *Enterobacter* sp. RC4

4.4: Comparision of COG of hydrocarbon utilization bacteria

Comparison of the N002 and RC4 genome with other available genomes of hydrocarbon degrading bacteria in IMG bacteria genome databaseand one-sample t-test was used to evaluate possible significant differences of the gene abundances of each COG categories between *P. aeruginosa* N002 and *Enterobacter* sp. RC4 with other genomes in the IMG genome database. From the comparison it was observed that strain *P. aeruginosa* N002 have appreciably higher number of amino acid metabolism (E), cell motility (N), unknown functionl gene (S), Inter cellular trafficking, secretion and transport (U), post translation modification, protein (O), signal transduction mechanism (T) and Transcription (K). On the other in *Enterobacter* sp. RC4 carbohydrate transport and metabolism (G), cell cycle control, cell division, chromosome portioning (D), cell wall membrane envelope biogenesis (M), coenzyme transport and metabolism(H),

inorganic ion transport and mechanism(P), Translation, ribosomal structure and biogenesis (J), Posttranslational modification, protein turnover, chaperones (0), Transcription (K) and Signal transduction mechanisms (T) are higher (Table 4.2).

Table 4.2: Comparison of COG categories between *P. aeruginosa* N002 and *Enterobacter* sp. RC4 with other genomes of hydrocarbon degrading bacteria in IMG database

COG list	*P. aeruginosa* N002			*Enterobacter* sp. RC4		
	Gene abundance (%)	Mean (%)	P-value (2-tailed)	Gene abundance (%)	Mean (%)	P-value (2-tailed)
Amino acid transport and metabolism (E)	9.35	8.652	3.72E-15	8.8	8.625	3.65E-15
Carbohydrate transport and metabolism (G)	4.31	4.831	5.41E-13	9.52	5.092	1.99E-11
Cell cycle control, cell division, chromosome partitioning (D)	0.67	0.961	1.52E-11	1.03	0.979	7.06E-12
Cell motility (N)	2.92	1.837	6.61E-07	3.95	1.888	1.08E-06
Cell wall/ membrane/ envelope biogenesis (M)	4.9	5.069	1.56E-13	6.36	5.142	1.91E-13
Chromatin structure and dynamics (B)	0.06	0.337	2.50E-01	0	0.318	0.25397
Coenzyme transport and metabolism (H)	4.03	4.122	3.53E-15	4.9	4.166	4.32E-15
Defence mechanisms (V)	1.41	1.671	1.68E-05	2.34	1.717	1.27E-05
Energy production and conversion (C)	6.3	6.715	4.80E-16	5.4	6.67	8.98E-16
Function unknown (S)	10.04	8.922	9.36E-16	5.24	8.682	9.36E-15
General function prediction only (R)	11.62	11.715	2.80E-16	7.05	11.486	3.54E-15
Inorganic ion transport and metabolism (P)	5.69	4.982	2.19E-12	6.22	5.008	2.52E-12
Intracellular trafficking, secretion, and vesicular transport (U)	3.24	2.764	1.58E-06	2.18	2.711	2.07E-06
Lipid transport and metabolism (I)	4.43	5.226	1.63E-10	3.28	5.168	2.86E-10
Nucleotide transport	2.04	2.263	4.44E-13	2.15	2.268	4.04E-13

and metabolism (F)						
Posttranslational modification, protein turnover, chaperones (0)	3.8	3.585	4.67E-14	3.9	3.59	4.73E-14
Replication, recombination and repair (L)	2.42	4.309	1.19E-09	3.4	4.358	6.61E-10
RNA pprocessing and modification (A)	0.06	0.048	3.51E-02	0.02	0.046	0.04285
Secondary metabolites biosynthesis, transport and catabolism (Q)	3.09	3.716	5.28E-08	2.03	3.663	8.47E-08
Signal transduction mechanisms (T)	6.56	5.402	4.09E-11	5.74	5.361	3.90E-11
Transcription (K)	9.27	8.230	1.14E-13	8.46	8.19	1.10E-13
Translation, ribosomal structure and biogenesis (J)	3.8	4.736	1.14E-12	5.83	4.837	8.07E-13

4.5: COG comparision with same species

Comparision of COG of *P. aeruginosa* N002 and *Enterobacter* sp. RC4 with the same group is described in the (Table 4.3). In both the species there were appreciably higher aboundance of COG. The higher abundance of reported common COG in other and within the hydrocarbon degrading genes signifies that the strain *P. aeruginosa* N002 and *Enterobacter* sp. RC4 sare the appropriate microbes to degraded hydrocarbon.

Table 4.3: Comparison of COG categories between *P. aeruginosa* N002 and *Enterobacter* sp. RC4 within the same group in available genome in IMG database.

COG list	***P. aeruginosa* N002**			***Enterobacter sp* RC4**		
	Gene abundance (%)	**Mean (%)**	**P-value (2-tailed)**	**Gene abundance of (%)**	**Mean (%)**	**P-value (2-tailed)**
Amino acid transport and metabolism (E)	9.35	9.288	2.38E-13	8.8	9.364	5.59E-11
Carbohydrate transport and metabolism (G)	4.31	4.228	1.45E-13	9.52	9.394	2.08E-09
Cell cycle control, cell division, chromosome partitioning (D)	0.67	0.741	1.28E-09	1.03	1.017	5.33E-12
Cell motility (N)	2.92	2.915	3.82E-13	3.95	2.872	1.14E-07

Cell wall/membrane/ envelope biogenesis (M)	4.9	4.932	1.84E-14	6.36	6.018	2.25E-08
Chromatin structure and dynamics (B)	0.06	0.058	5.16E-10	0.00	0.00	0.00
Coenzyme transport and metabolism (H)	4.03	3.942	3.32E-13	4.9	5.115	1.36E-13
Defence mechanisms (V)	1.41	1.526	1.65E-08	2.34	2.251	4.13E-11
Energy production and conversion (C)	6.3	6.117	1.95E-13	5.4	5.562	1.00E-14
Function unknown (S)	10.04	9.828	3.15E-13	5.24	5.587	1.42E-07
General function prediction only (R)	11.62	11.571	7.59E-13	7.05	7.31	3.67E-13
Inorganic ion transport and metabolism (P)	5.69	5.718	1.89E-12	6.22	6.332	5.04E-13
Intracellular trafficking, secretion, and vesicular transport (U)	3.24	3.401	2.72E-11	2.18	1.834	5.97E-08
Lipid transport and metabolism (I)	4.43	4.426	7.45E-14	3.28	3.16	1.15E-12
Nucleotide transport and metabolism (F)	2.04	2.037	3.22E-13	2.15	2.316	2.73E-10
Posttranslational modification, protein turnover, chaperones (0)	3.8	3.745	7.67E-15	3.9	4.049	1.04E-11
Replication, recombination and repair (L)	2.42	2.96	2.61E-07	3.4	3.338	9.16E-13
RNA pprocessing and modification (A)	0.06	0.046	3.47E-05	0.02	0.036	0.00777
Secondary metabolites biosynthesis, transport and catabolism (Q)	3.09	3.0615	3.41E-12	2.03	2.086	2.81E-09
Signal transduction mechanisms (T)	6.56	6.423	8.59E-15	5.74	5.531	2.14E-13
Transcription (K)	9.27	9.358	2.04E-15	8.46	8.556	2.33E-12
Translation, ribosomal structure and biogenesis (J)	3.8	3.985	1.68E-06	5.83	6.35	3.60E-10

4.6: Insertion sequences (IS) of *P. aeruginosa* N002 and *Enterobacter* sp. RC4

Details of ISs in *P. aeruginosa* N002 and *Enterobacter* sp. RC4 is described in (Table 4.4). Where the genome sequence of *P. aeruginosa* N002 having a total of 19 IS sequence in 4 different IS family while the genome of *Enterobacter* sp. RC4 having a

total of 25IS sequences in 17 different IS family genes. From the large number of insertion in *P. aeruginos* N002 and *Enterobacter* sp. RC4 it can be infer that both the strain can adopt to evolve based on environmental condition.

Table 4.4: Insertion sequences predicted in *P. aeruginosa* N002 and *Enterobacter* sp. RC4

***P. aeruginosa* N002**			***Enterobater* sp. RC4**		
IS family	**Start**	**End**	**IS family name**	**Start**	**End**
IS3 ssgr IS407	280566	279718	IS4_ssgr_IS4	825314	822946
IS3 ssgr IS407	280853	280590	ISL3	493428	498666
IS3 ssgr IS3	298567	298875	ISL3	3715947	3710876
IS3 ssgr IS3	298989	299717	ISL3	490704	494250
IS91	1366492	1367388	ISL3	3717025	3713934
IS3 ssgr IS407	2475088	2474828	IS3_ssgr_IS407	4489368	4486248
IS3 ssgr IS407	2475855	2475229	IS5_ssgr_IS903	3995833	3992910
IS3 ssgr IS3	2491650	2491802	IS5_ssgr_IS5	3688725	3686033
IS3 ssgr IS3	2491793	2492077	IS5_ssgr_IS5	3688017	3685745
IS3 ssgr IS3	2492102	2492317	IS3_ssgr_IS3	4488146	4484978
IS3 ssgr IS3	2492344	2492493	IS1	4269828	4267131
IS3 ssgr IS407	4583705	4583211	ISNCY_ssgr_IS1202	4259293	4262650
IS3 ssgr IS3	4587841	4588149	ISNCY_ssgr_IS1202	4260700	4263094
IS3 ssgr IS3	4588263	4588991	IS481	4482476	4485632
IS3 ssgr IS407	4892235	4891849	IS481	2184237	2187060
IS3 ssgr IS407	4892399	4892142	IS3_ssgr_IS150	3698101	3701127
IS5	4917507	4916905	ISAs1	375787	373437
IS91	6217527	6218438	Tn3	3719043	3715043
			IS3_ssgr_IS2	3710459	3712811
			IS630	4483953	4486470
			IS3_ssgr_IS51	3997801	3995232
			IS3_ssgr_IS51	4268355	4266143
			IS110_ssgr_IS1111	3991566	3994590
			IS110_ssgr_IS1111	4010567	4013570

	IS110	4268980	4266423

4.7: Genomic Island (GI) analysis of *P. aeruginosa* N002 and *Enterobacter* sp. RC4

Genome sequence of *P. aeruginosa* N002 is having a total of 40 GI sequence and the genome *Enterobacter* sp. RC4 having a total of 68 GI sequences. Similarly large number of GIs presents in *P. aeruginosa* N002 and *Enterobnacter* sp. RC4 indicates the selective advantages to the cell, which enhanced the bacterium changes of survival or colonization of new niche (Table 4.5).

Table 4.5: Genomic Island (GI) analysis of *P. aeruginosa* N002 and *Enterobacter* sp. RC4

***Enterobacter* sp. RC4**			***Pseudomonas aeruginosa* N002**		
GI No	**Start**	**End**	**GI No**	**Start**	**End**
GI1	9,943	14,291	GI1	282346	324180
GI2	21,764	26,052	GI2	856435	881136
GI3	142,227	151,598	GI3	1183607	1210067
GI4	304,839	309,158	GI4	1293079	1345156
GI5	364,132	368,646	GI5	1627864	1651715
GI6	391,805	396,609	GI6	1753981	1776765
GI7	495,538	502,781	GI7	1912348	1975013
GI8	504,476	509,425	GI8	2018692	2053149
GI9	1,358,151	1,363,711	GI9	2467219	2502719
GI10	2,059,916	2,075,786	GI10	2671500	2732599
GI11	2,452,798	2,461,085	GI11	2848568	2884416
GI12	2,464,195	2,468,335	GI12	2912099	2949247
GI13	2,829,333	2,833,530	GI13	3167743	3202104
GI14	2,835,371	2,845,352	GI14	3503576	3528155
GI15	2,857,624	2,876,484	GI15	4129186	4165730
GI16	2,879,925	2,888,842	GI16	4299922	4326833
GI17	3,679,479	3,687,687	GI17	4338173	4369203
GI18	3,712,876	3,719,174	GI18	4393260	4415837

GI19	3,999,046	4,006,386	GI19	4572028	4599246
GI20	4,010,572	4,016,120	GI20	4743296	4769072
GI21	4,026,496	4,031,669	GI21	4853383	4906629
GI22	4,032,973	4,038,222	GI22	4912805	4940105
GI23	4,460,673	4,476,099	GI23	5243628	5393030
GI24	4,480,121	4,485,282	GI24	6191799	6214676
GI25	4,861,538	4,865,915	GI25	288,290	298,547
GI26	124,652	136,290	GI26	299,726	312,954
GI27	140,105	159,455	GI27	1,801,156	1,811,016
GI28	364,291	372,431	GI28	1,919,508	1,941,373
GI29	484,837	490,543	GI29	2,474,727	2,483,492
GI30	497,719	505,202	GI30	2,482,125	2,503,467
GI31	505,695	509,800	GI31	2,712,726	2,721,391
GI32	713,141	718,282	GI32	2,847,774	2,861,105
GI33	883,097	887,778	GI33	2,867,555	2,876,537
GI34	1,339,667	1,347,994	GI34	2,926,345	2,937,198
GI35	1,365,641	1,371,356	GI35	3,329,542	3,337,557
GI36	1,844,738	1,849,338	GI36	3,788,182	3,768,530
GI37	2,881,514	2,888,786	GI37	4,313,536	4,326,917
GI38	3,549,098	3,553,655	GI38	4,569,922	4,588,991
GI39	3,693,452	3,703,007	GI39	4,860,224	4,895,095
GI40	3,711,479	3,717,197	GI40	5,347,338	5,358,104
GI41	3,723,858	3,729,160			
GI42	3,993,738	4,006,260			
GI43	4,007,543	4,015,864			
GI44	4,029,141	4,035,512			
GI45	4,267,003	4,271,977			
GI46	4,320,129	4,329,377			
GI47	4,475,878	4,480,451			
GI48	4,483,451	4,492,399			
GI49	4,679,007	4,689,140			
GI50	4,828,943	4,835,234			

GI51	4,878,684	4,883,886	
GI52	4,900,174	4,909,038	
GI53	95,162	110,947	
GI54	123,897	160,745	
GI55	361,218	396,218	
GI56	484,837	507,669	
GI57	702,171	726,173	
GI58	877,850	884,825	
GI59	1,236,144	1,247,071	
GI60	2,880,458	2,889,337	
GI61	2,905,924	2,922,481	
GI62	3,684,082	3,703,007	
GI63	3,710,756	3,728,121	
GI64	4,032,289	4,037,740	
GI65	4,462,393	4,464,955	
GI66	4,479,703	4,488,518	
GI67	4,867,556	4,883,886	
GI68	4,897,614	4,910,093	

Further, Genomic Island (GI) of both the bacteria showing in different colour islandPeak in green, IslandViewer in red, IslandPath-DIMOB in skyblue, SIGI-HMM in Orange and SeqWord sniffer in Pink. (Figure 4.3.A and 4.3.B).

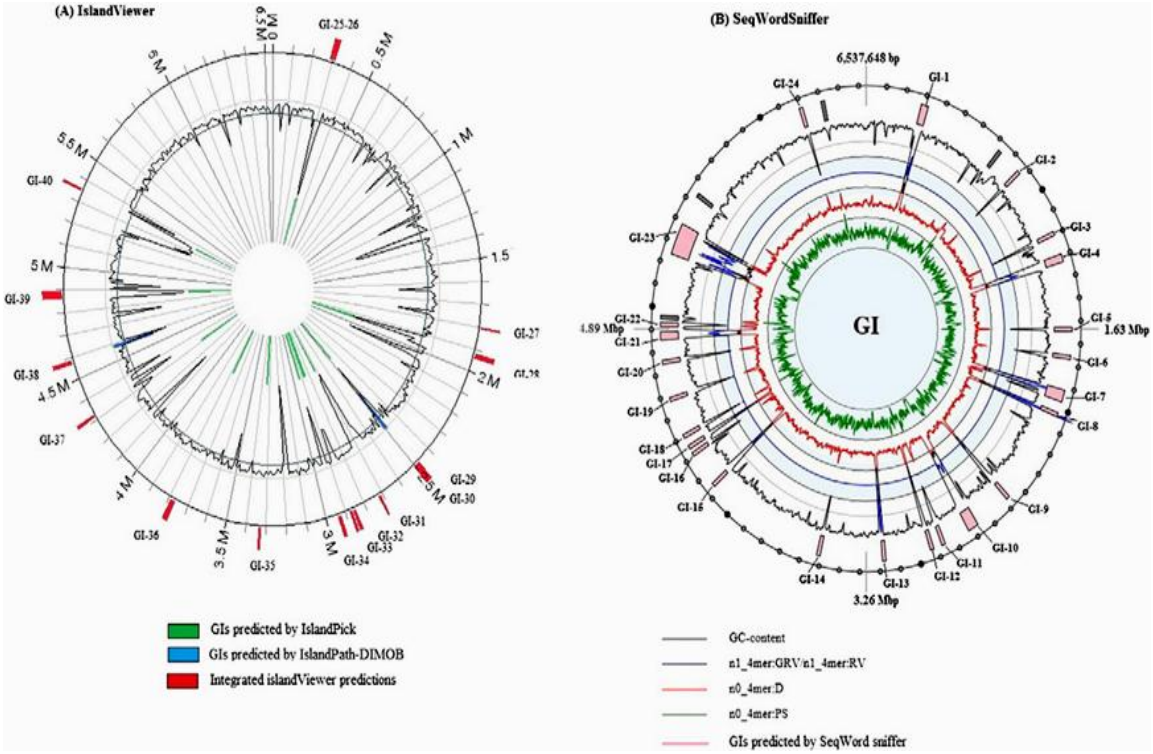


Figure 4.3.A: Genomic Island of *P. aeruginosa* N002 prediction by different methods

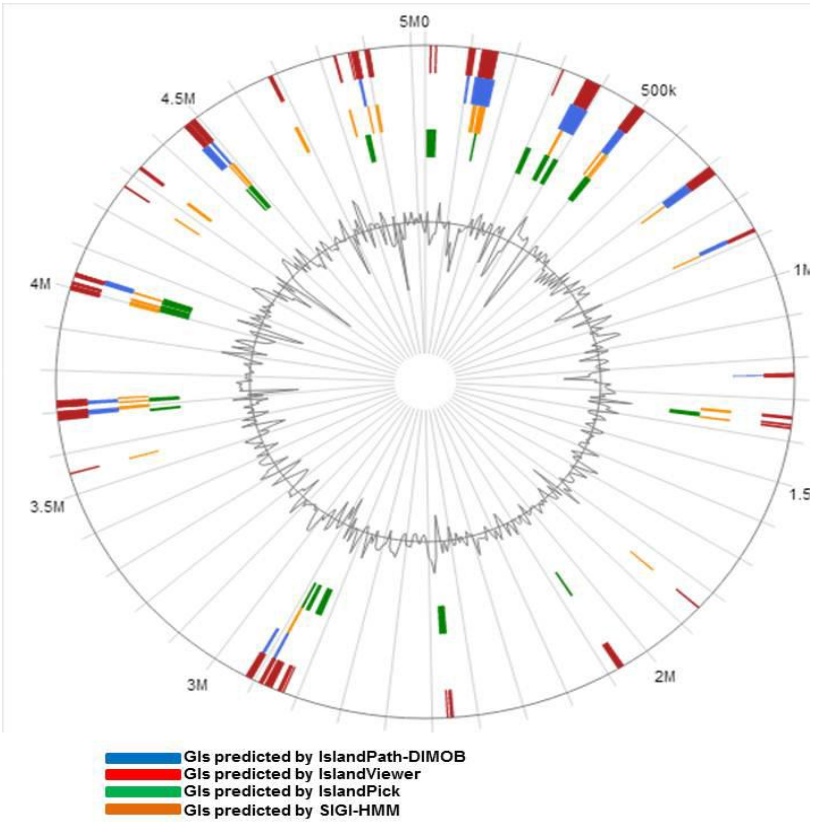


Figure 4.3.B: Genomic Island of *Enterobacter* sp. RC4 prediction by different methods

4.8: Two Component Regulatory system of *P. aeruginosa* N002 and *Enterobacter* sp. RC4

Bacterial communities in a wide range of environmental niches sense and respond to numerous external stimuli for their survival. Primarily, a source they require to follow up this communication is the two-component signal transduction system (TCS), which typically comprises a sensor Histidine kinase for receiving external input signals and a response regulator that conveys a proper change in the bacterial cell physiology. Several regulatory system through which a cell sense and then respond to environmental signals that is two component system is described in (Table 4.6). Altogether 213 and 164 TCS were found in *P. aeruginosa* N002 and *Enterobacter* sp. RC4 respectively. The presence of convincingly high numbers of TCS in both the strain in *P. aeruginosa* N002 and *Enterobacter* sp. RC4confirmed the responsibility of adaptation to oil degradation and adaptability.

Table 4.6: Two Component system (TCS) genes of *P. aeruginosa* N002 and *Enterobacter* sp. RC4 genome

***P. aeruginosa* N002 TCS genes**		***Enterobacter* sp.RC4 TCS Genes**	
Locus Tag	**Gene Product Name**	**Locus Tag**	**Gene Product Name**
A222_04932	3-ketoacyl-CoA thiolase	C4L14_04865	[citrate (pro-3S)-lyase] ligase
A222_03022	acetyl-CoA acetyltransferase	C4L14_03010	acetyl-CoA acetyltransferase
A222_01514	acetyl-CoA acyltransferase	C4L14_19370	acetyl-CoA C-acetyltransferase
A222_01054, A222_02469, A222_03303, A222_01395	acetyl-CoA C-acetyltransferase	C4L14_12385	amino acid ABC transporter substrate-binding protein (PAAT family) /L-glutamate-binding protein /L-aspartate-binding protein
A222_01492	acyl homoserine lactone synthase	C4L14_17370	anaerobic C4-dicarboxylate transporter DcuB

A222_00126, A222_03856	aerobic C4-dicarboxylate transport protein	C4L14_15010	carbon storage regulator CsrA
A222_05657	Alginate biosynthesis sensor protein KinB	C4L14_15215	CCS family citrate carrier protein
A222_01684	alkaline phosphatase	C4L14_17305, C4L14_21875	CheA signal transduction histidine kinase
A222_01069	alkaline phosphatase D	C4L14_17255, C4L14_21655	chemotaxis protein methyltransferase CheR
A222_02435	amino acid-binding domain sensor hybrid histidine kinase	C4L14_17315, C4L14_21865	chemotaxis protein MotA
A222_00868	beta-lactamase class C	C4L14_01090	chromosomal replication initiator protein DnaA
A222_05330	C4-dicarboxylate transporter, DctM subunit	C4L14_04855	citrate lyase subunit beta/citryl-CoA lyase
A222_04163, A222_05329	C4-dicarboxylate transporter, DctQ subunit	C4L14_04860	citrate lyase subunit gamma (acyl carrier protein)
	C4-dicarboxylate transporter, DctQ subunit	C4L14_11080	CRP/FNR family cyclic AMP-dependent transcriptional regulator
A222_05328, A222_04164, A222_04737	C4-dicarboxylate-binding protein DctP	C4L14_01910, C4L14_09855	Cu(I)/Ag(I) efflux system membrane fusion protein
A222_04141	carbon storage regulator, CsrA	C4L14_09850, C4L14_01915	Cu(I)/Ag(I) efflux system membrane protein CusA/SilA
A222_00433	chemosensory pili system protein ChpA (sensor histidine kinase/response regulator)	C4L14_01900, C4L14_09865	Cu(I)/Ag(I) efflux system outer membrane protein
A222_00434	chemosensory pili system protein ChpB (putative protein-glutamate methylesterase)	C4L14_01905, C4L14_09860	Cu(I)/Ag(I) efflux system protein CusF
A222_00435	chemosensory pili system protein ChpC	C4L14_23760, C4L14_03590, C4L14_12690	cytochrome bd-I ubiquinol oxidase subunit 1 apoprotein
A222_00183, A222_01627	chemotaxis protein methyltransferase CheR	C4L14_12695, C4L14_23755	cytochrome bd-I ubiquinol oxidase subunit 2 apoprotein
A222_01276	chemotaxis protein methyltransferase WspC	C4L14_03585	cytochrome d ubiquinol oxidase subunit II

A222_03597, A222_05107	chemotaxis protein MotA	C4L14_20105	DHA2 family multidrug resistance protein-like MFS transporter
A222_01275	chemotaxis-related protein WspB	C4L14_03135	DSF synthase
A222_01277	chemotaxis-related protein WspD	C4L14_12040	ferric enterobactin receptor
A222_00001	chromosomal replication initiator protein DnaA	C4L14_21860, C4L14_17320	flagellar transcriptional activator FlhC
A222_04663	conjugative transfer region protein, TIGR03748 family	C4L14_17325, C4L14_21855	flagellar transcriptional activator FlhD
A222_00671	CRP/FNR family transcriptional regulator, cyclic AMP receptor protein	C4L14_21755, C4L14_17520	Flagellin
A222_01049	cytochrome bd-I ubiquinol oxidase subunit 1 apoprotein	C4L14_15120	formate dehydrogenase (quinone-dependent) catalytic subunit
A222_01050	cytochrome bd-I ubiquinol oxidase subunit 2 apoprotein	C4L14_15115	formate dehydrogenase (quinone-dependent) iron-sulfur subunit
A222_00536	cytochrome c	C4L14_15110	formate dehydrogenase-N gamma subunit
A222_00119	cytochrome c oxidase assembly protein subunit 15	C4L14_21920	fumarate reductase flavoprotein subunit
A222_03498, A222_00845, A222_03173, A222_03494	cytochrome c oxidase cbb3-type subunit 1	C4L14_11380	fumarate reductase subunit C
A222_03499, A222_03495	cytochrome c oxidase cbb3-type subunit 2	C4L14_04845	holo-ACP synthase
A222_03497, A222_03501	cytochrome c oxidase cbb3-type subunit 3	C4L14_02685	Hpt sensor hybrid histidine kinase
A222_03500, A222_03496	cytochrome c oxidase cbb3-type subunit 4	C4L14_12575	K+-transporting ATPase ATPase A chain
A222_01199	D-alanyl-D-alanine dipeptidase	C4L14_12570	K+-transporting ATPase ATPase B chain
A222_02326, A222_05137, A222_01786, A222_04112, A222_04922	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain	C4L14_12565	K+-transporting ATPase ATPase C chain
A222_03946	Flagellin	C4L14_12580	K+-transporting ATPase ATPase F chain

A222_01446	GDP-mannose 6-dehydrogenase	C4L14_12370	L-glutamate ABC transporter ATP-binding protein /L-aspartate ABC transporter ATP-binding protein
A222_05696, A222_00308, A222_02982, A222_03486	glutamate--putrescine ligase	C4L14_12375, C4L14_12380	L-glutamate ABC transporter membrane protein /L-aspartate ABC transporter membrane protein
A222_03701	glutamin-(asparagin-)ase	C4L14_10390	L-glutamine synthetase
A222_04286	GTP-binding protein LepA	C4L14_22005	LuxR family capsular biosynthesis transcriptional activator
A222_00034	HPt (histidine-containing phosphotransfer) domain-containing protein	C4L14_17475	LuxR family transcriptional regulator
A222_04080	imipenem/basic amino acid-specific outer membrane pore	C4L14_17465, C4L14_01290, C4L14_20675, C4L14_08650, C4L14_16400, C4L14_04345	LuxR family two component transcriptional regulator
A222_03416	K+-transporting ATPase ATPase A chain	C4L14_15220	malate dehydrogenase (oxaloacetate-decarboxylating)
A222_03415	K+-transporting ATPase ATPase B chain	C4L14_21745	methyl-accepting chemotaxis protein
A222_03414	K+-transporting ATPase ATPase C chain	C4L14_04405	methyl-accepting chemotaxis protein-2 (aspartate sensor receptor)
A222_03699	L-glutamate ABC transporter ATP-binding protein /L-aspartate ABC transporter ATP-binding protein	C4L14_16450	methyl-accepting chemotaxis sensory transducer with Cache sensor
A222_03698, A222_03697	L-glutamate ABC transporter membrane protein /L-aspartate ABC transporter membrane protein	C4L14_04410	methyl-accepting chemotaxis sensory transducer with Pas/Pac sensor
A222_03696, A222_05240	L-glutamate-binding protein /L-aspartate-binding protein	C4L14_17260, C4L14_10055, C4L14_04975, C4L14_18700, C4L14_24060, C4L14_17265,	methyl-accepting chemotaxis sensory transducer with TarH sensor

		C4L14_23705, C4L14_06800	
A222_00311, A222_05682, A222_05276, A222_01619	L-glutamine synthetase	C4L14_10625, C4L14_20100, C4L14_20095	multidrug efflux pump
A222_03127	LuxR family transcriptional regulator	C4L14_20090	multidrug efflux system membrane fusion protein
A222_01491	LuxR family transcriptional regulator, transcriptional activator of rhlAB and lasB	C4L14_01725	Na+/H+-dicarboxylate symporter
A222_03004, A222_02495	membrane fusion protein, multidrug efflux system	C4L14_15075	NAD-dependent malic enzyme
A222_00197	Membrane-associated phospholipid phosphatase	C4L14_06205	nitrogen regulatory protein P-II family
A222_00188, A222_02123, A222_03402, A222_04396, A222_02459, A222_03790, A222_05066, A222_04758, A222_02068, A222_02446, A222_00184, A222_04638, A222_0523, A222_04414, A222_03441	methyl-accepting chemotaxis protein	C4L14_01305	OPA family hexose phosphate transport protein UhpT-like MFS transporter
A222_01274	methyl-accepting chemotaxis protein WspA	C4L14_01300	OPA family sugar phosphate sensor protein UhpC-like MFS transporter
A222_04416, A222_02362, A222_04413, A222_02360, A222_04417	methyl-accepting chemotaxis sensory transducer with Cache sensor	C4L14_20665	outer membrane pore protein C
A222_03491, A222_03090, A222_03633	methyl-accepting chemotaxis sensory transducer with Pas/Pac sensor	C4L14_05340	outer membrane pore protein F

A222_01033	multi-sensor hybrid histidine kinase	C4L14_03905	outer membrane protein
A222_02496, A222_03005, A222_02497,	multidrug efflux pump	C4L14_22705	PAS/PAC sensor hybrid histidine kinase
A222_01497	NAD-dependent malic enzyme	C4L14_14750, C4L14_10400	PAS/PAC sensor signal transduction histidine kinase
A222_05282	nitrogen metabolism transcriptional regulator, NtrC, Fis family	C4L14_07825	peptidase Do
A222_05361	Osmolarity sensor protein envZ	C4L14_01005	phosphate ABC transporter substrate-binding protein (PhoT family)
A222_05048, A222_04081, A222_02520	outer membrane porin, OprD family	C4L14_21880, C4L14_17300	purine-binding chemotaxis protein CheW
A222_05127, A222_00834	outer membrane protein	C4L14_20770, C4L14_20775, C4L14_20780	putative tricarboxylic transport membrane protein
A222_04110, A222_02324	outer membrane receptor for ferrienterochelin and colicins	C4L14_07990	RcsF protein
A222_05526, A222_05281, A222_03940	PAS/PAC sensor signal transduction histidine kinase	C4L14_15155, C4L14_16385	respiratory nitrate reductase alpha subunit apoprotein
A222_02475, A222_05534	phosphate ABC transporter substrate-binding protein, PhoT family	C4L14_15160, C4L14_16380	respiratory nitrate reductase beta subunit
A222_04122	phosphatidylglycerol lysyltransferase	C4L14_16375, C4L14_15165	respiratory nitrate reductase chaperone NarJ
A222_03594, A222_00185, A222_03593	purine-binding chemotaxis protein CheW	C4L14_15170, C4L14_16370	respiratory nitrate reductase gamma subunit
A222_04301, A222_04300, A222_04302	putative tricarboxylic transport membrane protein	C4L14_22740	RNA polymerase RpoN-/SigL-like sigma 54 subunit
A222_01108	respiratory nitrate reductase alpha subunit apoprotein	C4L14_21850	RNA polymerase sigma factor for flagellar operon FliA
A222_01109	respiratory nitrate reductase beta subunit	C4L14_17505	RNA polymerase sigma-28 (SigD/FliA/WhiG) subunit

A222_01110	respiratory nitrate reductase chaperone NarJ	C4L14_11390	succinate dehydrogenase subunit A
A222_01111	respiratory nitrate reductase gamma subunit	C4L14_11385	succinate dehydrogenase subunit B
A222_01280	response regulator receiver modulated diguanylate cyclase	C4L14_11375	succinate dehydrogenase subunit D
A222_03602	RNA polymerase, sigma 28 subunit, SigD/FliA/WhiG	C4L14_20765	transcriptional regulator
A222_04580	RNA polymerase, sigma 54 subunit, RpoN/SigL	C4L14_04840	triphosphoribosyl-dephospho-CoA synthase
A222_04673	Sensor protein PilS	C4L14_07060	two-component system aerobic respiration control protein ArcA
A222_05429	Sensory transduction protein kinase AlgZ	C4L14_01455	two-component system capsular synthesis response regulator RcsB
A222_04287	serine protease Do	C4L14_20680	two-component system capsular synthesis sensor histidine kinase RcsC
A222_03021	short-chain fatty acids transporter	C4L14_21650, C4L14_17250	two-component system chemotaxis response regulator CheB
A222_03941	sigma-54 specific transcriptional regulator, flagellar regulatory protein A	C4L14_20830	two-component system chemotaxis response regulator CheV
A222_02325, A222_04111	Signal transduction histidine kinase	C4L14_21645, C4L14_17245	two-component system chemotaxis response regulator CheY
A222_04161, A222_04162	TRAP transporter, DctM subunit	C4L14_04880	two-component system cit operon sensor histidine kinase CitA
A222_00430, A222_00431	twitching motility protein PilI	C4L14_09870, C4L14_01895, C4L14_01965	two-component system copper resistance phosphate regulon response regulator CusR
A222_00428	twitching motility two-component system response regulator PilG	C4L14_01890, C4L14_01970, C4L14_09875	two-component system heavy metal sensor histidine kinase CusS
A222_00429	twitching motility two-component system response regulator PilH	C4L14_12555	two-component system KDP operon response regulator KdpE

A222_01268, A222_00035, A222_01933, A222_02091, A222_01031, A222_00894, A222_01104	two component transcriptional regulator, LuxR family	C4L14_10630	two-component system nitrate/nitrite sensor histidine kinase NarQ
A222_05428	two component transcriptional regulator, LytTR family /Two-component response regulator AlgR	C4L14_16395	two-component system nitrate/nitrite sensor histidine kinase NarX
A222_05525	two component transcriptional regulator, winged helix family	C4L14_10405	two-component system nitrogen regulation response regulator GlnG
A222_05656	Two-component response regulator AlgB	C4L14_10900	two-component system osmolarity sensor histidine kinase EnvZ
A222_05362	Two-component response regulator OmpR	C4L14_10895	two-component system phosphate regulon response regulator OmpR
A222_04674	Two-component response regulator PilR	C4L14_14745	two-component system phosphate regulon response regulator PhoB
A222_03598, A222_00181	two-component system, chemotaxis family, response regulator CheB	C4L14_20115, C4L14_22845	two-component system response regulator BaeR
A222_01626	two-component system, chemotaxis family, response regulator CheV	C4L14_04885	two-component system response regulator CitB
A222_03601, A222_00187	two-component system, chemotaxis family, response regulator CheY	C4L14_10155	two-component system response regulator CpxR
A222_01279	two-component system, chemotaxis family, response regulator WspF	C4L14_21930	two-component system response regulator DcuR
A222_01278	two-component system, chemotaxis family, sensor histidine kinase and response regulator WspE	C4L14_01605	two-component system response regulator FimZ (fimbrial Z protein)
A222_03599, A222_00186	two-component system, chemotaxis family, sensor kinase CheA	C4L14_06200	two-component system response regulator GlrR
A222_01105	two-component system, NarL family, nitrate/nitrite sensor histidine kinase	C4L14_05870, C4L14_23320	two-component system response regulator PhoP

	NarX		
A222_04113	two-component system, NarL family, sensor histidine kinase BarA	C4L14_03805, C4L14_23365	two-component system response regulator QseB
A222_05327, A222_05685	two-component system, NtrC family, C4-dicarboxylate transport response regulator DctD	C4L14_18795	two-component system response regulator RstA
A222_05686, A222_05326	two-component system, NtrC family, C4-dicarboxylate transport sensor histidine kinase DctB	C4L14_20110	two-component system sensor histidine kinase BaeS
A222_00483	two-component system, OmpR family, catabolic regulation response regulator CreB	C4L14_22840	two-component system sensor histidine kinase BasS
A222_03620, A222_02193, A222_05035, A222_02501	two-component system, OmpR family, copper resistance phosphate regulon response regulator CusR	C4L14_10160	two-component system sensor histidine kinase CpxA
A222_02192, A222_05036, A222_02500, A222_03619	two-component system, OmpR family, heavy metal sensor histidine kinase CusS	C4L14_21935	two-component system sensor histidine kinase DcuS
A222_03412	two-component system, OmpR family, KDP operon response regulator KdpE	C4L14_04340	two-component system sensor histidine kinase EvgS
A222_03233	two-component system, OmpR family, response regulator ParR	C4L14_06190	two-component system sensor histidine kinase GlrK
A222_03860	two-component system, OmpR family, response regulator PhoP	C4L14_12560	two-component system sensor histidine kinase KdpD
A222_03882	two-component system, OmpR family, response regulator RstA	C4L14_05865	two-component system sensor histidine kinase PhoQ
A222_04298	two-component system, OmpR family, response regulator TctD	C4L14_03810	two-component system sensor histidine kinase QseC
A222_00484	two-component system, OmpR family, sensor histidine kinase CreC	C4L14_20670	two-component system sensor histidine kinase RcsD

A222_03413	two-component system, OmpR family, sensor histidine kinase KdpD	C4L14_18805	two-component system sensor histidine kinase RstB
A222_03859	two-component system, OmpR family, sensor histidine kinase PhoQ	C4L14_20760	two-component system sensor histidine kinase TctE
A222_04923	two-component system, OmpR family, sensor histidine kinase QseC	C4L14_01295	two-component system sensor histidine kinase UhpB
A222_03881	two-component system, OmpR family, sensor histidine kinase RstB	C4L14_02005	UDP-4-amino-4-deoxy-L-arabinose-oxoglutarate aminotransferase
A222_04297	two-component system, OmpR family, sensor histidine kinase TctE	C4L14_07845	UTP--GlnB (protein PII) uridylyltransferase GlnD
A222_03234	two-component system, OmpR family, sensor kinase ParS		
A222_03703	two-component system, response regulator AauR		
A222_03939	two-component system, response regulator FlrC		
A222_04611	two-component system, response regulator RegA		
A222_03702	two-component system, sensor histidine kinase AauS		
A222_04612	two-component system, sensor histidine kinase RegB		
A222_00432	type IV pilus assembly protein PilK		
A222_04545	ubiquinol-cytochrome c reductase cytochrome b subunit		
A222_04544	ubiquinol-cytochrome c reductase cytochrome c1 subunit		
A222_04546	ubiquinol-cytochrome c reductase iron-sulfur subunit		
A222_01434	UDP-4-amino-4-deoxy-L-arabinose-oxoglutarate aminotransferase		

A222_01324	UTP--GlnB (protein PII) uridylyltransferase, GlnD	
A222_04305	vancomycin resistance protein VanW	

Details of two component system signaling pathway of CitB, Lux, LuxR, Cell cycle, LytT, NarL, NtrC, OmpR, Sporilation, and Other family genes of *P. aeruginosa* N002 and *Enterobacter* sp. RC4 are described below.

4.8.a: Citrate utilization (CitB Family)

In *P. aeruginosa* N002 and *Enterobacter* sp. RC4 multiple genes are involved in citrate utilization process (Figure 4.4.A and 4.4.B). There are five systemes that regulate citrate utilization processes involving *CitA, CitS, DcuS, DctS* and *MalK* for environmental signals while *CitB, CitT, DcuR, DctR* and *MalR* for response regulator. The activity response relates to citrate utilization were citrate fomentation as well as anaerobic metabolism, citrate transport anaerobic respiration and C4-dicarboxyrate transport and malate utilization.

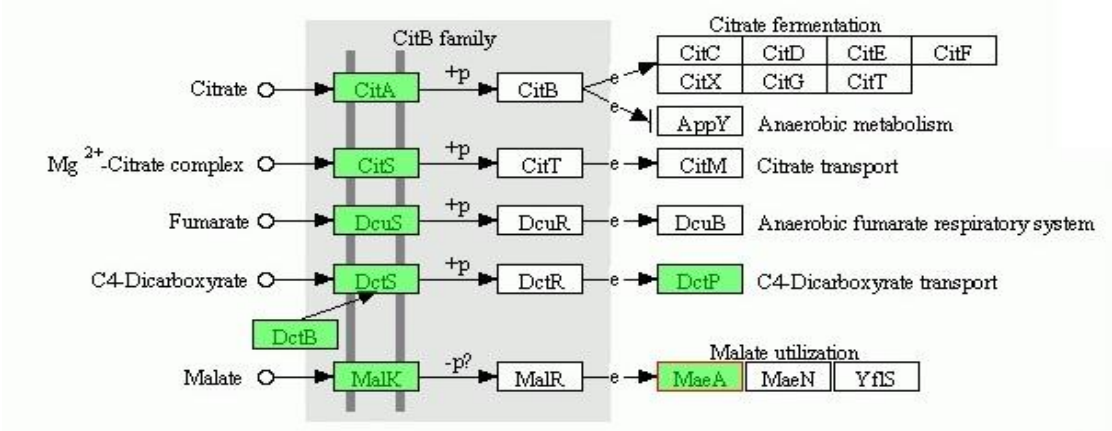


Figure 4.4.A: *P. aeruginosa* N002 CitB family two component system pathways

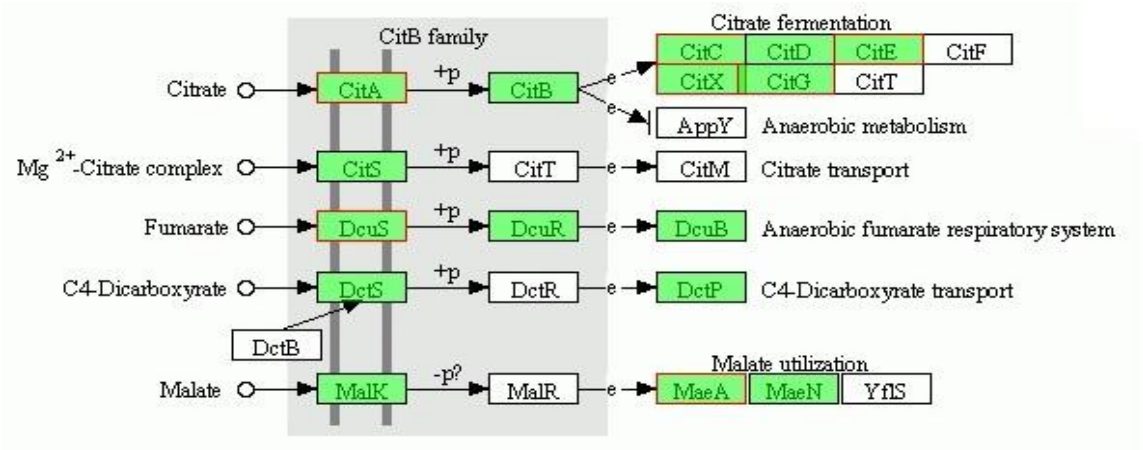


Figure 4.4.B: *Enterobacter* sp. RC4 CitB family two component system pathways

4.8.b: Mechanism of chemotaxis response

The mechanism of chemotaxis regulation in *P. aeruginosa* N002 and *Enterobacter* sp. RC4 there were systems which are senses of five sensory genes *MCP, PilJ, PixJ, WspA* and *Cph1* (Figure 4.4.C and 4.4.D). Finally the chemotaxis genes family regulates the activity response of flagella switch, motility, pototaltic and extracellular polysaccharide production.

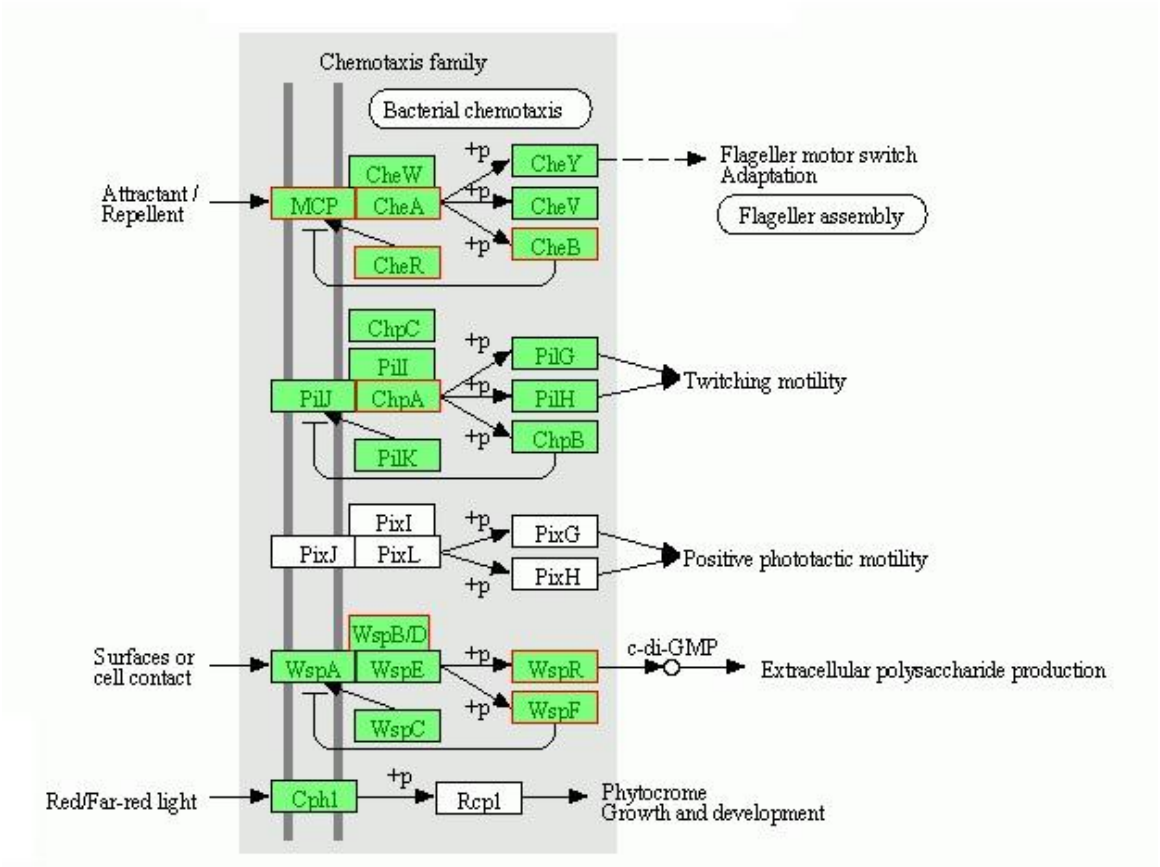


Figure 4.4.C: *P. aeruginosa* N002 Chemotaxis family two component system pathways

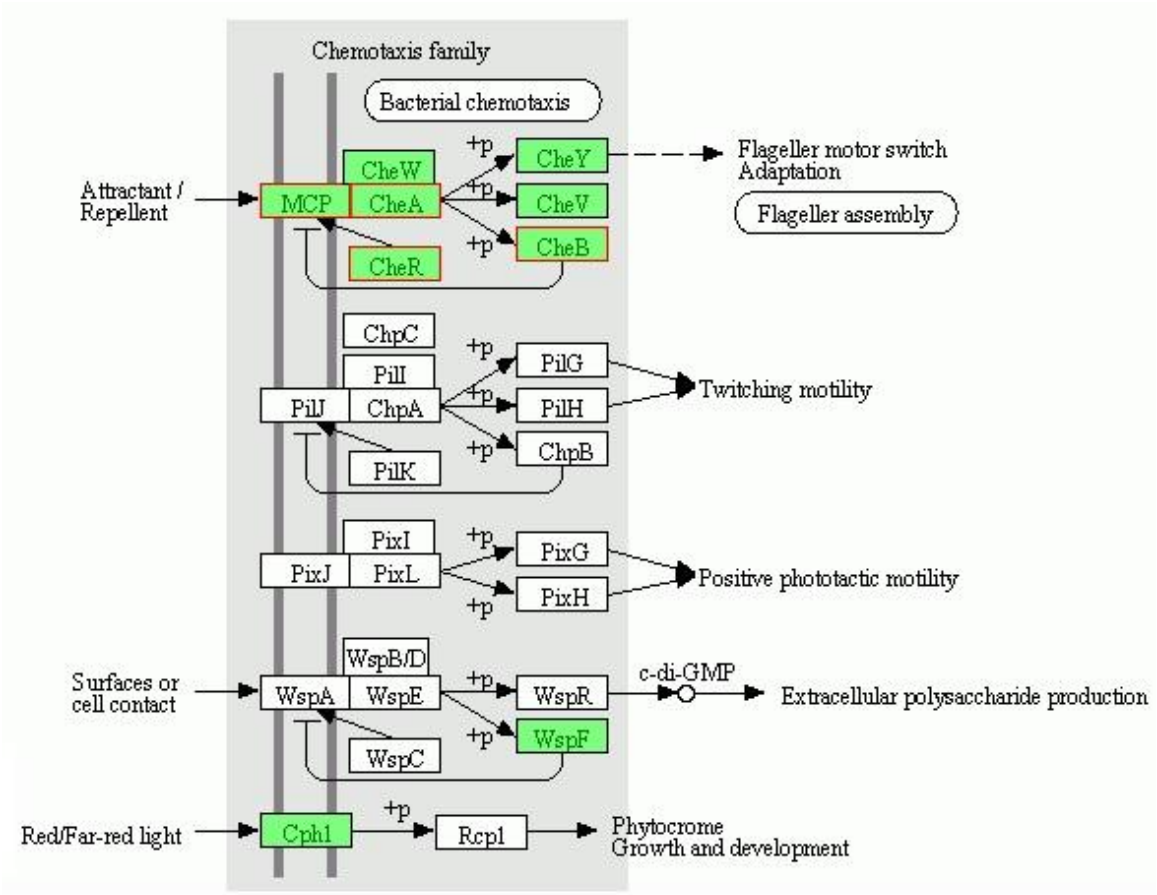


Figure 4.4.D: *Enterobacter* sp. RC4 Chemotaxis family two component system pathways

4.8.c: Lux, LuxR and Cell cycle family

The genes involved in activity response of cell cycle, *Lux* and *LuxR* family are described (Figure 4.4.E and 4.4.F). These are single feeding specific signals and three induce systems and carboxyurate, thioncte and O_2 limitation systems. To arrive in activity responses of cell cycle genes, *Lux* and *LuxR* altogether nine response genes were involved in the entire system of *P. eaeruginosa* N002 and *Enterobacter* sp. RC4 respectively.

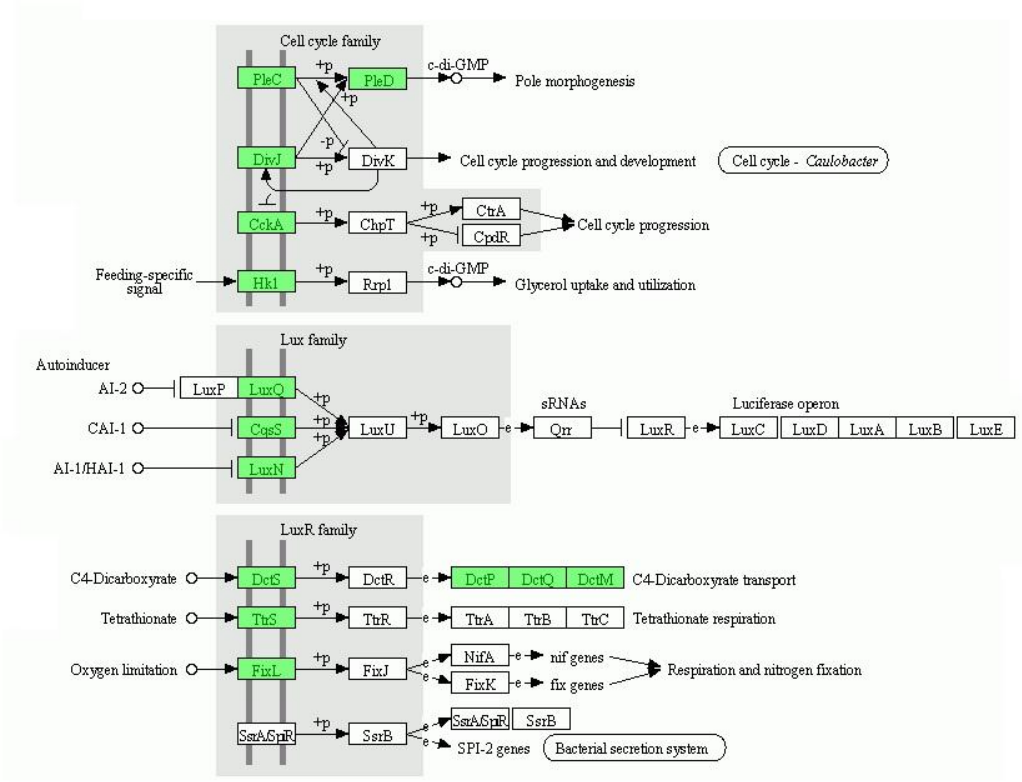


Figure 4.4.E: *P. aeruginosa* N002 Lux, LuxR, and, Cell cycle family two component system pathways

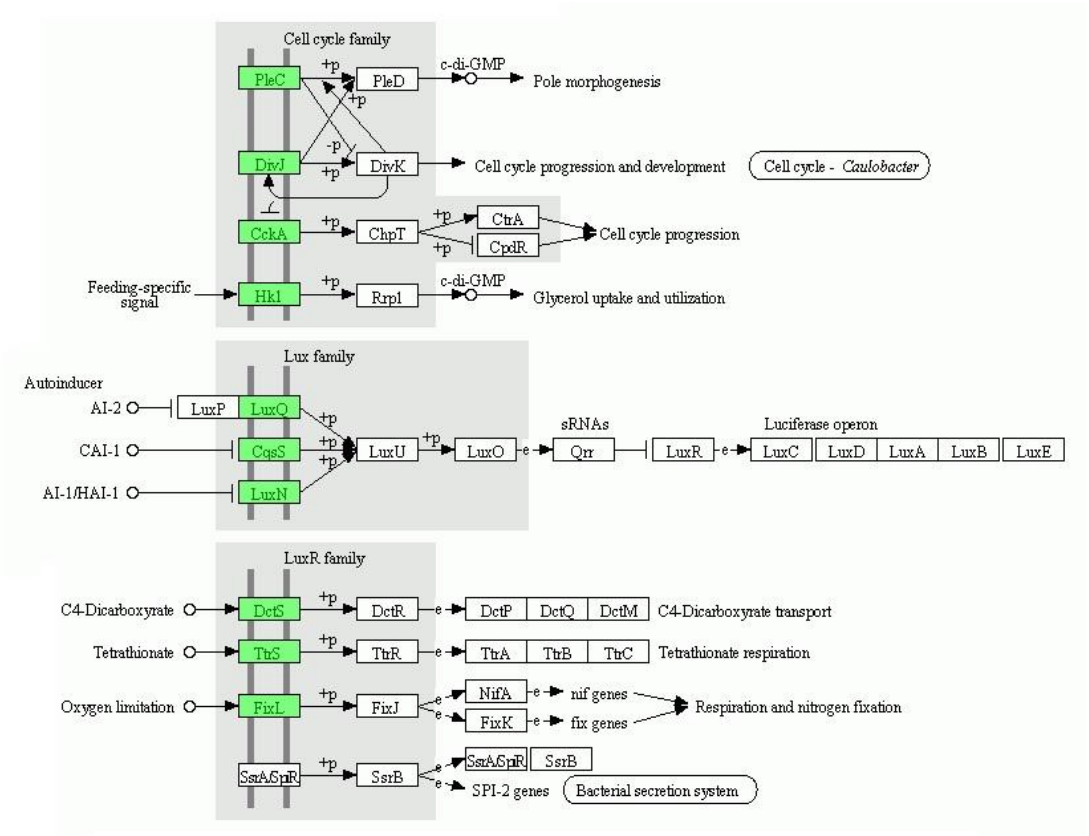


Figure 4.4.F: *Enterobacter* sp. RC4 Lux, LuxR, and, Cell cycle family two component system pathways

4.8.d: LytT family:

In *P. aeruginosa* N002 and *Enterobacter* sp. RC4 having *LytS, AlgZ, NatK* genes to control of pyruvate utilization which are showen in the figure 4.4.G and 4.4.H. The LytT family of two-component response regulator is mostly involved to control of pyruvate utilization.

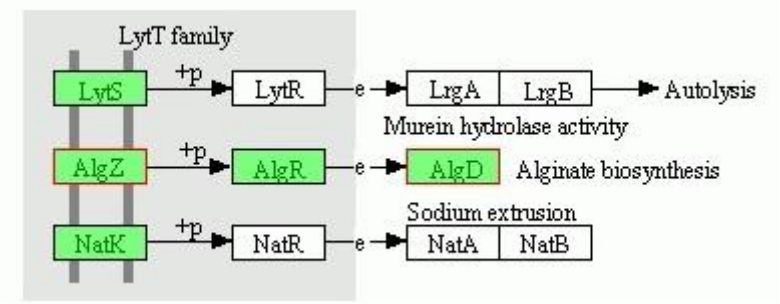


Figure 4.4.G: *P. aeruginosa* N002 LytT family two component system pathways

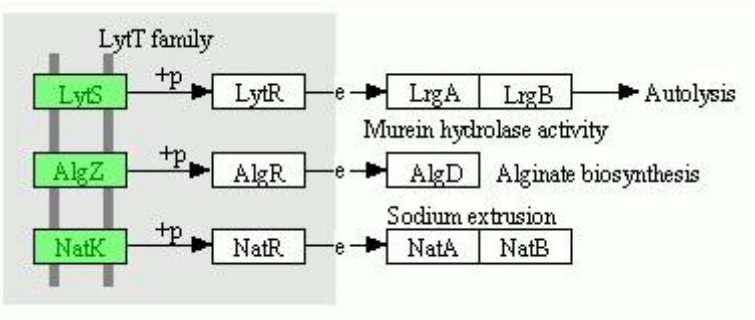


Figure 4.4.H: *Enterobacter* sp. RC4 LytT family two component system pathways

4.8.e: NtrLFamily

The NtrL family two component systems genes *NarX*, *NarQ* for nitrate/nitrite condition and *DegX* for salt stress condition assimilation of *P. aeruginosa* N002 and *Enterobacter* sp. RC4 are describe in figure 4.4.I and 4.4.J. In NarL regulation there are eleven different systems sense through fourteen individual genes which were regulated by fourteen response regulator genes to activate seventeen different responses.

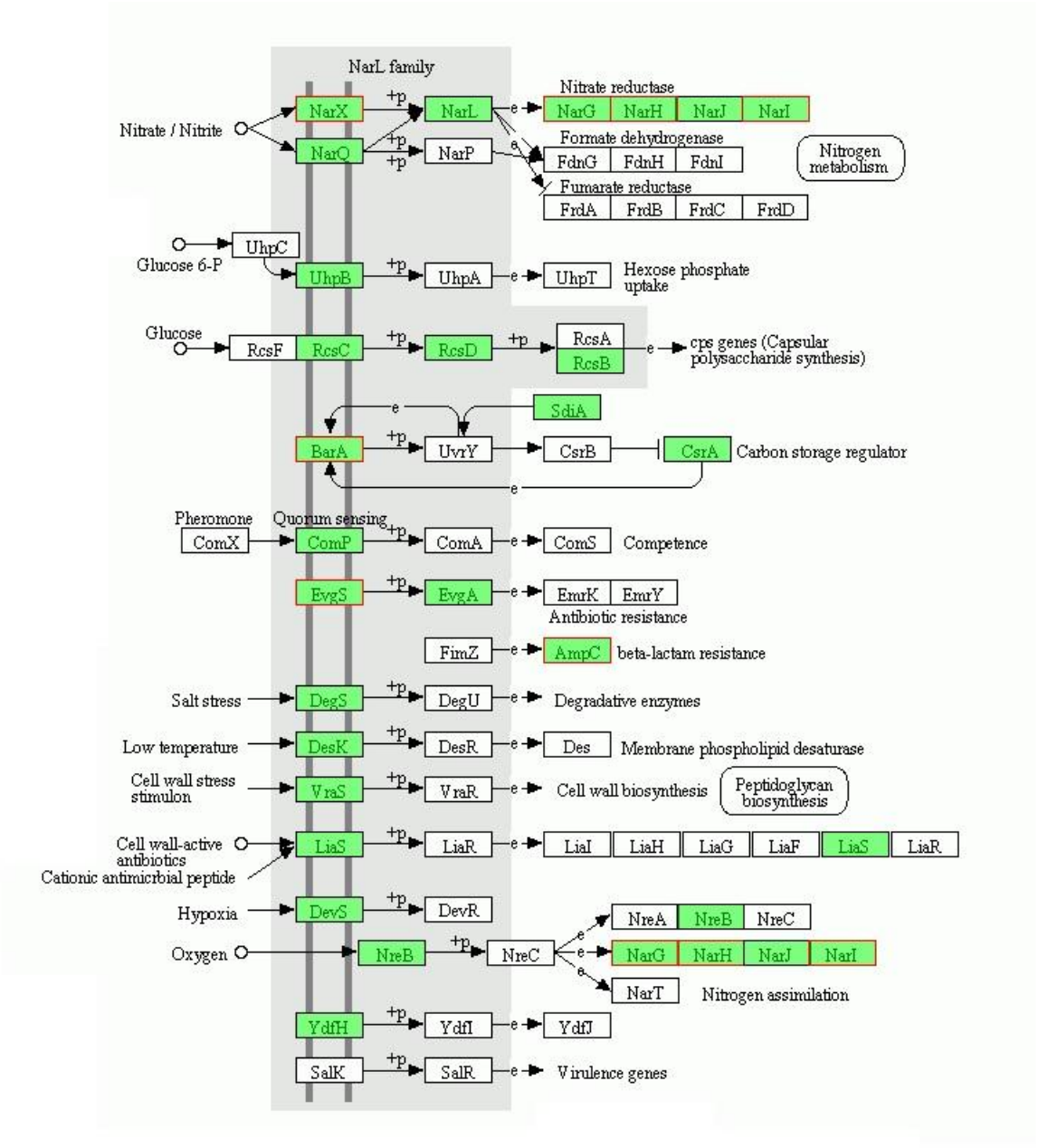


Figure 4.4.I: *P. aeruginosa* N002 NarL family two component system pathways

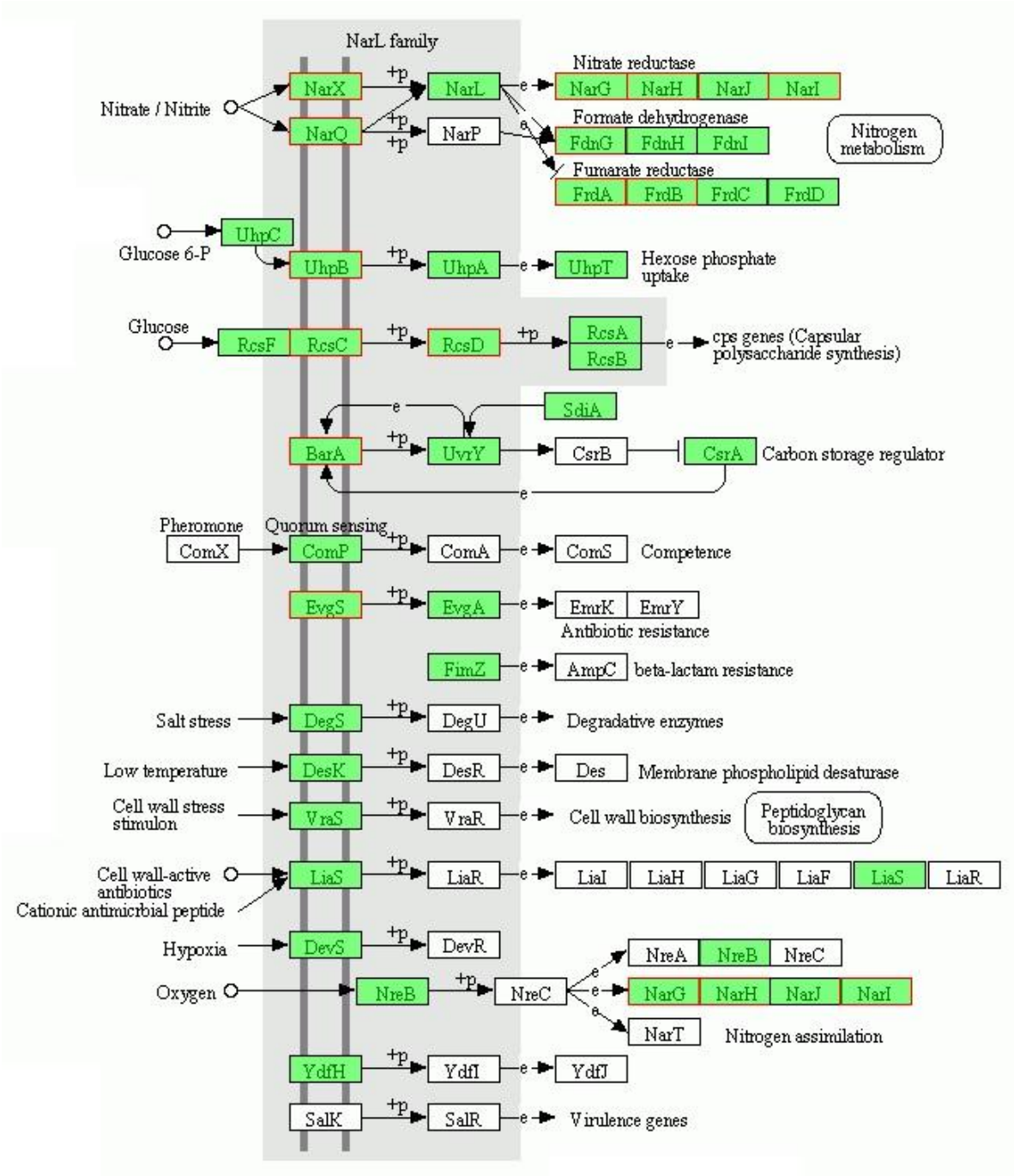


Figure 4.4.J: *Enterobacter* sp. RC4 NarL family two component system pathways

4.8.f: NtrC family

The NtrC family two component systems genes *NtrY, GlnL* for low nitrogen availability condition assimilation of *P. aeruginosa* N002 and *Enterobacter* sp. RC4 are describe in figure 4.4.K and 4.4.L. In NtrC regulation there are ten response regulator genes to activate different responses.

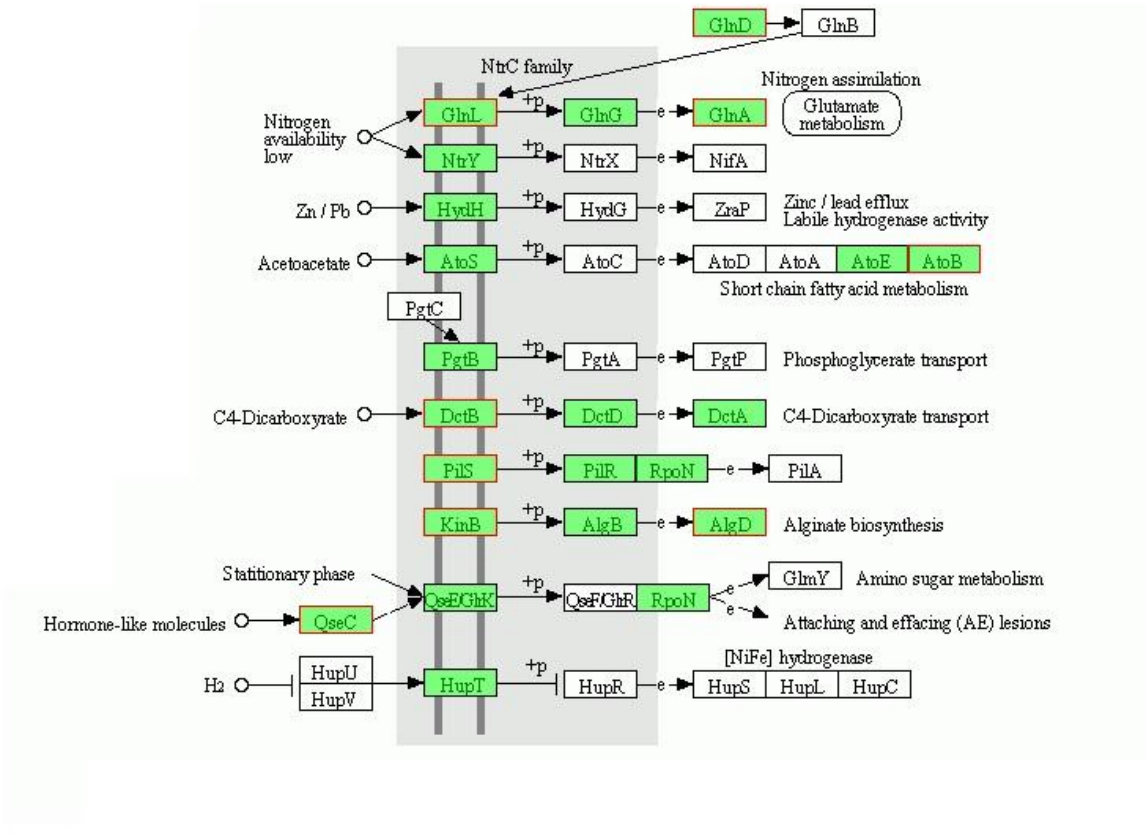


Figure 4.4.K: *P. aeruginosa* N002 NtrC family two component system pathways

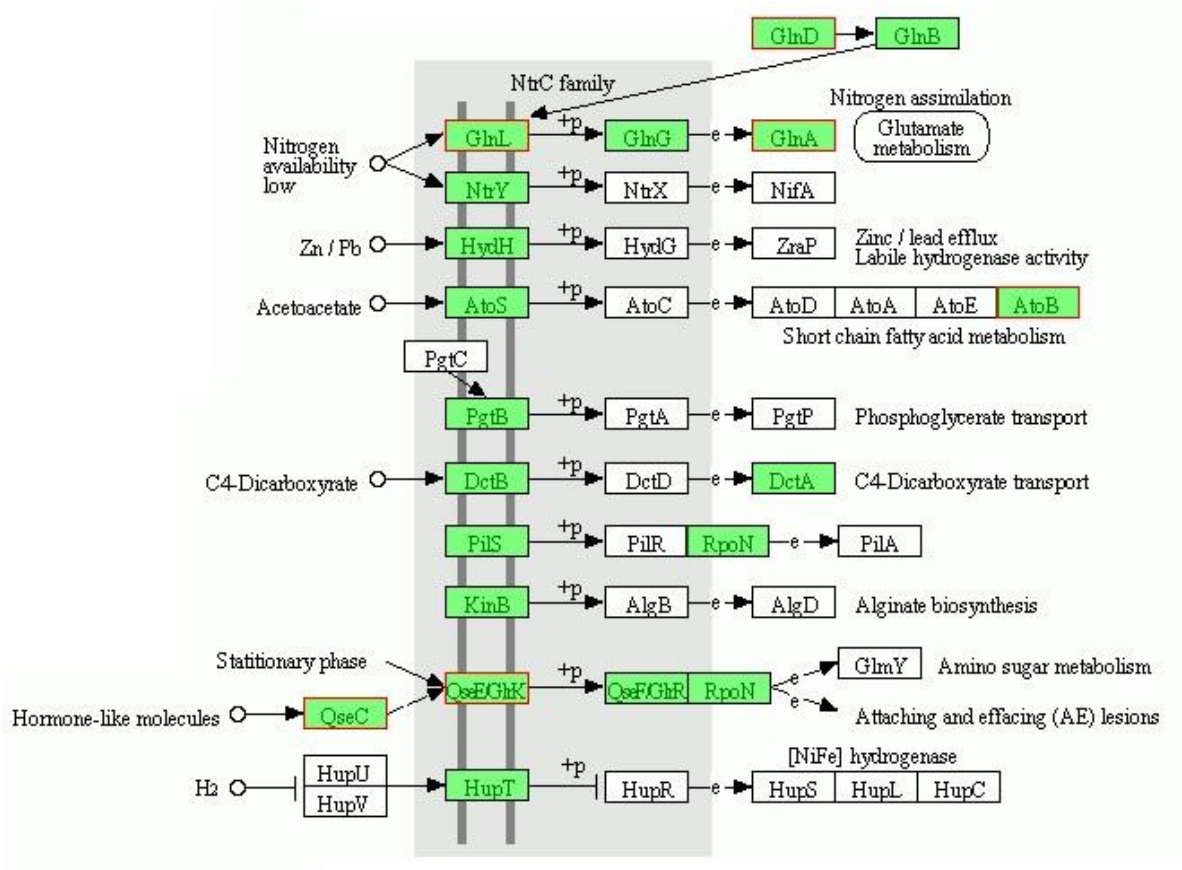


Figure 4.4.L: *Enterobacter* sp. RC4 NtrC family two component system pathways

4.8.g: Osmosis regulatory (OmpR) family

Osmoregulation of *P. aeruginosa* N002 and *Enterobacter* sp. RC4 are of complex in nature (Figure 4.4.M and 4.4.N). In osmotic regulation of *P. aeruginosa* N002 and *Enterobacter* sp. RC4 there were 31 systems that sense through 43 different genes 40 regulatory genes which actively participated in two component functioning. These genes either triggers to single functions or some more different activities.

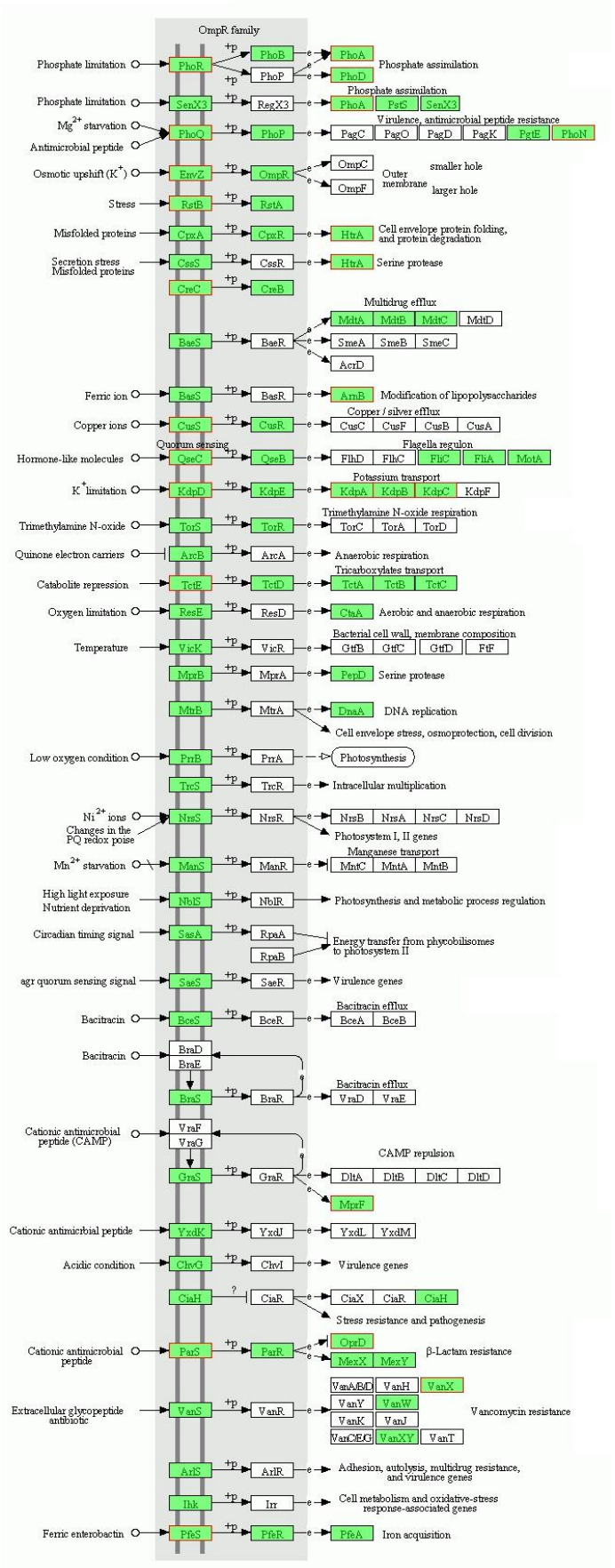


Figure 4.4.M: *P. aeruginosa* N002 OmpR family two component system pathwaysg

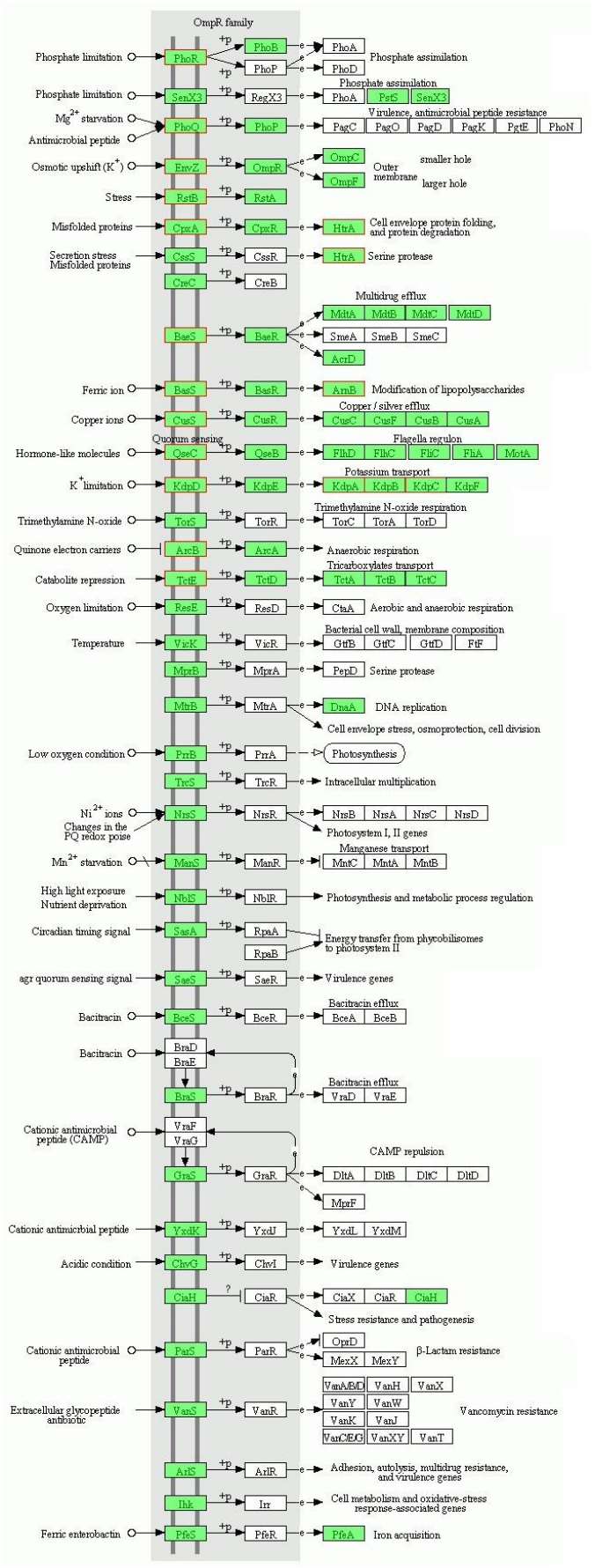


Figure 4.4.N: *Enterobacter* sp. RC4 OmpR family two component system pathways

4.8.h: Other genes family

There are some other genes which are also found in *P. aeruginosa* N002 and *Enterobacter* sp. RC4 (Figure 4.4.O and 4.4.P). The *SIn1, Chk1, Nik1, TcsA* genes are responsibule for hyperosmotic stress condition, *RegB* genes for Redox signal, *FitF* for insect signal.

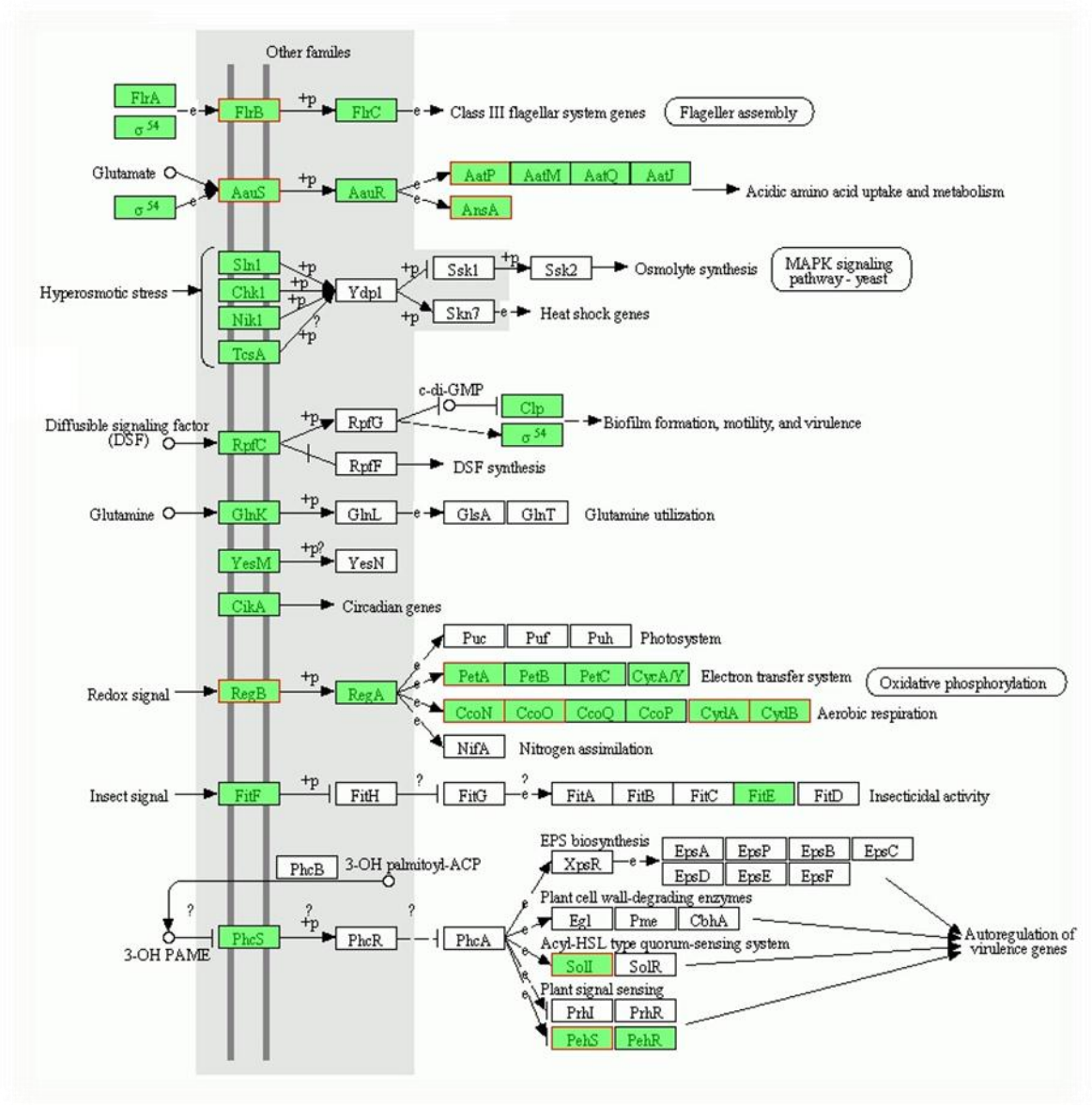


Figure 4.4.O: *P. aeruginosa* N002 other family two component system pathways

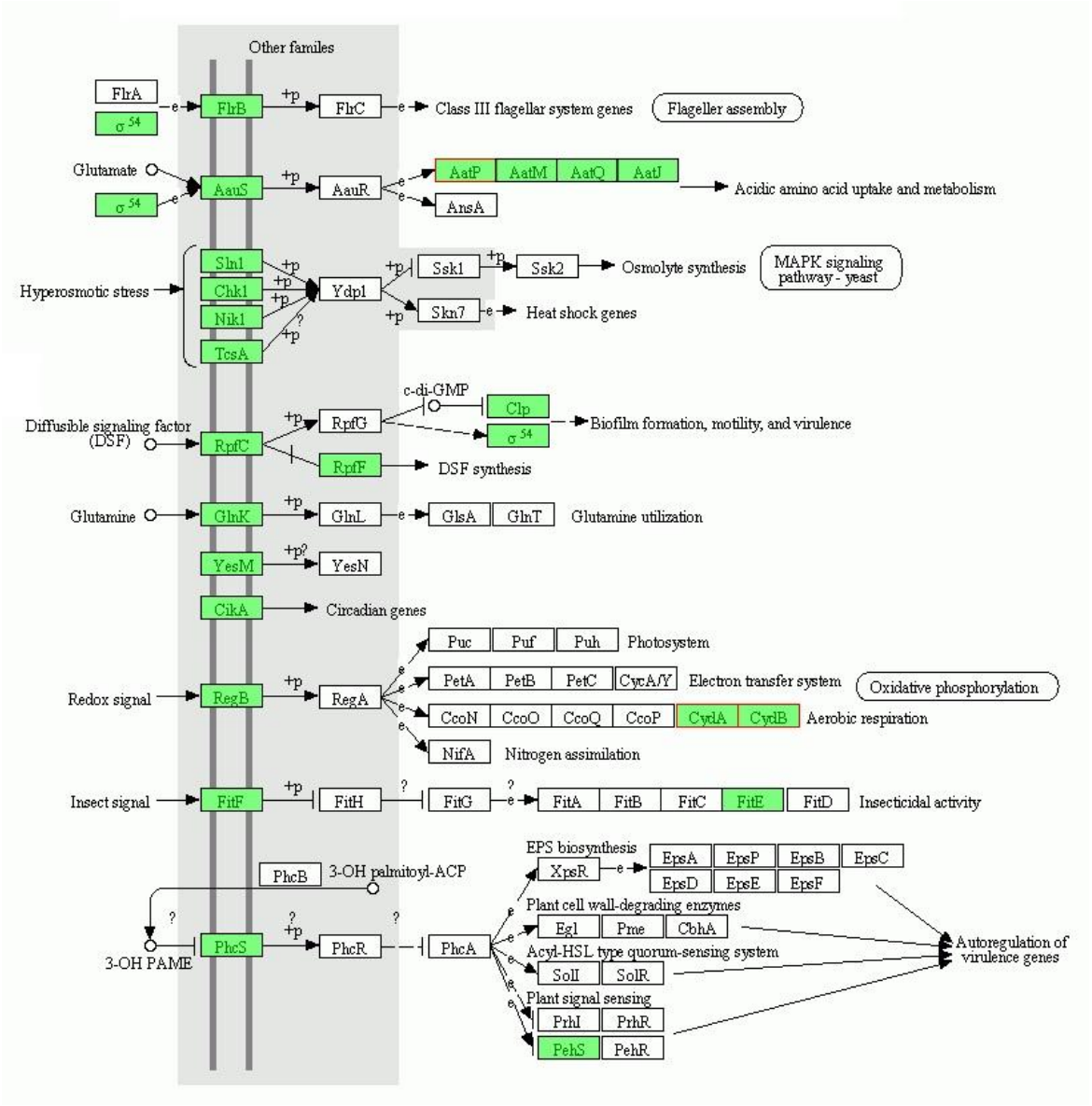


Figure 4.4.P: *Enterobacter* sp. RC4 other family two component system pathways

4.8.i: Sporulation family

Sporulation initiation, the key event is the accumulation of sufficient quantities of the response regulator *Spo0A* in its activated, phosphorylated form. Once phosphorylated, *Spo0A* represses genes expressed during post-exponential growth and activates genes required for sporulation. Phosphorylation of *Spo0A* is achieved by the multi-component phosphorelay, which is an expanded variant of the two-component signalling module. In the strain *P. aeruginosa* N002 and *Enterobacter* sp. RC4 having *KinA, KinB, KinC, KinD*

and *KinE*genes to encoding the sporulation phosphorelay sensor kinases which are showing in figure 4.4.Q and 4.4.R.

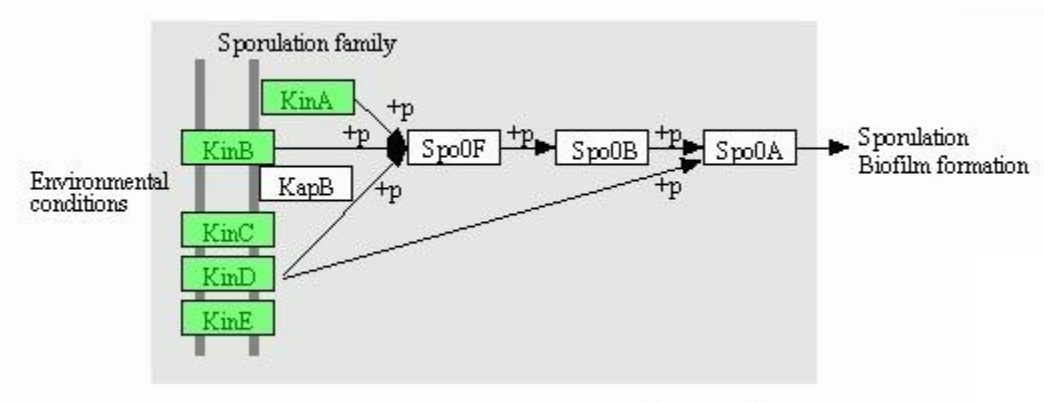


Figure 4.4.Q: *P. aeruginosa* N002 Sporilation family two component system pathways

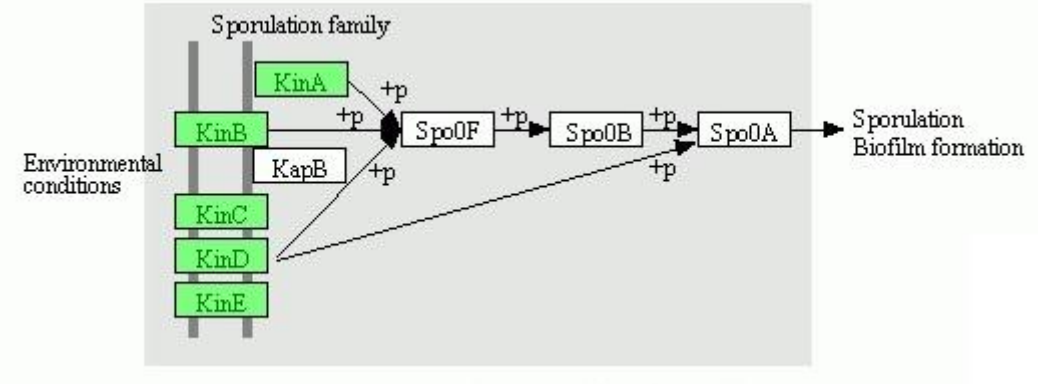


Figure 4.4.R: *Enterobacter* sp. RC4 Sporilation family two component system pathways

4.9. Regulatory genes for adaptation

Environmental adaptation category plant pathogen interaction pathway was analysis by using Kito encyclopedia of genes and genome (KEGG) database through IMG database online tool which are presented in (Table 4.7, Figure 4.4). Genome of *P. aeruginosa* N002 and *Enterobacter* sp. RC4 were found to propose ten and seven genes respectively.

Table 4.7: Adaptive genes found in *P. aeruginosa* N002 and *Enterobacter* sp. RC4

Locus tag	Gene product	COG ID	Position in Genome
***Pseudomonas aeruginosa* N002**			
A222_00835	Colicin V processing peptidase. Cysteine peptidase. MEROPS family C39	COG2274	4147bp-6306bp
A222_03946	Flagellin and related hook-associated protein FlgL	COG1344	108608bp-110348bp
A222_01402	Glycerol kinase	COG0554	118608bp-120125bp
A222_03571	Glycerol kinase	COG0554	24490bp-25905bp
A222_01405	Glycerol kinase	COG0554	121803bp-123287bp
A222_00836	Membrane fusion protein	COG1566	6316bp-7572bp
A222_03453	Molecular chaperone HtpG	COG0326	70141bp-72045bp
A222_03970	Molecular chaperone HtpG	COG0326	88769bp-90610bp
A222_00834	Outer membrane protein TolC	COG1538	2732bp-4147bp
A222_05127	Outer membrane protein TolC	COG1538	60281bp-61729bp
Enterobacter sp. RC4			
C4L14_01525	ATP-binding cassette subfamily B protein RaxB	COG2274	292295bp-294391bp
C4L14_17520	Flagellin and related hook-associated protein FlgL	COG1344	1999208bp-2000059bp
C4L14_21755	Flagellin and related hook-associated protein FlgL	COG1344	702254bp-703756bp
C4L14_15600	Glycerol kinase	COG0554	1619247bp-1620755bp
C4L14_01535	Membrane fusion protein	COG1566	295144bp-296394bp
C4L14_08915	Molecular chaperone HtpG	COG0326	309206bp-311080bp
C4L14_03905	Outer membrane protein TolC	COG1538	818296bp-819762bp

The pathway analysis of different adaptive genes suggests that both the bacteria strain follows respective adaptation processes to accustomed under different environment.

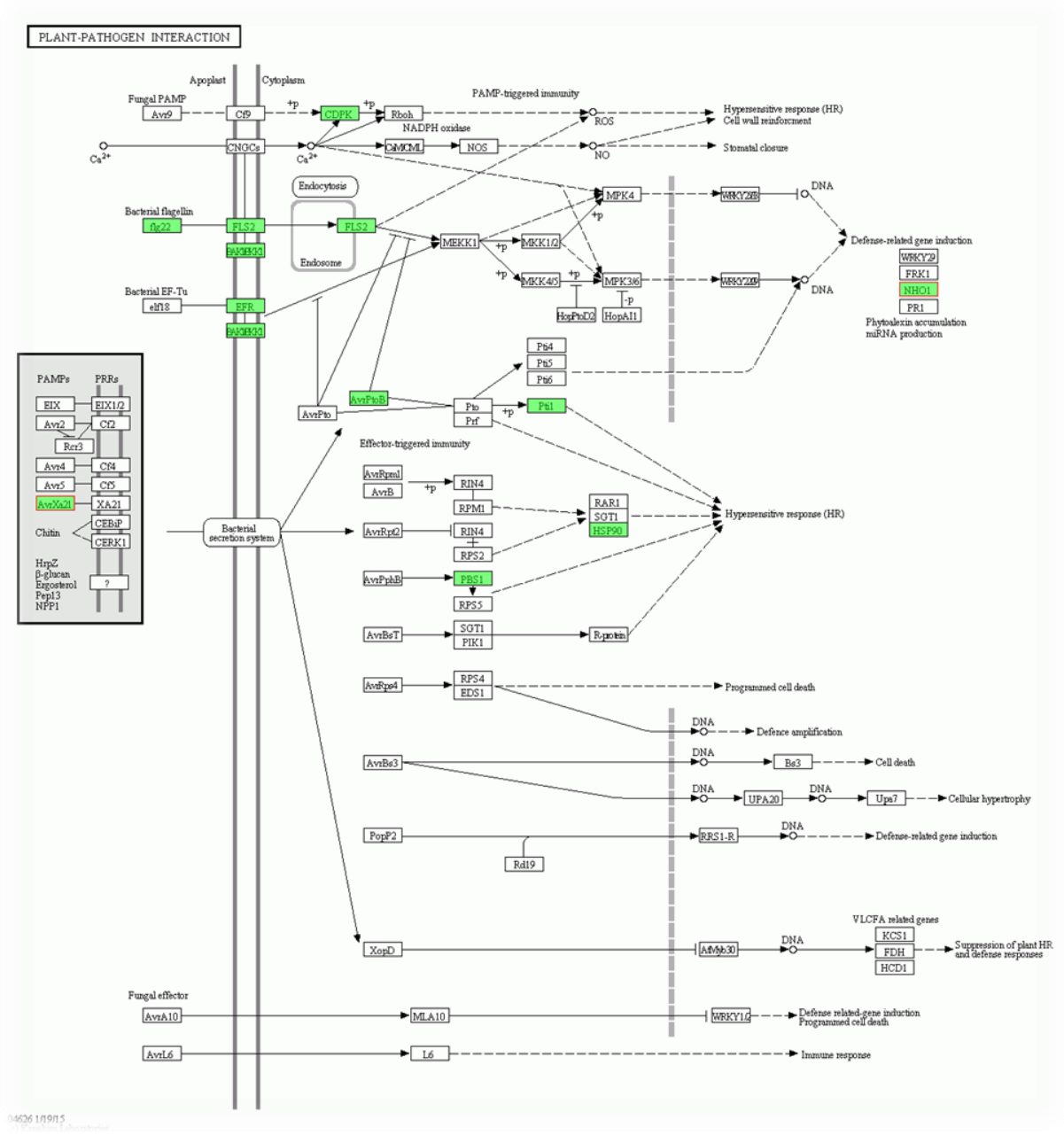


Figure 4.5: Environmental adaptation pathway of *P. aeruginosa* N002 and *Enterobacter* sp. RC4

4.10. Xenobiotic degradation genes

A total of 40 and 89 genes which are involped in diffrent xenobiotic degradation pathway such as Chloroalkane and chloroalkene degradation, Naphthalene degradation, Benzoate degradation, Aminobenzoate degradation, Benzoate degradation, Xylene degradation, Nitrotoluene degradation, Styrene degradation, Ethylbenzene degradation, Dioxin

degradation, Caprolactam degradation, Metabolism of xenobiotics by cytochrome P450, Atrazine degradation, Fluorobenzoate degradation, vanilate degradation, Toluene degradation and Naphthalene degradation pathways are described in (Table 4.8, Figure 4.6.A to Figure 4.6.N). The present of respective genes able to degreded diffren hydrocarbon suggest the suitability of *P. aeruginosa* N002 and *Enterobacter* sp. RC4 for crude oil degradation.

Table 4.8: Xenobiotic degradation genes of *P. aeruginosa* N002 and *Enterobacter* sp. RC4.

***P. aeruginosa* N002**		***Enterobacter* sp. RC4**	
Locus Tag	**Gene Product Name**	**Locus Tag**	**Gene Product Name**
Chloroalkane & chloroalkene degrading genes		**Chloroalkane, Chloroalkene and Naphthalene degrading genes**	
A222_00942	NAD-dependent aldehyde dehydrogenases (exaC)	C4L14_15085	propanol-preferring alcohol dehydrogenase
A222_01344	Transcriptional regulator, lysR	C4L14_16310	acetaldehyde dehydrogenase /alcohol dehydrogenase AdhE
A222_01345	S-(hydroxymethyl)glutathione dehydrogenase/class III alcohol dehydrogenase (frmA)	C4L14_06630	aldehyde dehydrogenase (NAD+)
A222_02834	3-Hydroxyisobutyrate dehydrogenase and related beta-hydroxyacid dehydrogenases (mmsB)	C4L14_23700	S-(hydroxymethyl) glutathione dehydrogenase/alcohol dehydrogenase
A222_02938	Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily) (dehH)	**Benzoate degrading genes**	
A222_03041	NAD-dependent aldehyde dehydrogenases (aldedh)	C4L14_03010	acetyl-CoA acetyltransferase
A222_03042	Cytochrome C	C4L14_13510	short chain enoyl-CoA hydratase
A222_03043	PQQ-dependent dehydrogenase, methanol/ethanol family (exaA)	C4L14_09380	short chain enoyl-CoA hydratase
A222_03895	Alcohol dehydrogenase, class IV (Fe-adh)	C4L14_16240	3-ketoacyl-CoA thiolase

A222_04019	NAD-dependent aldehyde dehydrogenases (Aldedh)	C4L14_13500	3-hydroxyacyl-CoA dehydrogenase
A222_04245	2-Haloalkanoic acid dehalogenase, type II (had-2)	C4L14_16235	3-hydroxyacyl-CoA dehydrogenase/enoyl-CoA hydratase/3-hydroxybutyryl-CoA epimerase
A222_05081	NAD-dependent aldehyde dehydrogenases (aldedh)	C4L14_09385	3-ketoacyl-CoA thiolase
A222_05624	Formaldehyde dehydrogenase, glutathione-independent (fdhA)	C4L14_15230	4-oxalocrotonate tautomerase
A222_05630	Zn-dependent alcohol dehydrogenases (adhP)	C4L14_19370	acetyl-CoA C-acetyltransferase
A222_05631	Transcriptional regulator, lysR	**Aminobenzoate degrading genes**	
Benzoate/Xylene degrading genes		C4L14_05120	vanillate O-demethylase monooxygenase subunit
A222_00159	Protocatechuate 3,4-dioxygenase, beta subunit (pcaH)	C4L14_05125	vanillate O-demethylase ferredoxin subunit
A222_00160	Protocatechuate 3,4-dioxygenase, alpha subunit (pcaG)	C4L14_05175	Acylphosphatase
A222_00161	3-hydroxyacyl-CoA dehydrogenase (fadA)	C4L14_08335	acid phosphatase
A222_00226	3-oxoadipyl-CoA thiolase (pcaD)	C4L14_13510	short chain enoyl-CoA hydratase
A222_00227	Metabolite-proton symporter (pcaT MFS transporter, MHS family, dicarboxylic acid transporter PcaT)	**Nitrotoluene degrading genes**	
A222_00228	3-Carboxy-cis, cis-muconate cycloisomerase (pcaB)	C4L14_08780	nitroreductase/dihydropteridine reductase
A222_00229	3-Oxoadipate enol-lactonase (pcaD)	C4L14_13360	Nitroreductase
A222_00230	4-Carboxymuconolactone decarboxylase (pcaC)	C4L14_19000	N-ethylmaleimide reductase
A222_00231	Transcriptional regulator, lysR	**Styrene degrading genes**	
A222_00245	4-Hydroxybenzoate 3-monooxygenase (pobA)	C4L14_13570	phenylacetaldehyde dehydrogenase
A222_00491	3-Oxoadipate enol-lactonase (pcaD)	**Ethylbenzene degrading genes**	
A222_01040	Acetyl-CoA acetyltransferases (atoB)	C4L14_09385	3-ketoacyl-CoA thiolase

A222_01385	3-hydroxyacyl-CoA dehydrogenase (fadB)	C4L14_16240	3-ketoacyl-CoA thiolase
A222_01386	Acetyl-CoA acetyltransferases (atoB)	**Xylene degradation and Dioxin degrading genes**	
A222_01504	Acetyl-CoA acetyltransferases (fadA)	C4L14_15230	4-oxalocrotonate tautomerase
A222_01532	Enoyl-CoA hydratase/carnithine racemase (paaF)	**Toluene, Fluorobenzoate, Chlorocyclohexane and Chlorobenzene degrading genes**	
A222_01953	Fatty oxidation complex, beta subunit FadA (fadA)	C4L14_09450	Carboxymethylenebutenolidase
A222_02033	Acetyl-CoA acetyltransferases (fadA)	**Caprolactam degradation**	
A222_02483	Acetyl-CoA acetyltransferases (atoB)	C4L14_09380	short chain enoyl-CoA hydratase
A222_02520	AraC-type DNA-binding domain-containing proteins	C4L14_13510	3-hydroxyacyl-CoA dehydrogenase/enoyl-CoA hydratase/3-hydroxybutyryl-CoA epimerase
A222_02521	Benzoate 1,2-dioxygenase, large subunit (benA/xylX)	C4L14_16235	3-hydroxyacyl-CoA dehydrogenase/enoyl-CoA hydratase/3-hydroxybutyryl-CoA epimerase
A222_02522	Benzoate 1,2-dioxygenase, small subunit (benB/xylY)	**cytochrome P450 metabolism of xenobiotics degrading genes**	
A222_02523	2-Polyprenylphenol hydroxylase and related flavodoxin oxidoreductases (benC/xylZ)	C4L14_08240	glutathione S-transferase
A222_02524	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases) (bend/xylL)	C4L14_13280	glutathione S-transferase
A222_02529	Transcriptional regulator, lysR	C4L14_13500	3-hydroxyacyl-CoA dehydrogenase
A222_02530	Muconate and chloromuconate cycloisomerases (catB)	C4L14_15085	propanol-preferring alcohol dehydrogenase
A222_02531	Muconolactone delta-isomerase (catC)	C4L14_18915	glutathione S-transferase
A222_02532	Catechol 1,2-dioxygenase, proteobacterial (catA)	C4L14_20985	glutathione S-transferase
A222_03208	Enoyl-CoA hydratase/carnithine	C4L14_23700	S-(hydroxymethyl)glutathi

	racemase (paaF)		one dehydrogenase/alcohol dehydrogenase
A222_03288	Enoyl-CoA hydratase/carnithine racemase (echA)	**Atrazine degradation**	
A222_03300	Acetyl-CoA acetyltransferases (atoB)	C4L14_04005	urease subunit gamma
A222_03417	3-Hydroxyacyl-CoA dehydrogenase (fadB)	C4L14_04010	urease subunit beta
A222_04077	Tol-pal system-associated acyl-CoA thioesterase (4-hbt)	C4L14_04015	urease subunit alpha
A222_04961	Acetyl-CoA acetyltransferases (atoB)	C4L14_13715	allophanate hydrolase
A222_05588	3-Hydroxyacyl-CoA dehydrogenase (fadB)		
Fluorobenzoate degrading genes			
A222_02328	Dienelactone hydrolase and related enzymes (ysgA)		
A222_02520	AraC-type DNA-binding domain-containing proteins		
A222_02521	Benzoate 1,2-dioxygenase, large subunit (benA)		
A222_02522	Benzoate 1,2-dioxygenase, small subunit (benB)		
A222_02523	2-Polyprenylphenol hydroxylase and related flavodoxin oxidoreductases (benC)		
A222_02524	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases) (benD)		
A222_02530	Muconate and chloromuconate cycloisomerases (catB)		
A222_02532	Catechol 1,2-dioxygenase, proteobacterial (catA)		
Aminobenzoate & vanilate degrading genes			
A222_00803	Asp-tRNAAsn/Glu-tRNAGln amidotransferase A		

	subunit and related amidases (amiE)	
A222_00868	Rieske [2Fe-2S] domain (vanA)	
A222_00869	Flavodoxin reductases (ferredoxin-NADPH reductases) family 1 (vanB)	
A222_01055	Phosphodiesterase/alkaline phosphatase D (phoD)	
A222_01532	Enoyl-CoA hydratase/carnithine racemase (paaF)	
A222_01597	Predicted amidohydrolase (amiE)	
A222_01632	Cytochrome P450 (p450)	
A222_01670	Alkaline phosphatase (phoA)	
A222_02525	Flavodoxin reductases (ferredoxin-NADPH reductases) family 1 (antC)	
A222_02526	Anthranilate 1,2-dioxygenase, small subunit (antB)	
A222_02527	Anthranilate 1,2-dioxygenase, large subunit (antA)	
A222_02528	AraC-type DNA-binding domain-containing proteins	
A222_02564	Cytochrome P450 (p450)	
A222_03208	Enoyl-CoA hydratase/carnithine racemase (paaF)	
A222_03288	Enoyl-CoA hydratase/carnithine racemase (echA)	
A222_04091	Acylphosphatases (acyP)	
A222_04347	Asp-tRNAAsn/Glu-tRNAGln amidotransferase A subunit and related amidases (amiE)	
A222_04470	Asp-tRNAAsn/Glu-tRNAGln amidotransferase A subunit and related amidases (amiE)	

A222_05083	Thiamine pyrophosphate-requiring enzymes [acetolactate synthase, pyruvate dehydrogenase (cytochrome), glyoxylate carboligase, phosphonopyruvate decarboxylase] (mdlC)	
A222_05086	Phenylpropionate dioxygenase and related ring-hydroxylating dioxygenases, large terminal subunit (vanA)	
A222_05087	Flavodoxin reductases (ferredoxin-NADPH reductases) family 1 (vanB)	
A222_05088	Transcriptional regulator, VanR, GntR family	
Toluene degrading genes		
A222_02328	Dienelactone hydrolase and related enzymes (dlh)	
A222_02530	Muconate and chloromuconate cycloisomerases (catB)	
A222_02532	Catechol 1,2-dioxygenase, proteobacterial (catA)	
A222_03463	Succinate dehydrogenase and fumarate reductase iron-sulfur protein (sdhB)	
A222_03464	Succinate dehydrogenase, flavoprotein subunit, E. coli/mitochondrial subgroup (sdhA)	
A222_03465	Succinate dehydrogenase, hydrophobic membrane anchor protein (sdhD)	
A222_03466	Succinate dehydrogenase, cytochrome b556 subunit (sdhC)	
Nitrotoluene degrading genes		
A222_02042	NADH:flavin oxidoreductases, Old Yellow Enzyme family (nemA)	
A222_03701	NADH:flavin oxidoreductases, Old Yellow Enzyme family (nemA)	

Styrene degrading genes		
A222_00223	XRE family with cupin sensor transcriptional regulator	
A222_00224	Acyl CoA: acetate/3-ketoacid CoA transferase, alpha subunit (gctA)	
A222_00225	Acyl CoA: acetate/3-ketoacid CoA transferase, beta subunit (gctB)	
A222_00803	Asp-tRNAAsn/Glu-tRNAGln amidotransferase A subunit and related amidases (amiE)	
A222_00890	NAD-dependent aldehyde dehydrogenases (feaB)	
A222_01597	Predicted amidohydrolase (amiE)	
A222_02566	maleylacetoacetate isomerase (maiA)	
A222_03015	Transcriptional regulator, IclR family	
A222_03016	Homogentisate 1,2-dioxygenase (HgmA)	
A222_03017	Fumarylacetoacetase (fahA)	
A222_03018	Maleylacetoacetate isomerase (maiA)	
A222_04347	Asp-tRNAAsn/Glu-tRNAGln amidotransferase A subunit and related amidases (amiE)	
A222_04470	Asp-tRNAAsn/Glu-tRNAGln amidotransferase A subunit and related amidases (amiE)	
Ethylbenzene degrading genes		
A222_01504	Acetyl-CoA acetyltransferases (fadA)	
A222_01953	Fatty oxidation complex, beta subunit FadA (fadA)	
A222_02033	Acetyl-CoA acetyltransferases (fadA)	
Naphthalene degrading genes		
A222_01345	S-(hydroxymethyl) glutathione	

	dehydrogenase/class III alcohol dehydrogenase (frmA)	
A222_05630	Zn-dependent alcohol dehydrogenases (adhP)	
Chlorocyclohexane and chlorobenzene degrading genes		
A222_02328	Dienelactone hydrolase and related enzymes (dlh)	
A222_02530	Muconate and chloromuconate cycloisomerases (catB)	
A222_02532	Catechol 1,2-dioxygenase, proteobacterial (catA)	
A222_02938	Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily) (dehH)	
A222_04245	2-Haloalkanoic acid dehalogenase, type II (had-2)	
Caprolactam degrading genes		
A222_01532	Enoyl-CoA hydratase/carnithine racemase (paaF)	
A222_01952	Fatty oxidation complex, alpha subunit FadB (fadB)	
A222_02816	NAD-dependent aldehyde dehydrogenases (aldH)	
A222_03208	Enoyl-CoA hydratase/carnithine racemase (paaF)	
A222_03288	Enoyl-CoA hydratase/carnithine racemase (echA)	
A222_03299	3-Hydroxyacyl-CoA dehydrogenase (fadJ)	
A222_03414	Acyl-CoA dehydrogenases (DCAA)	
A222_03785	NAD-dependent aldehyde dehydrogenases (aldH)	

All details chloroalkane and chloroalkene degrading pathway, polycyclic aromatic hydrocarbon degradation pathway, Chlorocyclohexane and chelrobenzene degrading pathway, fluorobenzoate degradatioon pathway, xylene degradation pathway, Styrene degradation pathway, Aminobenzoate degradation pathway , caprolactam degradation

pathway., Toluene degradation pathway, Chlorocyclohexane and chelrobenzene degradation pathway, and Benzoate degradation pathway are showing in the (Figure 4.6.A to Figure 4.6.N)

4.10.a: Chloroalkane and Chloroalkene utilization

In Chloroalkane and Chloroalkene degradation pathways both *P. aeruginosa* N002 and *Enterobacter* sp. RC4 led to utilize in pyruvate metabolism as well as glycolate (Figure 4.8.A and 4.8.B). In *P. aeruginosa* N002 enzymes 1.1.1.1, 1.2.13 (Alcohol dehydrogenase, aldehyde dehydrogenase) from *dhP, ALDH* genes were involved in pyruvate cycle while enzyme 1.1.2.8, 1.2.1.3, 3.8.1.3 (Alcohol dehydrogenase cytochrome c, Aldehyde dehydrogenase NAD+,Haloacetate dehalogenase) from *exaA, ALDH, dehH* were involved in glycolate systems. Instead *Enterobacter* sp. RC4 enzymes 1.1.1.1, 1.2.13, 4.1.1 (Alcohol dehydrogenase, aldehyde dehydrogenase) from *dhP, ALDH, msaD* was involved in pyruvatemetabolism while enzyme 1.2.1.3 from *ALDH* genes in glycolate.

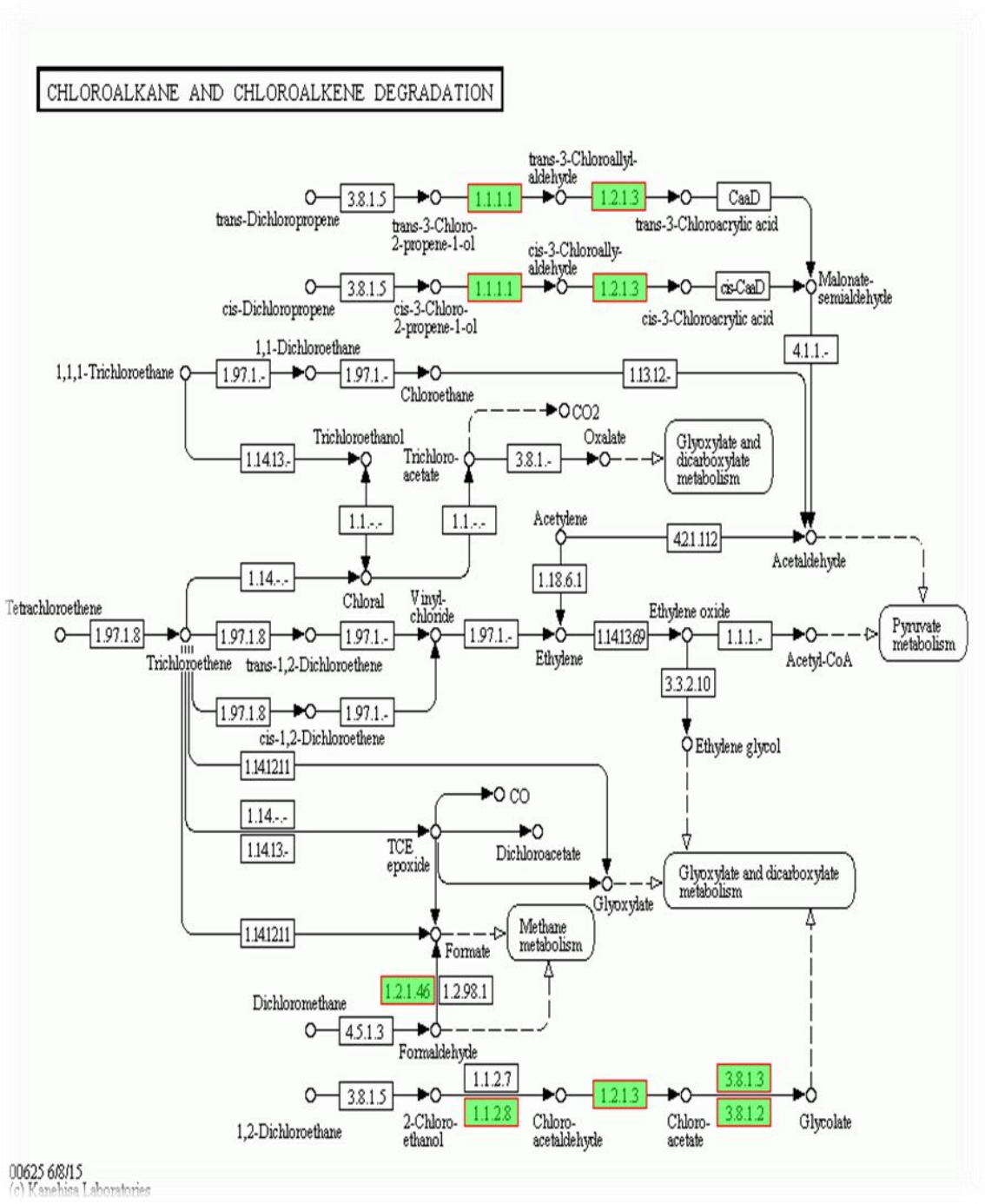


Figure 4.6.A: ***P. aeruginosa*** **N002 genome chloroalkane and chloroalkene degrading pathway**

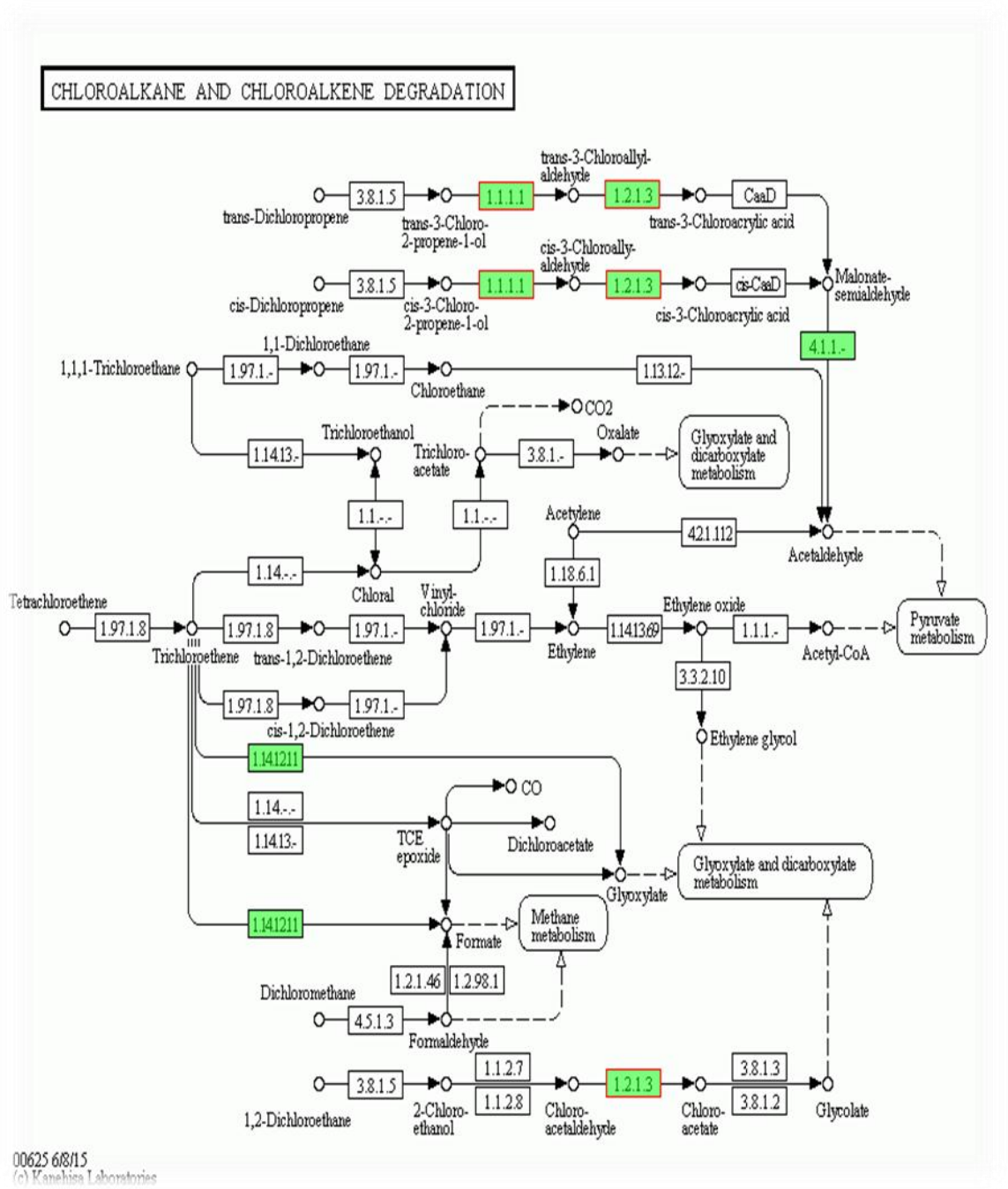


Figure 4.6.B: *Enterobacter* sp. RC4 genome chloroalkane and chloroalkene degrading pathway

4.10.b: Polycyclic aromatic hydrocarbon utilization

In Polycyclic aromatic hydrocarbon degradation pathwaysin *P. aeruginosa* N002 enzyme 1.13.11, 1.2.1, 1.14.12.7,1.3.1 (Phthalate 4,5-dioxygenase) from *nidA, phdF, nidD, pht2, phtB* genes utilized phenanthrene and *DbfA* genes involved in fluorine

degradation. *Enterobacter* sp. RC4 led to utilize enzyme 1.13.11, 1.2.1, 1.14.12.7, 1.3.1, 4.1.2 (Phthalate 4,5-dioxygenase) from *nidA, phdF, nidD, pht2, phtB, phdG* genes utilized phenanthrene which are showing in the figure 4.6.C and 4.6.D.

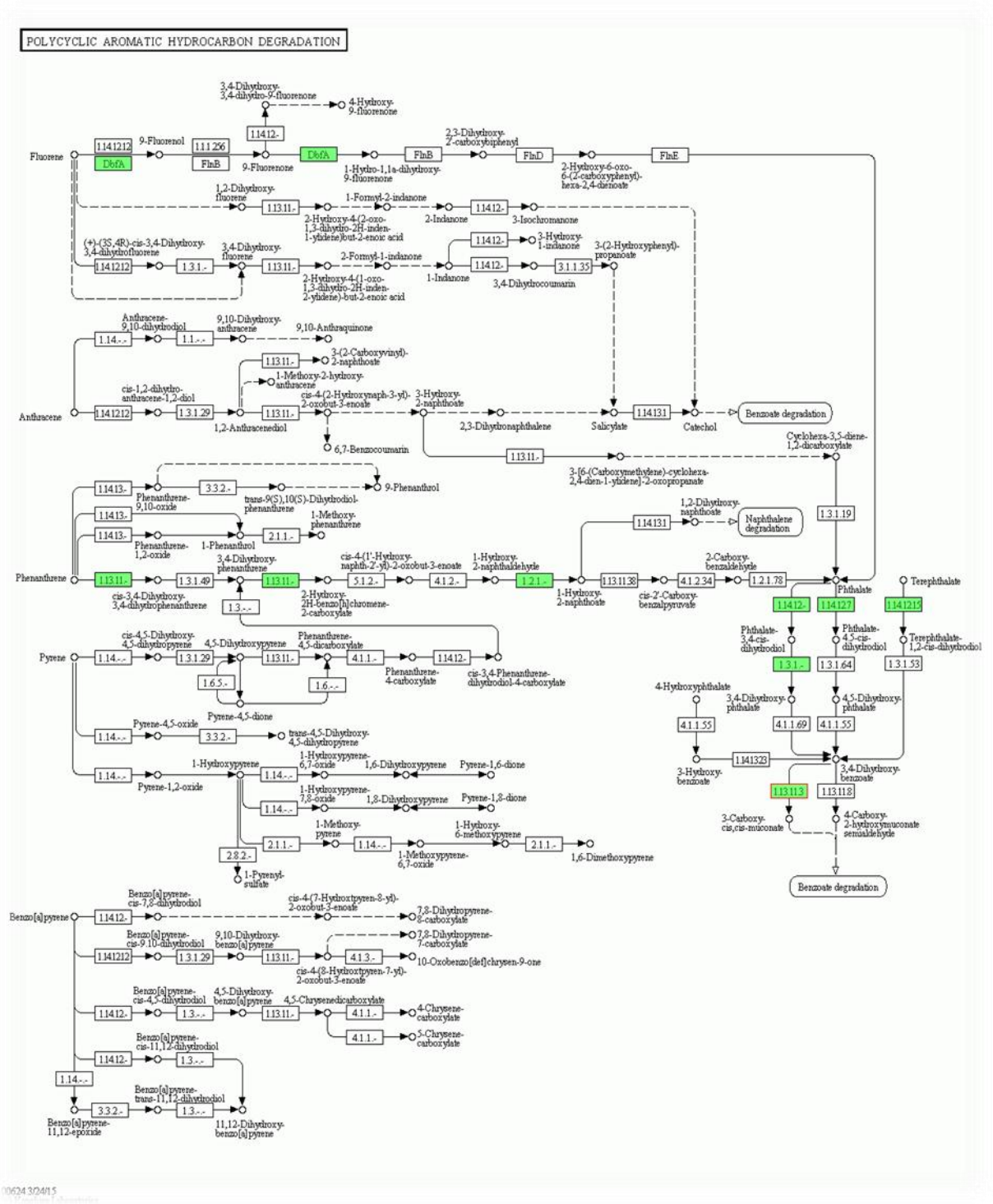


Figure 4.6.C: ***P aeruginosa*** **N002 genome polycyclic aromatic hydrocarbon degradation pathway**

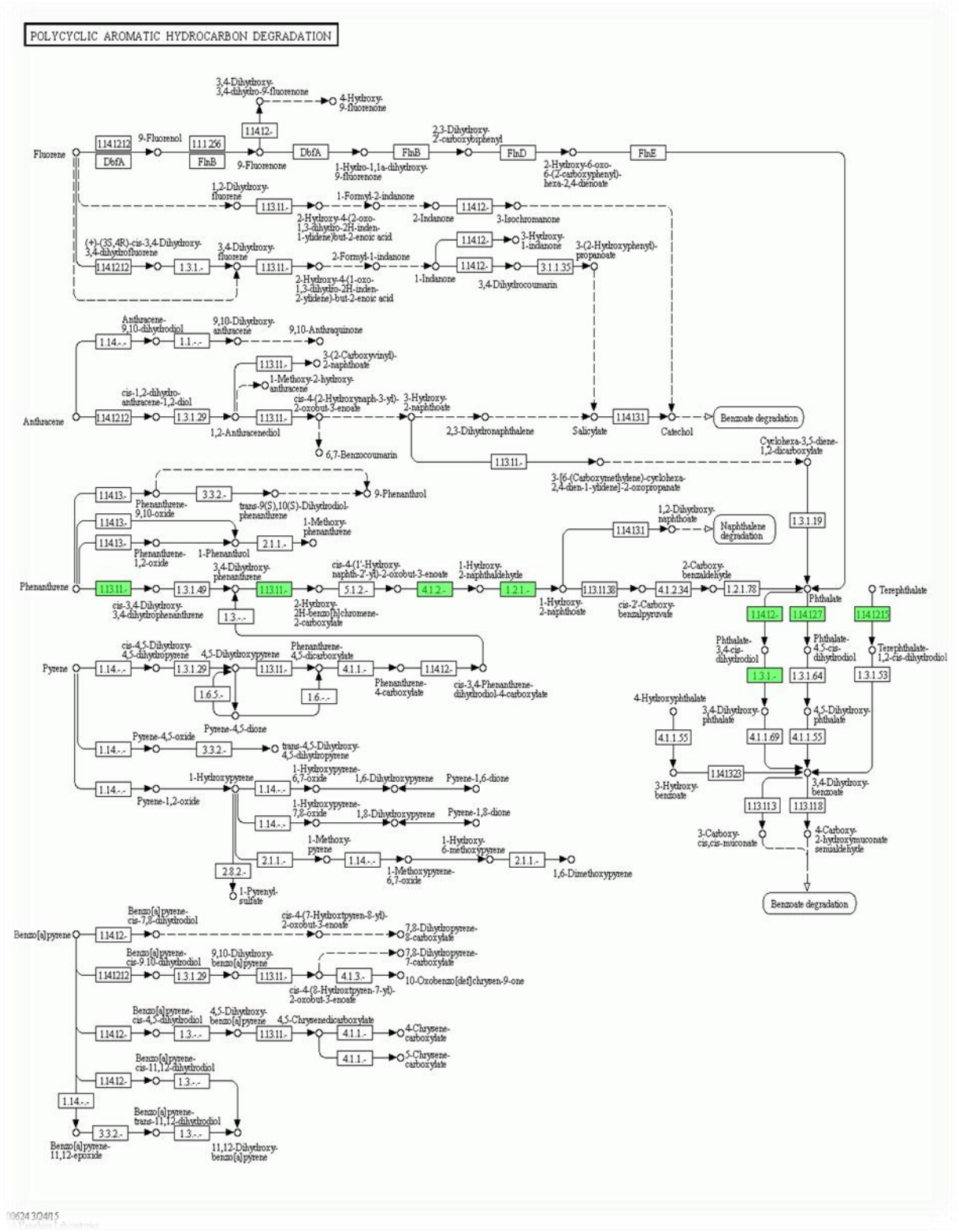


Figure 4.6.D: ***Enterobacter*** **sp. RC4 genome polycyclic aromatic hydrocarbon degrading pathway**

4.10.c: Chlorocyclohexane and chelrobenzene utilization

In Chlorocyclohexane and chelrobenzene degradation pathways in both the strain *P. aeruginosa* N002 and *Enterobacter sp.* RC4 utilize 2,4-Dichloro-phenoxyaceate with the help of genes TidA,TdfC and convert into maleylacetate which are showing in the figure 4.6.E and 4.6.F.

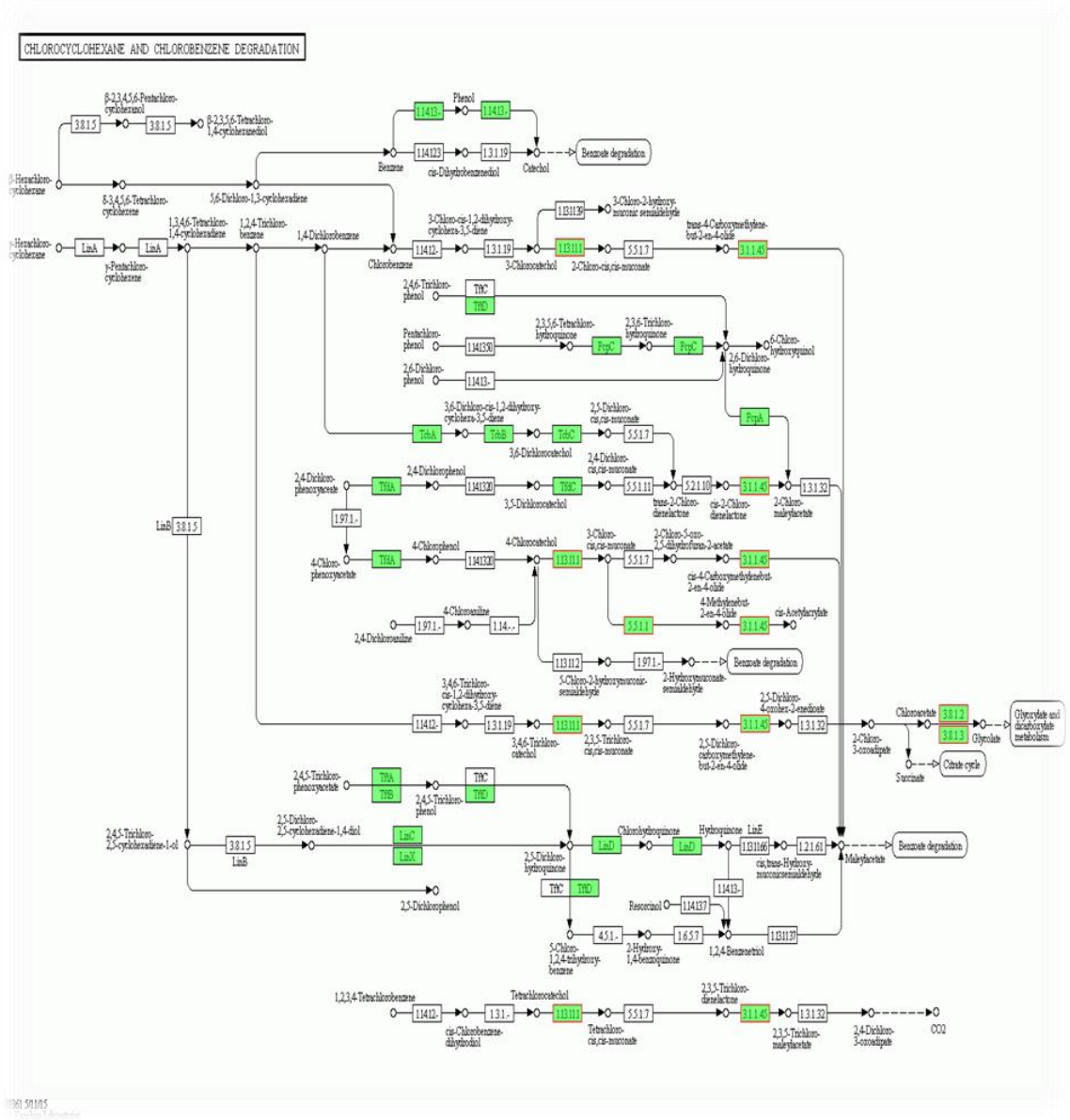


Figure 4.6.E: ***P. aeruginosa*** **N002 genome Chlorocyclohexane and chelrobenzene degrading pathway**

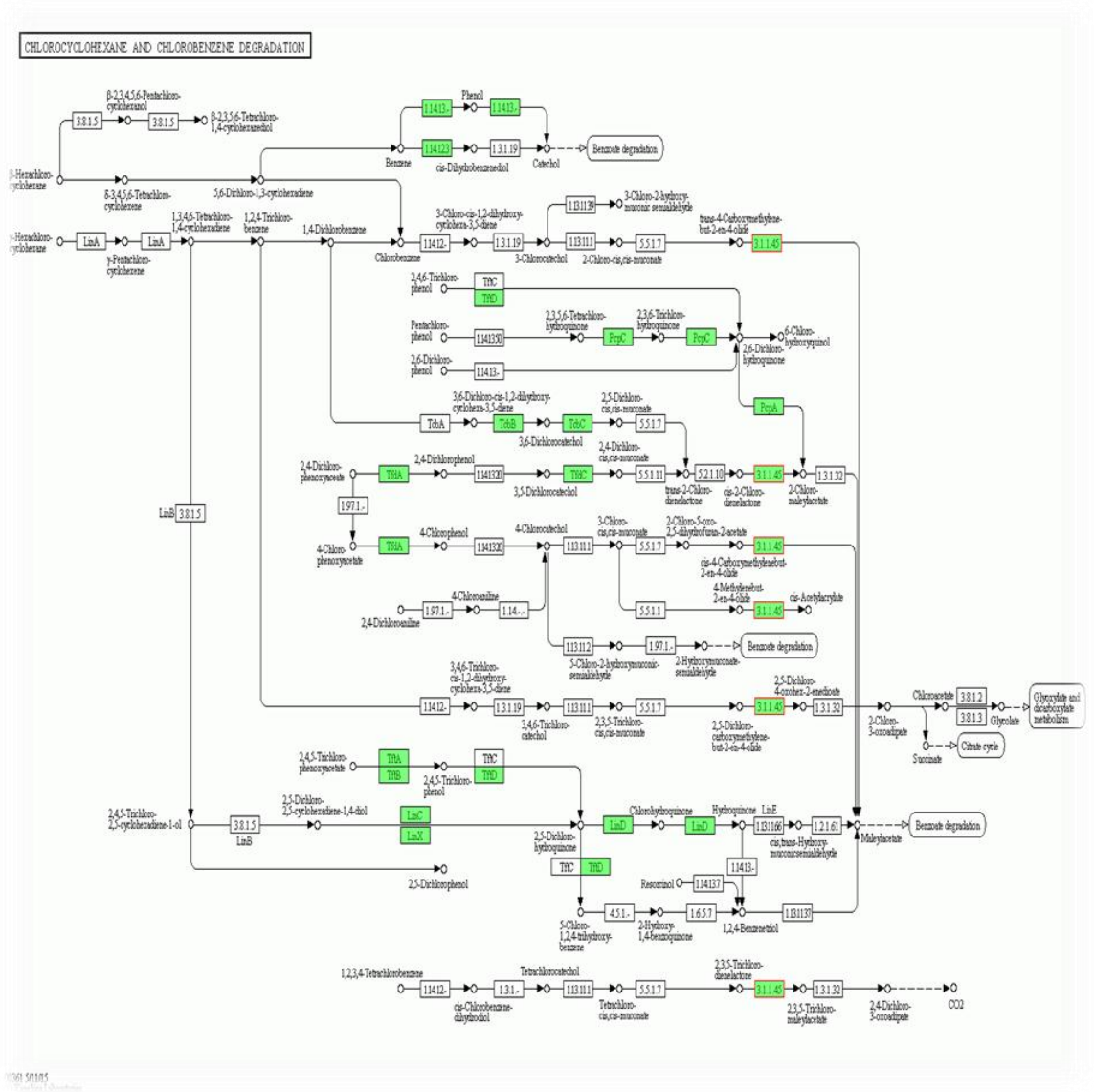


Figure 4.6.F: ***Enterobacter*** **sp. RC4 genome Chlorocyclohexane and chelrobenzene degradation pathway**

4.10.d: Benzoate utilization

In Benzoate degradation pathways the strain *P. aeruginosa* N002 utilize benzoate with the help of enzyme 1.14.12.10, 1.3.1.25, 1.12.11.1, 5.5.1.1, 5.3.3.4, 3.1.1.24, 2.3.1.16, 23.1.174 (Benzoate 1,2-dioxygenase, 1,6-dihydroxycyclohexa-2,4-diene-1-carboxylate dehydrogenase, Muconate cycloisomerase) from *benA, benD, catA, catB, catC, pcaD, fadA, pcaF* genes to converted into succinyl-CoA and strain *Enterobacter* sp. RC4 utilize benzoate with the help of enzyme 1.14.12.10, 1.3.1.25 (Benzoate-1,2-

dioxygenase, 1,6-dihydroxycyclohexa-2,4-diene-1-carboxylate dehydrogenase) from *benA, benD* genes to convert catechol which are described in Figure 4.6.G and 4.6.H.

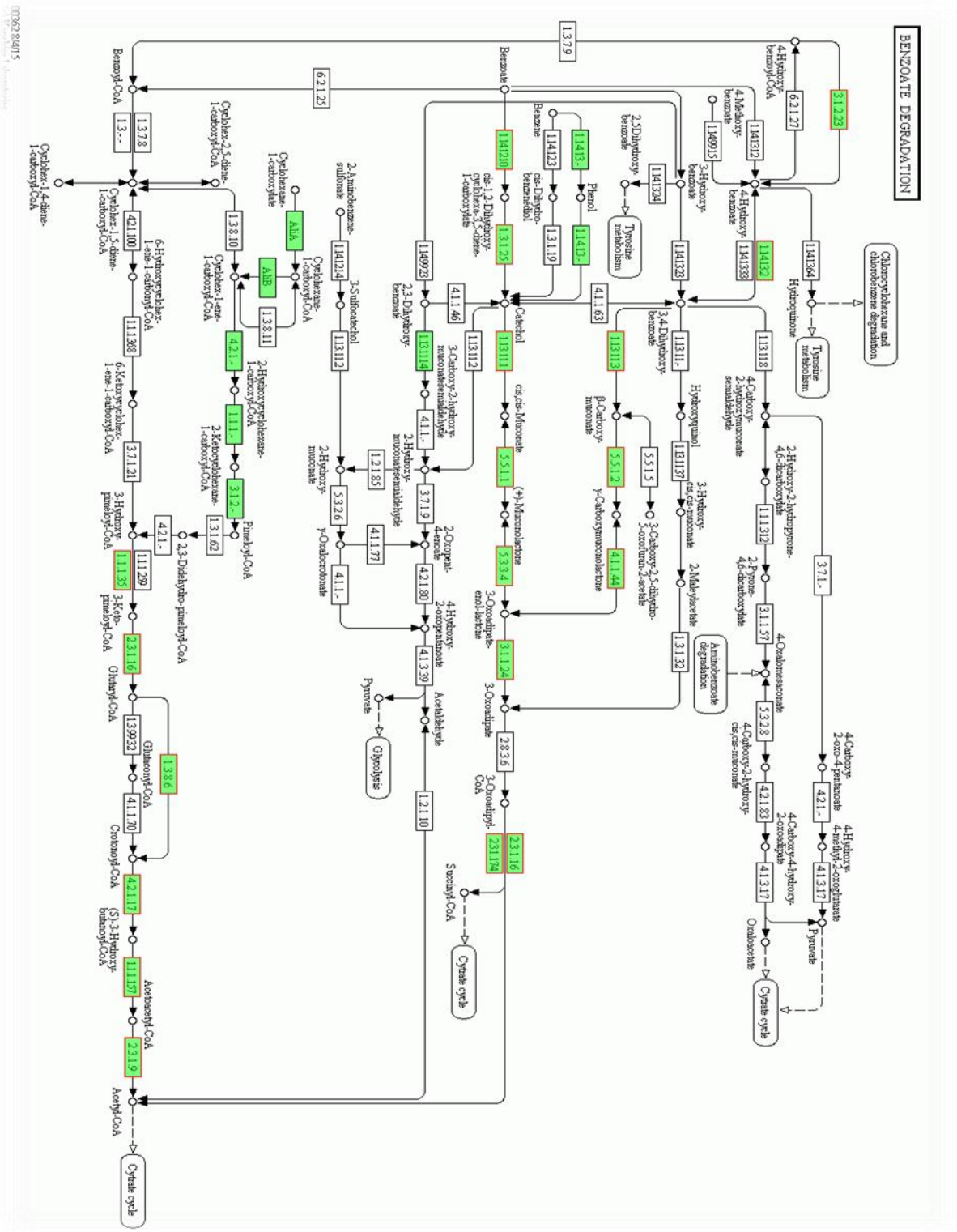


Figure 4.6.G: ***P. aeruginosa*** **N002 Benzoate degradation pathway**

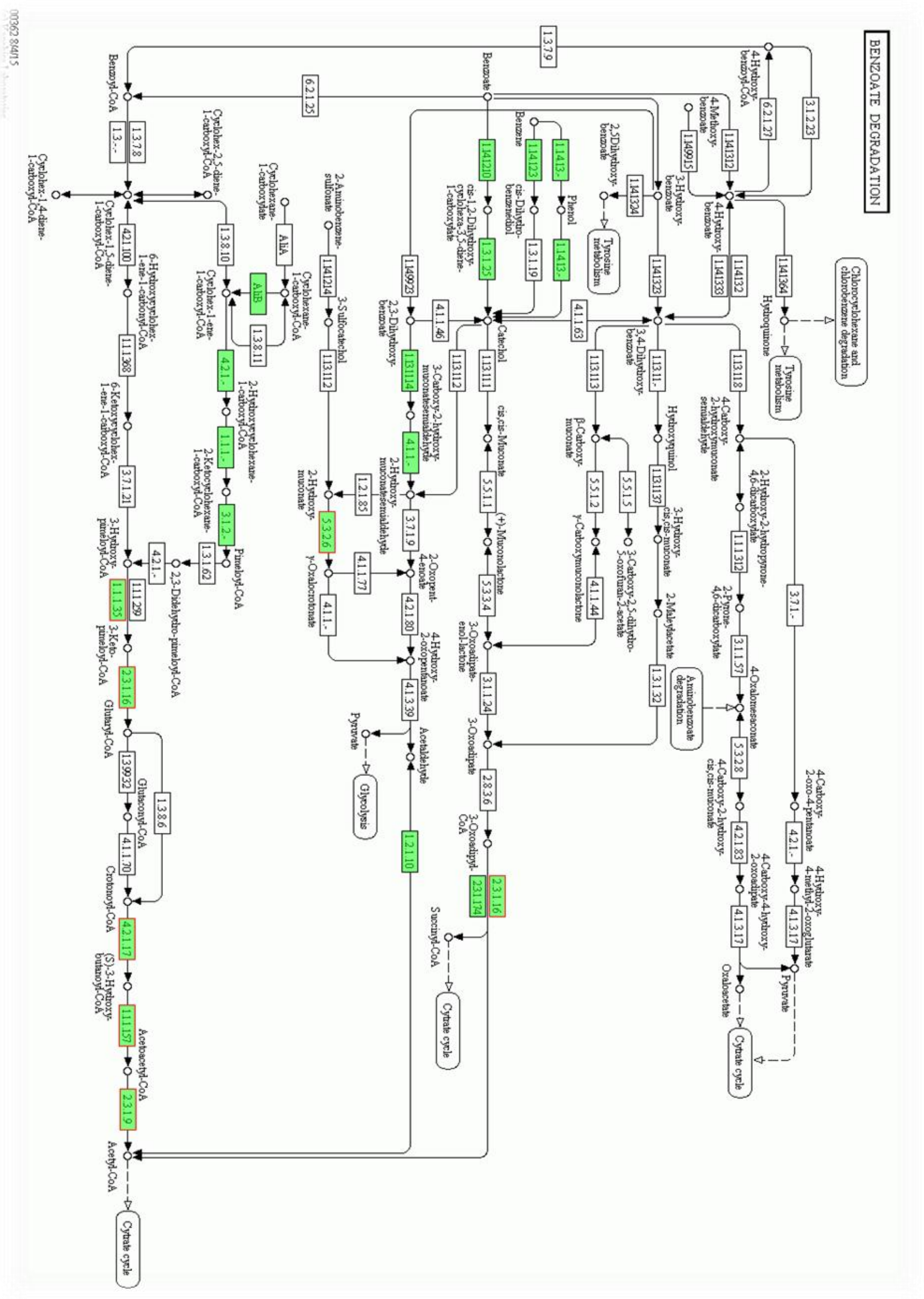


Figure 4.6.H: *Enterobacter* **ps. RC4 genome benzoate degradation pathway**

4.10.e: Fluorobenzoate utilization

In Fluorobenzoate degradation pathways *P. aeruginosa* N002 utilize 2-Fluorobenzoate, 3-Fluorobenzoate, 4-Fluorobenzoate with the help of enzyme 1.14.12.10, 1.3.1.25, 1.13.11.1, 5.5.1.1 (Benzoate 1,2-dioxygenase, 1,6-dihydroxycyclohexa-2,4-diene-1-carboxylate dehydrogenase, Catechol 1,2-dioxygenase) from *benA*, *bend*, *catA*, *catB* genes to convert Maleylacetate.

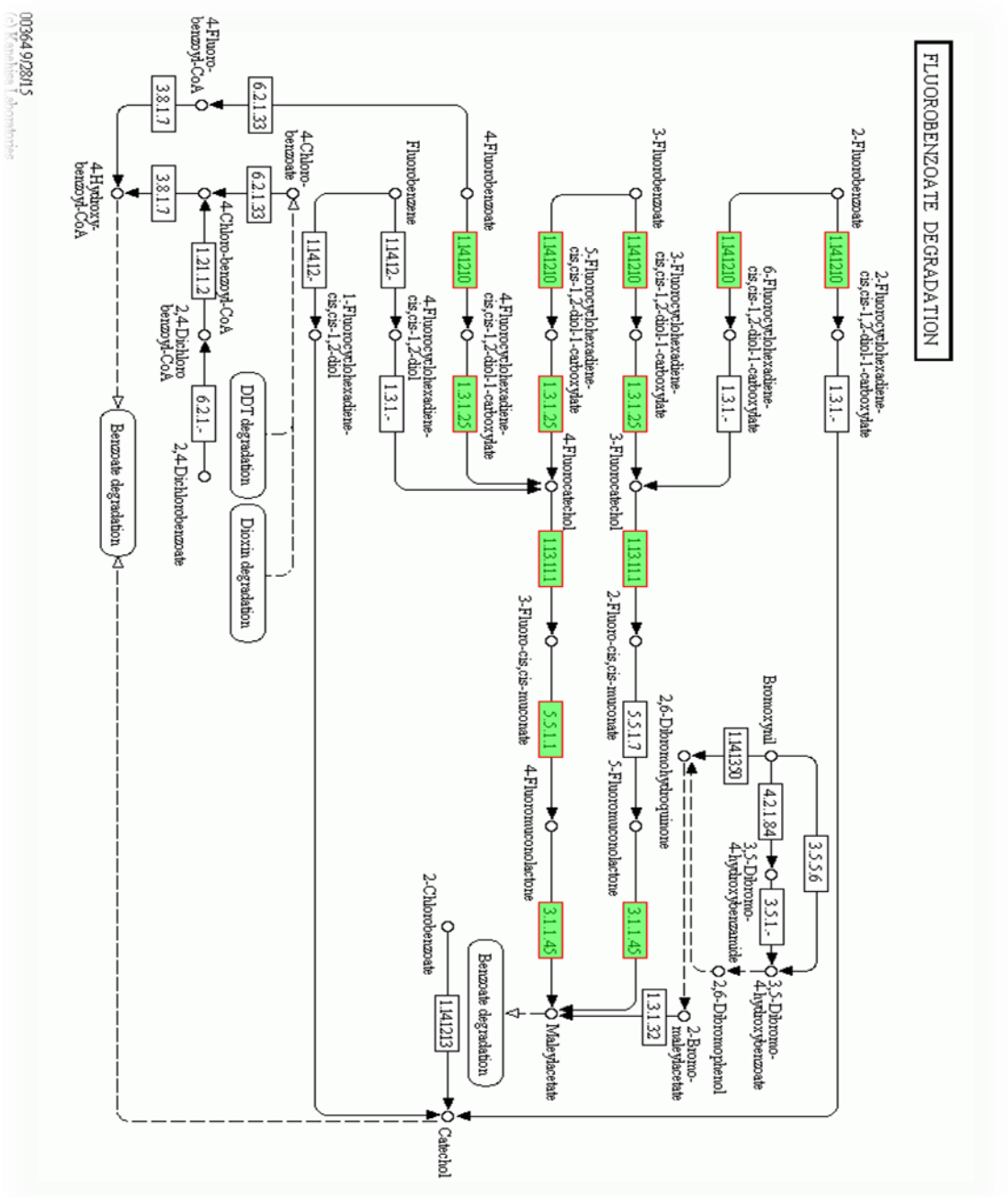


Figure 4.6.I: *P. aeruginosa* N002 genome fluorobenzoate degradatioon pathway

4.10.f: Xylene utilization:

In Xylene degradation pathways *P. aeruginosa* N002 utilize O-Xylene, P-Xylene, m-Xylene and Cymene with the help of *XylMA* and *CymA* geneswhich are described in figure 4.6.J.

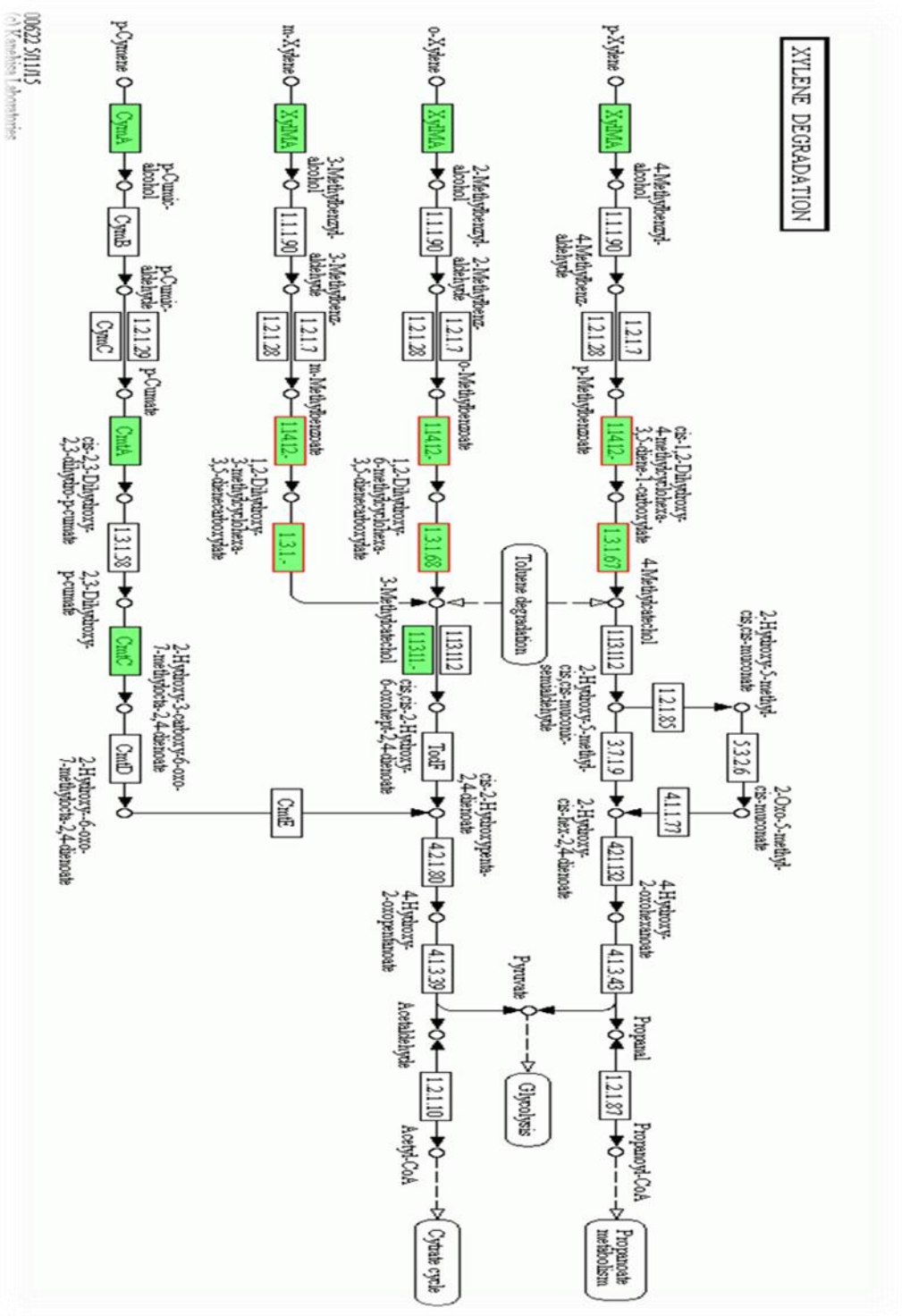


Figure 4.6.J: *P. aeruginosa* N002 genome xylene degradation pathway

4.10.g: Styrene utilization

In Styrene degradation pathways *P. aeruginosa* N002 utilize enzyme 3.5.1.4, 2.8.3.12, 1.2.1.39 (Glutaconate CoA-transferase, Phenylacetaldehyde dehydrogenase) from *gctA, feadB* genes in Pyruvate metabolism which are described in figure 4.6.K.

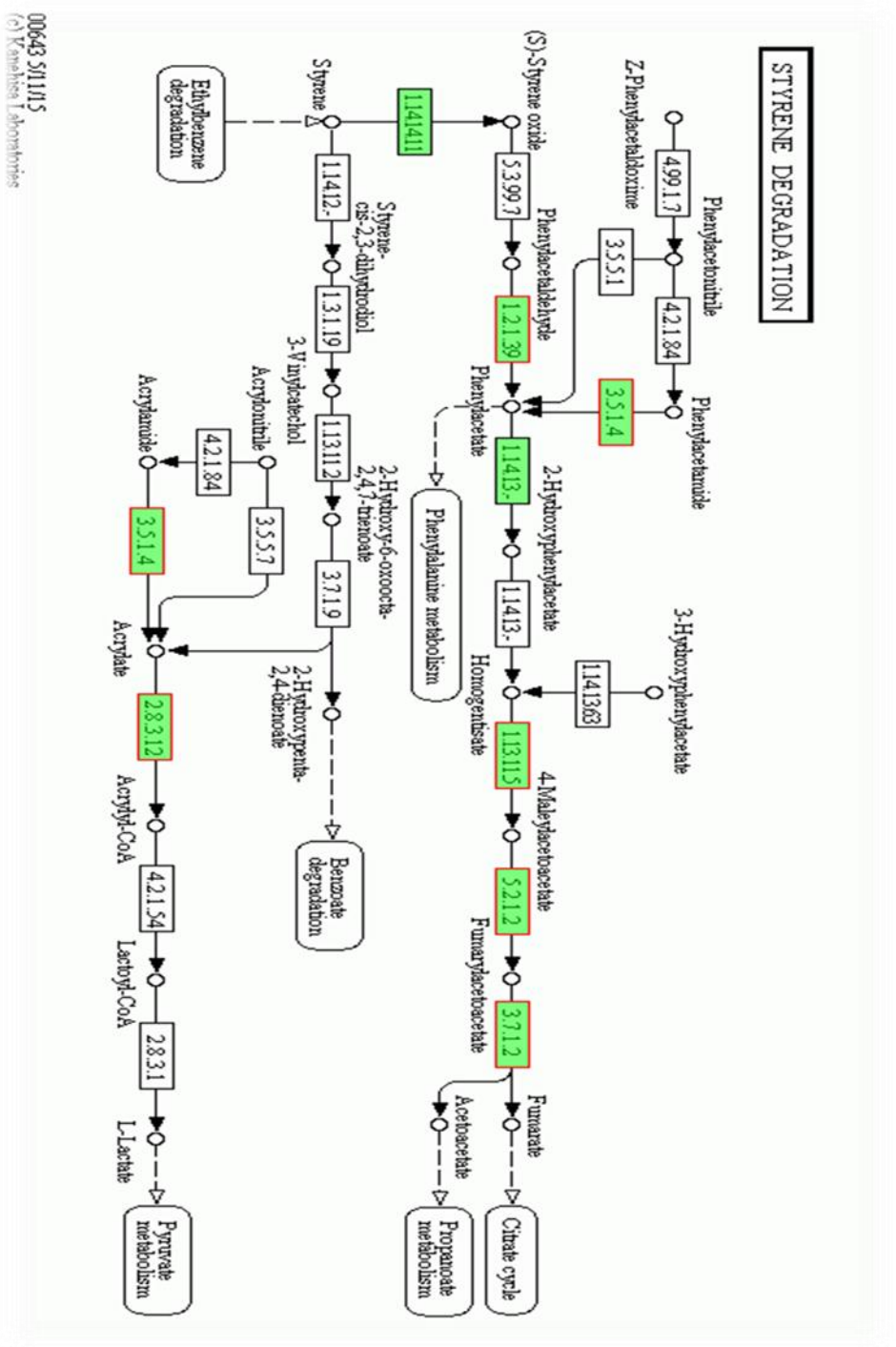


Figure 4.6.K: *P. aeruginosa* N002 genome Styrene degradation pathway

4.10.h: Aminobenzoate utilization

In aminobenzoate degradation pathways *Enterobactrer* sp. RC4 utilize enzyme 3.6.17, 1.14.12.1 (Androstene-3,6,17-trione, Anthranilate hydroxylase) from *acyP*, *antA*, *DesA*, *DesZ* genes involved in aminobenzoate friction degradation which are described in figure 4.6.L.

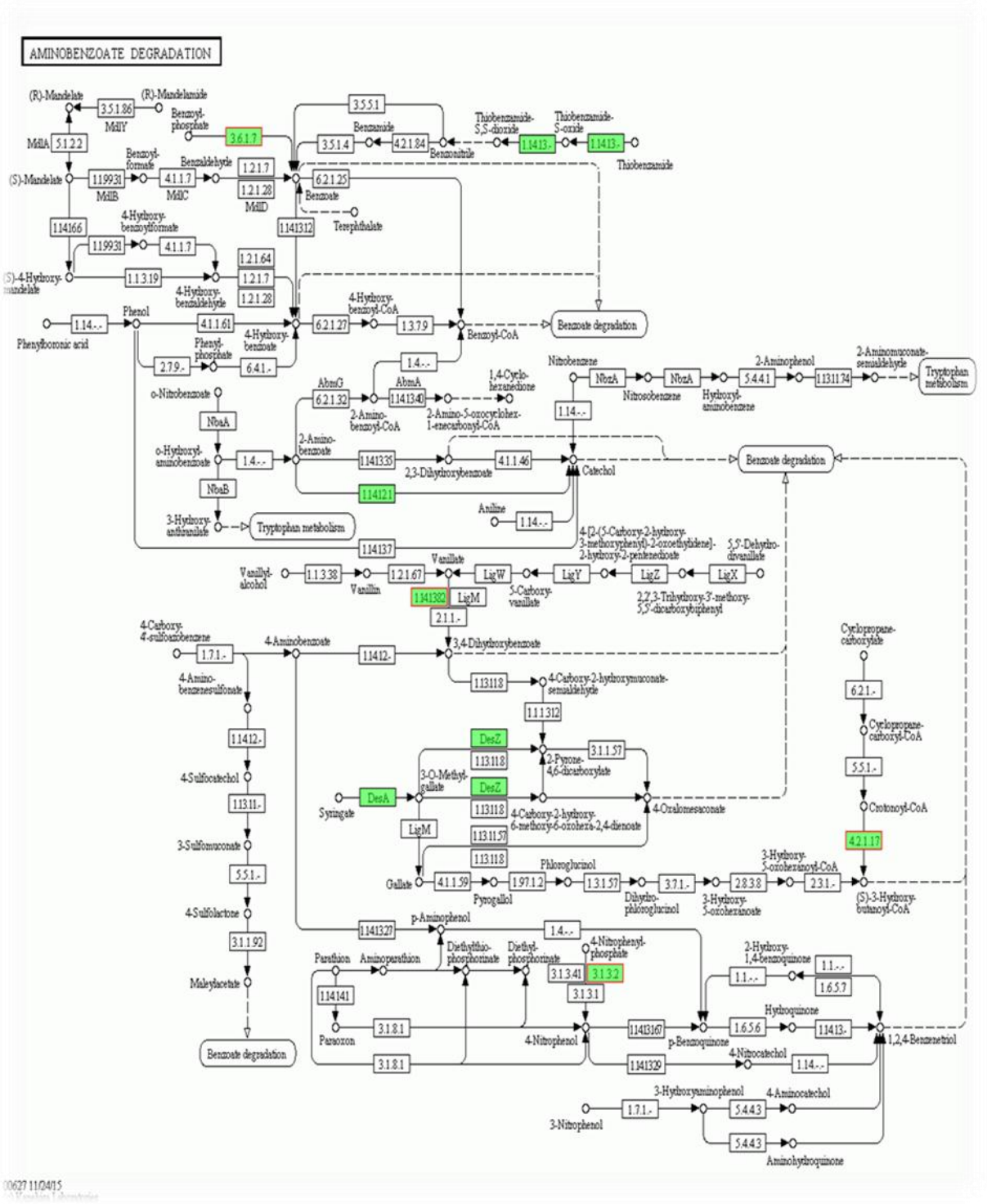


Figure 4.6.L: *Enterobacter* sp. RC4 genome aminobenzoate degradation pathway

4.10.i: Caprolactam utilization

In Caprolactam degradation pathways *Enterobactrer* sp. RC4 utilize enzyme 1.3.99, 4.2.1.17, 1.1.1.35 (3-oxosteroid 1-dehydrogenase, Enoyl-CoA hydratase, 3-hydroxyacyl-CoA dehydrogenase.) from DcaA, DcaE, fadB genes involved in Caprolactam friction degradation which are described in figure 4.6.M.

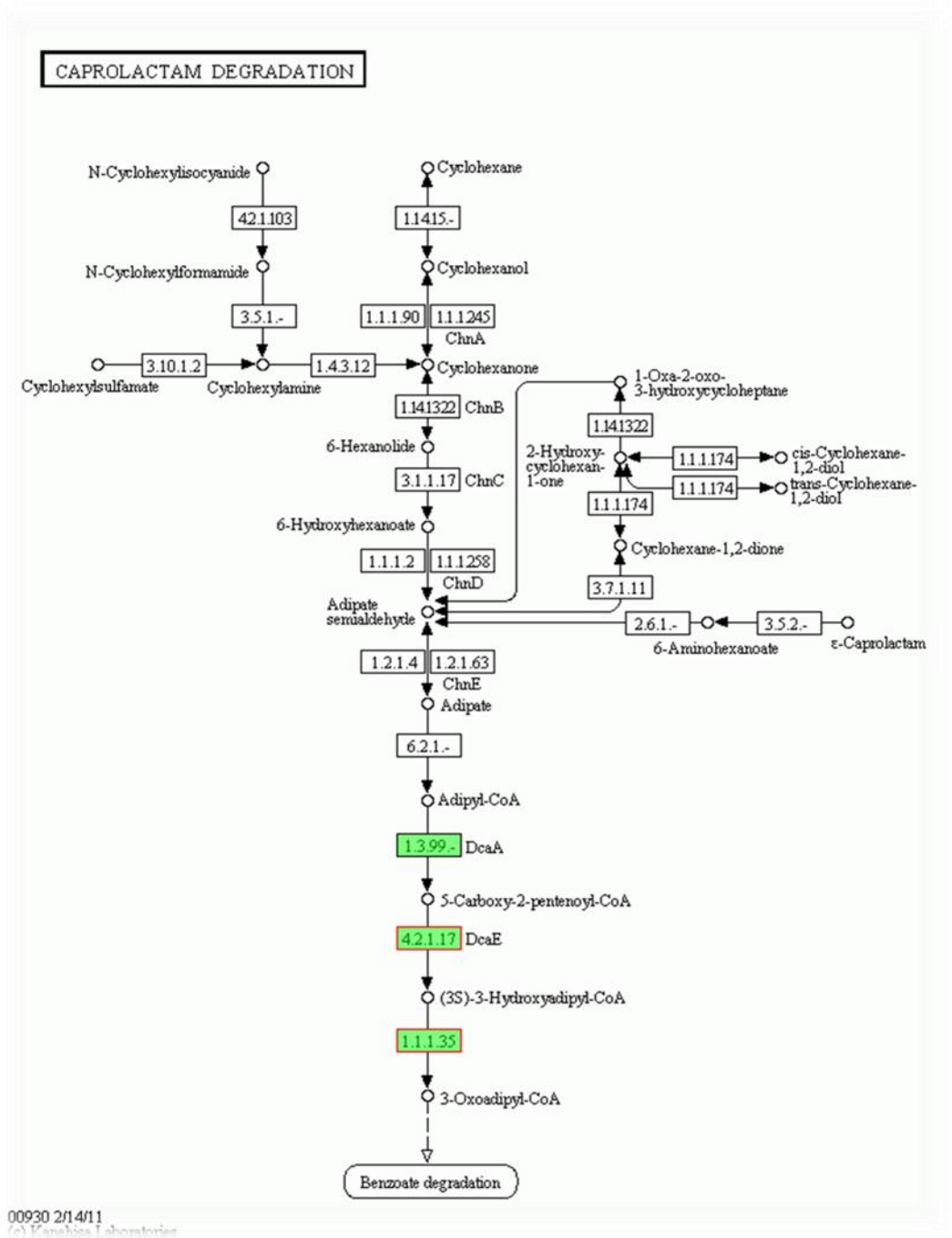


Figure 4.6.M: *Enterobacter* sp. RC4 genome caprolactam degradation pathway

4.10.j: Toluene utilization

In toluenedegradation pathways *Enterobacter* sp. RC4 utilize enzyme 1.14.13, 1.14.12.11, 1.1.1.35 (3-hydroxyacyl-CoA dehydrogenase, Toluene dioxygenase) from dmpK, todC1, bbsC genes involved in toluene friction degradation which are described in figure 4.6.N.

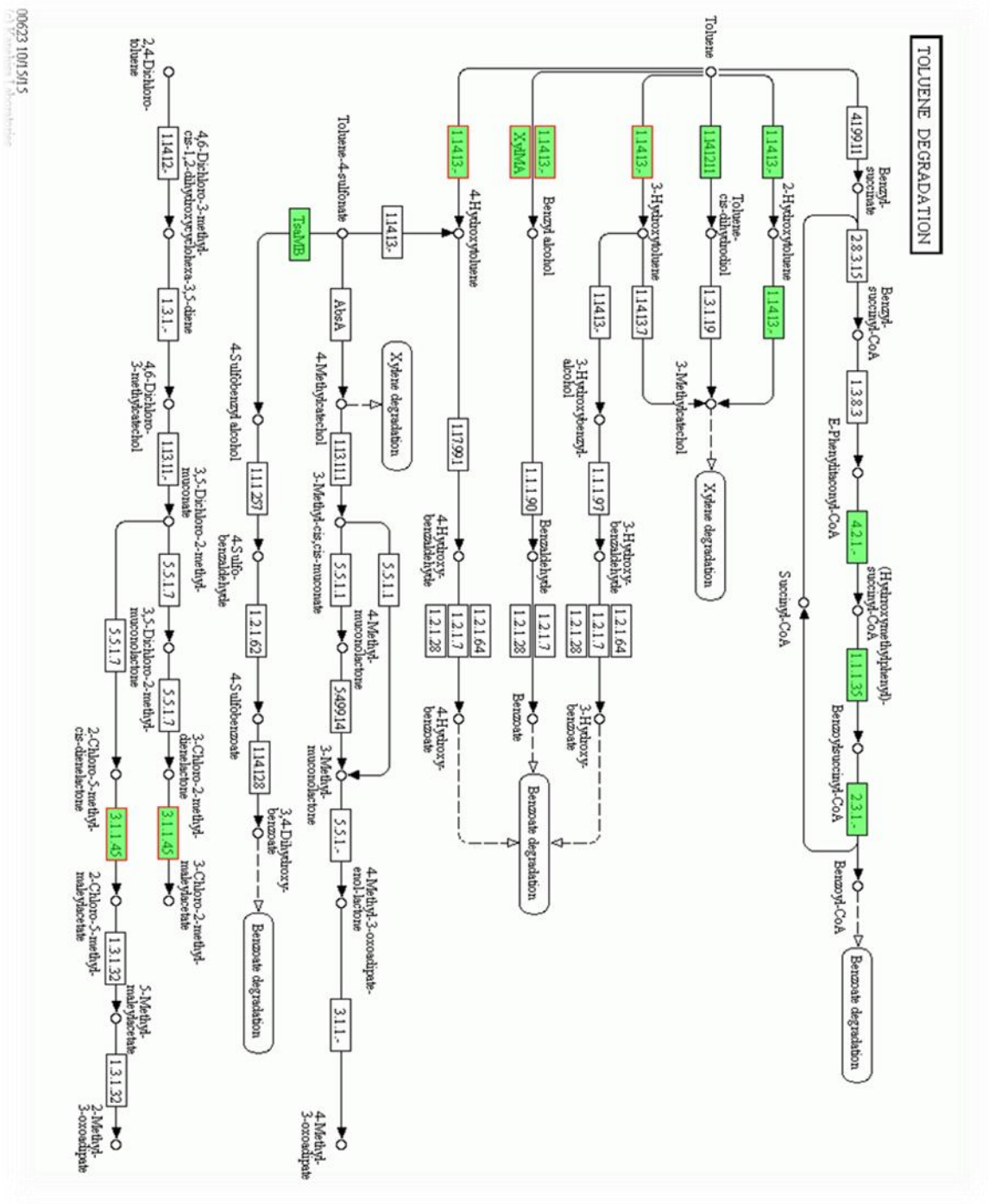


Figure 4.6.N: *Enterobacter* sp. RC4 genome toluene degradation pathway

4.11: Cell motility and chemotaxis genes

Alltogether 77 genes in *P. aeruginosa* and 102 genes in *Enterobacter* sp. RC4 was found for Cell motility and chemotaxis genes which help in servable of different environment condition which are showing in the (Table 4.9).

Table 4.9: *P. aeruginosa* N002 and *Enterobacter* sp. RC4 genomes cell motility and chemotaxis genes

***P. aeruginosa* N002**		***Enterobacter* sp. RC4**	
Locus Tag	**Gene Product Name**	**Locus Tag**	**Gene Product Name**
A222_04635	ABC-type dipeptide transport system, periplasmic component	C4L14_17305	CheA signal transduction histidine kinase
A222_04631	ABC-type dipeptide transport system, periplasmic component	C4L14_21875	CheA signal transduction histidine kinase
A222_03080	ABC-type sugar transport system, periplasmic component	C4L14_17240	chemotaxis protein CheZ
A222_03594	Chemotaxis protein	C4L14_21640	chemotaxis protein CheZ
A222_00183	Chemotaxis protein histidine kinase and related kinases (EC:2.7.13.3)	C4L14_21655	chemotaxis protein methyltransferase CheR
A222_03593	Chemotaxis protein histidine kinase and related kinases (EC:2.7.13.3)	C4L14_17255	chemotaxis protein methyltransferase CheR
A222_00179	Chemotaxis protein; stimulates methylation of MCP proteins (EC:3.5.1.44)	C4L14_21865	chemotaxis protein MotA
A222_00178	Chemotaxis response regulator containing a CheY-like receiver domain and a methylesterase domain (EC:3.1.1.61)	C4L14_17315	chemotaxis protein MotA
A222_03592	Chemotaxis response regulator containing a CheY-like receiver domain and a methylesterase domain (EC:3.1.1.61)	C4L14_21870	chemotaxis protein MotB

A222_03587	Chemotaxis signal transduction protein	C4L14_17310	chemotaxis protein MotB
A222_00182	Chemotaxis signal transduction protein	C4L14_23840	dipeptide transport system substrate-binding protein
A222_03588	Chemotaxis signal transduction protein	C4L14_24875	dipeptide transport system substrate-binding protein
A222_01614	Chemotaxis signal transduction protein	C4L14_21785	flagella basal body P-ring formation protein FlgA
A222_03964	fagellar hook-basal body proteins	C4L14_05565	flagella basal body P-ring formation protein FlgA
A222_01613	flagella basal body P-ring formation protein FlgA	C4L14_05555	flagella synthesis protein FlgN
A222_03961	Flagellar basal body L-ring protein	C4L14_21775	flagella synthesis protein FlgN
A222_03943	flagellar basal-body M-ring protein/flagellar hook-basal body protein (fliF)	C4L14_22060	flagellar assembly protein FliH
A222_03960	Flagellar basal-body P-ring protein	C4L14_21675	flagellar assembly protein FliH
A222_03967	flagellar basal-body rod protein FlgB	C4L14_21800	flagellar basal-body rod modification protein FlgD
A222_03966	flagellar basal-body rod protein FlgC	C4L14_05585	flagellar basal-body rod modification protein FlgD
A222_03963	flagellar basal-body rod protein FlgF	C4L14_21790	flagellar basal-body rod protein FlgB
A222_03962	flagellar basal-body rod protein FlgG, Gram-negative bacteria	C4L14_05575	flagellar basal-body rod protein FlgB
A222_01612	flagellar biosynthesis anti-sigma factor FlgM	C4L14_21795	flagellar basal-body rod protein FlgC
A222_03599	flagellar biosynthesis protein FlhA	C4L14_05580	flagellar basal-body rod protein FlgC
A222_01611	Flagellar biosynthesis/type III secretory pathway chaperone	C4L14_05595	flagellar basal-body rod protein FlgF
A222_03941	Flagellar biosynthesis/type III secretory pathway protein	C4L14_21810	flagellar basal-body rod protein FlgF
A222_03602	flagellar biosynthetic protein FlhB	C4L14_21815	flagellar basal-body rod protein FlgG
A222_03606	flagellar biosynthetic protein FliO	C4L14_05600	flagellar basal-body rod protein FlgG

A222_03605	flagellar biosynthetic protein FliP	C4L14_17230	flagellar biosynthesis protein FlhA
A222_03604	flagellar biosynthetic protein FliQ	C4L14_21630	flagellar biosynthesis protein FlhA
A222_03603	flagellar biosynthetic protein FliR	C4L14_21635	flagellar biosynthetic protein FlhB
A222_03949	flagellar biosynthetic protein FliS	C4L14_17235	flagellar biosynthetic protein FlhB
A222_03950	Flagellar capping protein	C4L14_21715	flagellar biosynthetic protein FliP
A222_03939	flagellar export protein FliJ	C4L14_22020	flagellar biosynthetic protein FliP
A222_03965	Flagellar hook capping protein	C4L14_21720	flagellar biosynthetic protein FliQ
A222_03957	flagellar hook-associated protein 3	C4L14_22015	flagellar biosynthetic protein FliQ
A222_03958	flagellar hook-associated protein FlgK	C4L14_21725	flagellar biosynthetic protein FliR
A222_03944	flagellar hook-basal body complex protein FliE	C4L14_22010	flagellar biosynthetic protein FliR
A222_03610	Flagellar hook-length control protein	C4L14_22050	flagellar FliJ protein
A222_03591	Flagellar motor component	C4L14_21685	flagellar FliJ protein
A222_05137	Flagellar motor protein	C4L14_05590	flagellar hook protein FlgE
A222_03590	Flagellar motor protein	C4L14_21805	flagellar hook protein FlgE
A222_05138	flagellar motor stator protein MotA	C4L14_21835	flagellar hook-associated protein 1 FlgK
A222_03942	flagellar motor switch protein FliG	C4L14_05620	flagellar hook-associated protein 1 FlgK
A222_03608	flagellar motor switch protein FliM	C4L14_21760	flagellar hook-associated protein 2
A222_03607	flagellar motor switch protein FliN	C4L14_17525	flagellar hook-associated protein 2
A222_03940	flagellar protein export ATPase FliI (EC:3.6.3.14)	C4L14_05625	flagellar hook-associated protein 3 FlgL
A222_03952	Flagellin and related hook-associated proteins	C4L14_21840	flagellar hook-associated protein 3 FlgL
A222_00185	Methyl-accepting chemotaxis protein	C4L14_21660	flagellar hook-basal body complex protein FliE
A222_04433	Methyl-accepting chemotaxis protein	C4L14_22075	flagellar hook-basal body complex protein FliE
A222_02441	Methyl-accepting chemotaxis protein	C4L14_21690	flagellar hook-length control protein FliK

A222_02473	Methyl-accepting chemotaxis protein	C4L14_22045	flagellar hook-length control protein FliK
A222_04416	Methyl-accepting chemotaxis protein	C4L14_05605	flagellar L-ring protein precursor FlgH
A222_03437	Methyl-accepting chemotaxis protein	C4L14_21820	flagellar L-ring protein precursor FlgH
A222_03787	Methyl-accepting chemotaxis protein	C4L14_21665	flagellar M-ring protein FliF
A222_05097	Methyl-accepting chemotaxis protein	C4L14_22070	flagellar M-ring protein FliF
A222_02055	Methyl-accepting chemotaxis protein	C4L14_21670	flagellar motor switch protein FliG
A222_02359	Methyl-accepting chemotaxis protein	C4L14_22065	flagellar motor switch protein FliG
A222_04436	Methyl-accepting chemotaxis protein	C4L14_21700	flagellar motor switch protein FliM
A222_04437	Methyl-accepting chemotaxis protein	C4L14_22035	flagellar motor switch protein FliM
A222_05260	Methyl-accepting chemotaxis protein	C4L14_21705	flagellar motor switch protein FliN/FliY
A222_03398	Methyl-accepting chemotaxis protein	C4L14_22030	flagellar motor switch protein FliN/FliY
A222_04655	Methyl-accepting chemotaxis protein	C4L14_21825	flagellar P-ring protein precursor FlgI
A222_00181	Methyl-accepting chemotaxis protein	C4L14_05610	flagellar P-ring protein precursor FlgI
A222_04790	Methyl-accepting chemotaxis protein	C4L14_22025	flagellar protein FliO/FliZ
A222_02357	Methyl-accepting chemotaxis protein	C4L14_21710	flagellar protein FliO/FliZ
A222_01615	Methylase of chemotaxis methyl-accepting proteins (EC:2.1.1.80)	C4L14_21765	flagellar protein FliS
A222_00180	Methylase of chemotaxis methyl-accepting proteins (EC:2.1.1.80)	C4L14_17530	flagellar protein FliS
A222_03091	PAS domain S-box	C4L14_17535	flagellar protein FliT
A222_03627	PAS domain S-box	C4L14_17320	flagellar transcriptional activator FlhC
A222_03487	PAS domain S-box	C4L14_21860	flagellar transcriptional activator FlhC

A222_03595	Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains	C4L14_17325	flagellar transcriptional activator FlhD
A222_00184	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	C4L14_21855	flagellar transcriptional activator FlhD
		C4L14_17520	Flagellin
		C4L14_21755	Flagellin
		C4L14_22055	flagellum-specific ATP synthase
		C4L14_21680	flagellum-specific ATP synthase
		C4L14_05560	FlgM family anti-sigma-28 factor
		C4L14_21780	FlgM family anti-sigma-28 factor
		C4L14_20415	glucose-binding protein /galactose-binding protein
		C4L14_08435	maltooligosaccharide-binding protein
		C4L14_21745	methyl-accepting chemotaxis protein
		C4L14_04405	methyl-accepting chemotaxis protein-2 (aspartate sensor receptor)
		C4L14_16450	methyl-accepting chemotaxis sensory transducer with Cache sensor
		C4L14_04410	methyl-accepting chemotaxis sensory transducer with Pas/Pac sensor
		C4L14_17260	methyl-accepting chemotaxis sensory transducer with TarH sensor
		C4L14_10055	methyl-accepting chemotaxis sensory transducer with TarH sensor
		C4L14_04975	methyl-accepting chemotaxis sensory transducer with TarH

		sensor
	C4L14_18700	methyl-accepting chemotaxis sensory transducer with TarH sensor
	C4L14_24060	methyl-accepting chemotaxis sensory transducer with TarH sensor
	C4L14_17265	methyl-accepting chemotaxis sensory transducer with TarH sensor
	C4L14_23705	methyl-accepting chemotaxis sensory transducer with TarH sensor
	C4L14_06800	methyl-accepting chemotaxis sensory transducer with TarH sensor
	C4L14_08145	monosaccharide ABC transporter substrate-binding protein (CUT2 family)
	C4L14_21880	purine-binding chemotaxis protein CheW
	C4L14_17300	purine-binding chemotaxis protein CheW
	C4L14_18380	ribose-binding protein
	C4L14_21650	two-component system chemotaxis response regulator CheB
	C4L14_17250	two-component system chemotaxis response regulator CheB
	C4L14_20830	two-component system chemotaxis response regulator CheV
	C4L14_21645	two-component system chemotaxis response regulator CheY
	C4L14_17245	two-component system chemotaxis response regulator CheY

The details of *P. aeruginosa* N002 and *Enterobacter* sp. RC4 bacterial chemotaxis pathway was showing in (Fgure 4.7.A and 4.7.B) where green colour showing the genes

which are present in to our study bacteria.

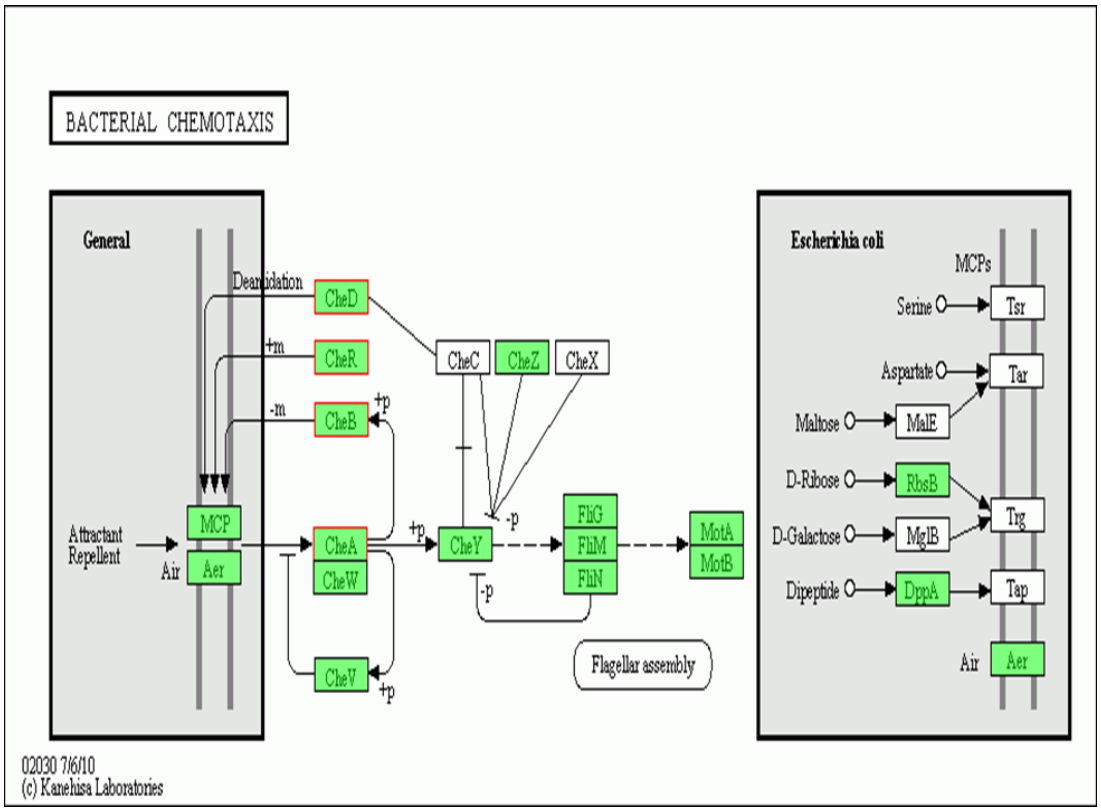


Figure 4.7.A:*P. aeruginosa* N002 genomes bacterial chemotaxis pathway

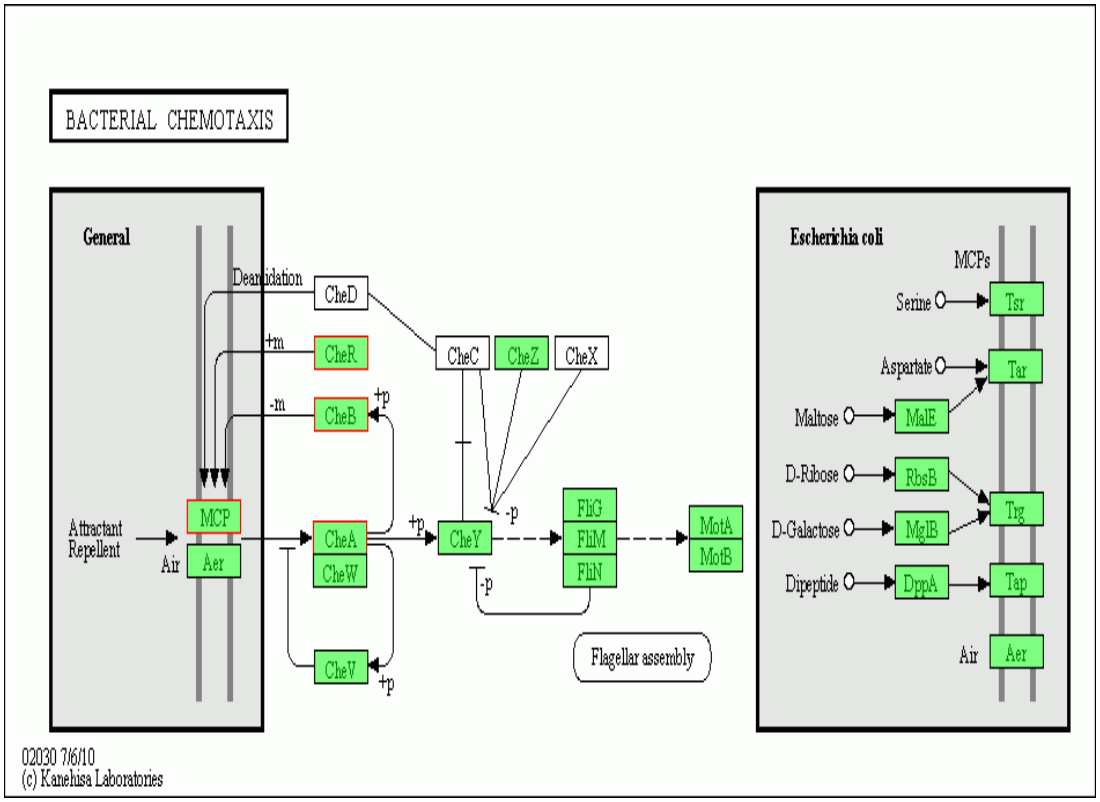


Figure 4.7.B: *Enterobacter* sp. RC4genomes bacterial chemotaxis pathway

4.12. Signal transduction genes

A total of 344 Signal transduction genes from *P. aeruginosa*N002 and 240 Signal transduction genes from *Enterobacter* sp. RC4 genomes are founed to involved in diffrent function of Signal transduction mechanism which are showing intable 4.10.

Table 4.10: Signal transduction genes of *P. aeruginosa* N002 and *Enterobacter* sp. RC4 genome

***P. aeruginosa* N002**		***Enterobacter* sp. RC4**	
Locus Tag	**Gene Product Name**	**Locus Tag**	**Gene Product Name**
A222_04113	(p)ppGpp synthetase, RelA/SpoT family (EC:2.7.6.5)	C4L14_00100	protein-tyrosine phosphatase
A222_05534	(p)ppGpp synthetase, RelA/SpoT family (EC:3.1.7.2)	C4L14_01290	LuxR family two component transcriptional regulator
A222_01107	ABC-type amino acid transport/signal transduction systems, periplasmic component/domain	C4L14_01295	two-component system sensor histidine kinase UhpB
A222_00462	ABC-type amino acid transport/signal transduction systems, periplasmic component/domain	C4L14_01355	transcriptional regulator with PAS, ATPase and Fis domain
A222_01006	ABC-type amino acid transport/signal transduction systems, periplasmic component/domain	C4L14_01445	EAL domain-containing protein (putative c-di-GMP-specific phosphodiesterase class I)
A222_00915	ABC-type amino acid transport/signal transduction systems, periplasmic component/domain	C4L14_01450	DNA-binding NarL/FixJ family response regulator
A222_01710	ABC-type amino acid transport/signal transduction systems, periplasmic component/domain	C4L14_01455	two-component system capsular synthesis response regulator RcsB
A222_03518	ABC-type amino acid transport/signal transduction systems, periplasmic	C4L14_01605	two-component system response regulator FimZ (fimbrial Z protein)

	component/domain		
A222_05327	ABC-type amino acid transport/signal transduction systems, periplasmic component/domain	C4L14_01720	diguanylate cyclase/phosphodiesterase
A222_04280	ABC-type amino acid transport/signal transduction systems, periplasmic component/domain	C4L14_01740	EAL domain-containing protein (putative c-di-GMP-specific phosphodiesterase class I)
A222_05326	ABC-type amino acid transport/signal transduction systems, periplasmic component/domain	C4L14_01770	DNA-binding NarL/FixJ family response regulator
A222_05344	ABC-type amino acid transport/signal transduction systems, periplasmic component/domain	C4L14_01815	universal stress protein A
A222_05325	ABC-type amino acid transport/signal transduction systems, periplasmic component/domain	C4L14_01890	two-component system heavy metal sensor histidine kinase CusS
A222_05466	ABC-type amino acid transport/signal transduction systems, periplasmic component/domain	C4L14_01895	two-component system copper resistance phosphate regulon response regulator CusR
A222_00320	ABC-type amino acid transport/signal transduction systems, periplasmic component/domain	C4L14_01965	two-component system copper resistance phosphate regulon response regulator CusR
A222_05167	ABC-type amino acid transport/signal transduction systems, periplasmic component/domain	C4L14_01970	two-component system heavy metal sensor histidine kinase CusS
A222_02052	ABC-type amino acid transport/signal transduction systems, periplasmic component/domain	C4L14_02675	GTP pyrophosphokinase
A222_05677	ABC-type amino acid transport/signal transduction systems, periplasmic	C4L14_02685	Hpt sensor hybrid histidine kinase

	component/domain		
A222_05264	ABC-type amino acid transport/signal transduction systems, periplasmic component/domain	C4L14_02875	transcriptional regulator
A222_02829	ABC-type amino acid transport/signal transduction systems, periplasmic component/domain	C4L14_03175	two-component system response regulator AdeR
A222_03778	ABC-type amino acid transport/signal transduction systems, periplasmic component/domain	C4L14_03180	two-component system sensor histidine kinase AdeS
A222_02075	ABC-type amino acid transport/signal transduction systems, periplasmic component/domain	C4L14_03240	psp operon transcriptional activator
A222_00937	ABC-type amino acid transport/signal transduction systems, periplasmic component/domain	C4L14_03245	phage shock protein A (PspA) family protein
A222_05270	ABC-type amino acid transport/signal transduction systems, periplasmic component/domain	C4L14_03255	phage shock protein C (PspC) family protein
A222_03441	ABC-type amino acid transport/signal transduction systems, periplasmic component/domain	C4L14_03490	PAS domain S-box-containing protein/diguanylate cyclase (GGDEF)-like protein
A222_03693	ABC-type amino acid transport/signal transduction systems, periplasmic component/domain	C4L14_03495	diguanylate cyclase (GGDEF)-like protein
A222_01483	ABC-type amino acid transport/signal transduction systems, periplasmic component/domain (EC:4.2.1.91, EC:4.2.1.51)	C4L14_03500	methyl-accepting chemotaxis protein
A222_02431	Activator of osmoprotectant	C4L14_03805	two-component system response regulator QseB

	transporter ProP		
A222_01756	Adenylate cyclase, family 3 (some proteins contain HAMP domain) (EC:4.6.1.-)	C4L14_03810	two-component system sensor histidine kinase QseC
A222_01616	Anti-anti-sigma regulatory factor (antagonist of anti-sigma factor)	C4L14_03890	Icc protein
A222_04586	Anti-anti-sigma regulatory factor (antagonist of anti-sigma factor)	C4L14_04045	LuxR family two component transcriptional regulator
A222_02192	Anti-anti-sigma regulatory factor (antagonist of anti-sigma factor)	C4L14_04340	two-component system sensor histidine kinase EvgS
A222_00850	Bacteriophytochrome (light-regulated signal transduction histidine kinase)	C4L14_04345	LuxR family two component transcriptional regulator
A222_00602	bis(5'-nucleosyl)-tetraphosphatase (symmetrical) (EC:3.6.1.41)	C4L14_04405	methyl-accepting chemotaxis protein-2 (aspartate sensor receptor)
A222_00661	cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases	C4L14_04410	methyl-accepting chemotaxis sensory transducer with Pas/Pac sensor
A222_00535	cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases	C4L14_04880	two-component system cit operon sensor histidine kinase CitA
A222_00278	cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases	C4L14_04885	two-component system response regulator CitB
A222_02171	cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases	C4L14_04915	signal transduction histidine kinase

A222_03505	cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases	C4L14_04920	LuxR family two component transcriptional regulator
A222_04761	Carbon starvation protein, predicted membrane protein	C4L14_04975	methyl-accepting chemotaxis sensory transducer with TarH sensor
A222_05376	Carbon storage regulator (could also regulate swarming and quorum sensing)	C4L14_05050	amino acid ABC transporter substrate-binding protein (PAAT family)
A222_04146	carbon storage regulator (csrA)	C4L14_05825	EAL and modified HD-GYP domain-containing signal transduction protein
A222_03173	CBS-domain-containing membrane protein	C4L14_05865	two-component system sensor histidine kinase PhoQ
A222_03594	Chemotaxis protein	C4L14_05870	two-component system response regulator PhoP
A222_01269	Chemotaxis protein histidine kinase and related kinases	C4L14_05880	high frequency lysogenization protein
A222_00423	Chemotaxis protein histidine kinase and related kinases	C4L14_05905	LuxR family two component transcriptional regulator
A222_00183	Chemotaxis protein histidine kinase and related kinases (EC:2.7.13.3)	C4L14_05925	sensor c-di-GMP phosphodiesterase-like protein
A222_03593	Chemotaxis protein histidine kinase and related kinases (EC:2.7.13.3)	C4L14_05950	diguanylate cyclase
A222_00179	Chemotaxis protein; stimulates methylation of MCP proteins (EC:3.5.1.44)	C4L14_06000	putative serine protein kinase PrkA
A222_00424	Chemotaxis response regulator containing a CheY-like receiver domain and a methylesterase domain	C4L14_06090	RseA-like anti sigma(E) protein
A222_01270	Chemotaxis response regulator containing a CheY-like receiver domain and a methylesterase domain (EC:3.1.1.61)	C4L14_06095	MucB/RseB-like sigma(E) regulatory protein

A222_00178	Chemotaxis response regulator containing a CheY-like receiver domain and a methylesterase domain (EC:3.1.1.61)	C4L14_06100	RseC/MucC-like positive regulator of sigma(E)
A222_03592	Chemotaxis response regulator containing a CheY-like receiver domain and a methylesterase domain (EC:3.1.1.61)	C4L14_06175	membrane-bound lytic murein transglycosylase F
A222_03587	Chemotaxis signal transduction protein	C4L14_06190	two-component system sensor histidine kinase GlrK
A222_00182	Chemotaxis signal transduction protein	C4L14_06200	two-component system response regulator GlrR
A222_00425	Chemotaxis signal transduction protein	C4L14_06205	nitrogen regulatory protein P-II family
A222_01266	Chemotaxis signal transduction protein	C4L14_06390	predicted negative regulator of RcsB-dependent stress response
A222_00420	Chemotaxis signal transduction protein	C4L14_06555	LuxR family two component transcriptional regulator
A222_01268	Chemotaxis signal transduction protein	C4L14_06735	carbon starvation protein CstA
A222_03588	Chemotaxis signal transduction protein	C4L14_06800	methyl-accepting chemotaxis sensory transducer with TarH sensor
A222_01614	Chemotaxis signal transduction protein	C4L14_06855	DNA-binding NarL/FixJ family response regulator
A222_04871	Cyclic nucleotide-binding domain.	C4L14_06860	DNA-binding NarL/FixJ family response regulator
A222_04460	Diguanylate cyclase (GGDEF) domain	C4L14_06880	diguanylate cyclase (GGDEF)-like protein
A222_01620	Diguanylate cyclase (GGDEF) domain	C4L14_07055	CreA protein
A222_03937	Diguanylate cyclase (GGDEF) domain	C4L14_07060	two-component system aerobic respiration control protein ArcA
A222_03176	Diguanylate cyclase (GGDEF) domain	C4L14_07230	bis(5'nucleosyl)-tetraphosphatase ApaH
A222_03922	Diguanylate cyclase (GGDEF) domain	C4L14_07610	LuxR family two component transcriptional regulator
A222_02107	Diguanylate cyclase (GGDEF) domain	C4L14_07740	sugar fermentation stimulation protein
A222_05023	Diguanylate cyclase (GGDEF) domain	C4L14_07830	CdaR family transcriptional regulator
A222_05492	Diguanylate cyclase (GGDEF) domain	C4L14_07845	UTP--GlnB (protein PII) uridylyltransferase GlnD

A222_01714	Diguanylate cyclase (GGDEF) domain	C4L14_08170	DNA-binding response OmpR family regulator
A222_05111	Diguanylate cyclase (GGDEF) domain	C4L14_08185	hypothetical protein
A222_03618	Diguanylate cyclase (GGDEF) domain	C4L14_08260	sensor c-di-GMP phosphodiesterase-like protein
A222_05692	Diguanylate cyclase (GGDEF) domain	C4L14_08275	FecR family protein
A222_03310	Diguanylate cyclase (GGDEF) domain	C4L14_08395	SOS-response transcriptional repressor LexA
A222_04496	Diguanylate cyclase (GGDEF) domain	C4L14_08650	LuxR family two component transcriptional regulator
A222_01652	Diguanylate cyclase (GGDEF) domain	C4L14_08660	fimbrial protein FimY
A222_00174	Diguanylate cyclase (GGDEF) domain	C4L14_08665	EAL domain-containing protein (putative c-di-GMP-specific phosphodiesterase class I)
A222_01797	Diguanylate cyclase (GGDEF) domain	C4L14_09020	sensor c-di-GMP phosphodiesterase-like protein
A222_01271	Diguanylate cyclase (GGDEF) domain (EC:2.7.7.65)	C4L14_09090	nitrogen regulatory protein P-II family
A222_04730	DnaK suppressor protein	C4L14_09185	BolA protein family transcriptional regulator
A222_05077	Fe2+-dicitrate sensor, membrane component	C4L14_09425	2-octaprenylphenol hydroxylase
A222_03671	Fe2+-dicitrate sensor, membrane component	C4L14_09680	exopolyphosphatase/guanosine -5'-triphosphate,3'-diphosphate pyrophosphatase
A222_03111	Fe2+-dicitrate sensor, membrane component	C4L14_09780	two-component system OmpR family response regulator
A222_01066	Fe2+-dicitrate sensor, membrane component	C4L14_09785	signal transduction histidine kinase
A222_02972	Fe2+-dicitrate sensor, membrane component	C4L14_09870	two-component system copper resistance phosphate regulon response regulator CusR
A222_02930	Fe2+-dicitrate sensor, membrane component	C4L14_09875	two-component system heavy metal sensor histidine kinase CusS
A222_04393	Fe2+-dicitrate sensor, membrane component	C4L14_09950	serine/threonine-protein kinase RsbW
A222_00482	Fe2+-dicitrate sensor, membrane component	C4L14_09955	anti-sigma B factor antagonist
A222_03735	Fe2+-dicitrate sensor, membrane component	C4L14_09960	sigma-B regulation protein RsbU (phosphoserine phosphatase)
A222_01550	Fe2+-dicitrate sensor, membrane component	C4L14_10000	PAS domain S-box-containing protein

A222_00155	Fe2+-dicitrate sensor, membrane component	C4L14_10055	methyl-accepting chemotaxis sensory transducer with TarH sensor
A222_02652	Fe2+-dicitrate sensor, membrane component	C4L14_10085	EAL domain-containing protein (putative c-di-GMP-specific phosphodiesterase class I)
A222_02572	Fe2+-dicitrate sensor, membrane component	C4L14_10155	two-component system response regulator CpxR
A222_01018	FOG: EAL domain	C4L14_10160	two-component system sensor histidine kinase CpxA
A222_01143	FOG: EAL domain	C4L14_10380	GTP-binding protein TypA/BipA
A222_02466	FOG: EAL domain	C4L14_10400	PAS/PAC sensor signal transduction histidine kinase
A222_02897	FOG: EAL domain	C4L14_10405	two-component system nitrogen regulation response regulator GlnG
A222_02205	FOG: GAF domain	C4L14_10445	Ser/Thr protein kinase RdoA (MazF antagonist)
A222_01618	FOG: HPt domain	C4L14_10485	exopolyphosphatase/guanosine -5'-triphosphate,3'-diphosphate pyrophosphatase
A222_00034	FOG: HPt domain	C4L14_10630	two-component system nitrate/nitrite sensor histidine kinase NarQ
A222_02156	GAF domain-containing protein	C4L14_10740	phosphocarrier protein HPr
A222_05200	Glutamine synthetase adenylyltransferase (EC:2.7.7.42)	C4L14_10895	two-component system phosphate regulon response regulator OmpR
A222_05305	GTP-binding protein TypA/BipA	C4L14_10900	two-component system osmolarity sensor histidine kinase EnvZ
A222_05139	HDOD domain.	C4L14_11060	cell filamentation protein
A222_02515	Heavy metal response regulator	C4L14_11080	CRP/FNR family cyclic AMP-dependent transcriptional regulator
A222_02180	Heavy metal response regulator	C4L14_11350	DNA-binding NarL/FixJ family response regulator
A222_03614	Heavy metal response regulator	C4L14_11465	RNA-binding protein Hfq
A222_05067	Heavy metal response regulator	C4L14_11520	methyl-accepting chemotaxis protein
A222_02179	Heavy metal sensor kinase (EC:2.7.13.3)	C4L14_11560	PTS system ascorbate-specific IIA component (L-Asc family)
A222_05068	Heavy metal sensor kinase (EC:2.7.13.3)	C4L14_12115	carbon starvation protein CstA
A222_03613	Heavy metal sensor	C4L14_12230	universal stress protein G

	kinase (EC:2.7.13.3)		
A222_02514	Heavy metal sensor kinase (EC:2.7.13.3)	C4L14_12305	hypothetical protein
A222_02443	His Kinase A (phosphoacceptor) domain./Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase.	C4L14_12385	amino acid ABC transporter substrate-binding protein (PAAT family) /L-glutamate-binding protein /L-aspartate-binding protein
A222_04702	His Kinase A (phosphoacceptor) domain./Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase. (EC:2.7.13.3)	C4L14_12405	phosphate starvation-inducible protein PhoH
A222_02095	His Kinase A (phosphoacceptor) domain./Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase./CBS domain.	C4L14_12555	two-component system KDP operon response regulator KdpE
A222_03046	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase./Histidine kinase.	C4L14_12560	two-component system sensor histidine kinase KdpD
A222_03127	Hypothetical protein	C4L14_13105	L-glutamine-binding protein
A222_02892	Hypothetical protein	C4L14_13410	DNA-binding response OmpR family regulator
A222_00441	Hypothetical protein	C4L14_13420	amino acid ABC transporter substrate-binding protein (PAAT family) /L-arginine-binding protein
A222_03661	Isocitrate dehydrogenase kinase/phosphatase (EC:3.1.3.-, EC:2.7.11.5)	C4L14_13435	amino acid ABC transporter substrate-binding protein (PAAT family) /L-arginine-binding protein
A222_05194	Lipopolysaccharide kinase (Kdo/WaaP) family.	C4L14_13630	universal stress protein G
A222_04163	Lysine-arginine-ornithine-binding periplasmic protein	C4L14_13825	hypothetical protein
A222_01100	Lysine-arginine-ornithine-binding periplasmic protein	C4L14_14185	diguanylate cyclase (GGDEF)-like protein
A222_00185	Methyl-accepting chemotaxis protein	C4L14_14560	diguanylate cyclase (GGDEF)-like protein
A222_04433	Methyl-accepting	C4L14_14630	CRP-like cAMP-binding

	chemotaxis protein		protein
A222_02110	Methyl-accepting chemotaxis protein	C4L14_14690	diguanylate cyclase
A222_02203	Methyl-accepting chemotaxis protein	C4L14_14745	two-component system phosphate regulon response regulator PhoB
A222_02441	Methyl-accepting chemotaxis protein	C4L14_14750	PAS/PAC sensor signal transduction histidine kinase
A222_02473	Methyl-accepting chemotaxis protein	C4L14_14810	formate hydrogenlyase transcriptional activator
A222_04416	Methyl-accepting chemotaxis protein	C4L14_14930	anaerobic nitric oxide reductase transcription regulator
A222_03437	Methyl-accepting chemotaxis protein	C4L14_15010	carbon storage regulator CsrA
A222_03787	Methyl-accepting chemotaxis protein	C4L14_15055	S-ribosylhomocysteine lyase /quorum-sensing autoinducer 2 (AI-2) synthesis protein LuxS
A222_05097	Methyl-accepting chemotaxis protein	C4L14_15240	amino acid ABC transporter substrate-binding protein (PAAT family)
A222_01265	Methyl-accepting chemotaxis protein	C4L14_15260	amino acid ABC transporter substrate-binding protein (PAAT family)
A222_02055	Methyl-accepting chemotaxis protein	C4L14_15460	phosphate starvation-inducible protein PhoH
A222_02359	Methyl-accepting chemotaxis protein	C4L14_15510	LuxR family transcriptional regulator of csgAB operon
A222_04436	Methyl-accepting chemotaxis protein	C4L14_16110	L-histidine-binding protein
A222_04437	Methyl-accepting chemotaxis protein	C4L14_16115	L-arginine-binding protein /L-lysine-binding protein /L-ornithine-binding protein
A222_05260	Methyl-accepting chemotaxis protein	C4L14_16230	phosphohistidine phosphatase SixA
A222_03398	Methyl-accepting chemotaxis protein	C4L14_16395	two-component system nitrate/nitrite sensor histidine kinase NarX
A222_05024	Methyl-accepting chemotaxis protein	C4L14_16400	LuxR family two component transcriptional regulator
A222_04655	Methyl-accepting chemotaxis protein	C4L14_16450	methyl-accepting chemotaxis sensory transducer with Cache sensor
A222_00181	Methyl-accepting chemotaxis protein	C4L14_16480	regulator of sirC expression with transglutaminase-like and TPR domain
A222_00421	Methyl-accepting chemotaxis protein	C4L14_16590	ankyrin repeat protein
A222_04790	Methyl-accepting	C4L14_16600	serine/threonine protein kinase

	chemotaxis protein		
A222_02357	Methyl-accepting chemotaxis protein	C4L14_16635	serine/threonine protein phosphatase PrpC
A222_04434	Methyl-accepting chemotaxis protein	C4L14_16640	FHA domain protein
A222_01267	Methylase of chemotaxis methyl-accepting proteins	C4L14_16895	EAL domain-containing protein (putative c-di-GMP-specific phosphodiesterase class I)
A222_00422	Methylase of chemotaxis methyl-accepting proteins	C4L14_17005	ProP effector
A222_01615	Methylase of chemotaxis methyl-accepting proteins (EC:2.1.1.80)	C4L14_17010	GAF domain-containing protein
A222_00180	Methylase of chemotaxis methyl-accepting proteins (EC:2.1.1.80)	C4L14_17240	chemotaxis protein CheZ
A222_05193	Mn2+-dependent serine/threonine protein kinase	C4L14_17250	two-component system chemotaxis response regulator CheB
A222_01482	N-acyl-L-homoserine lactone synthetase (EC:2.3.1.184)	C4L14_17255	chemotaxis protein methyltransferase CheR
A222_03619	N-acyl-L-homoserine lactone synthetase (EC:2.3.1.184)	C4L14_17260	methyl-accepting chemotaxis sensory transducer with TarH sensor
A222_04294	Negative regulator of sigma E activity	C4L14_17265	methyl-accepting chemotaxis sensory transducer with TarH sensor
A222_04293	Negative regulator of sigma E activity	C4L14_17300	purine-binding chemotaxis protein CheW
A222_05313	nitrogen regulation protein NR(I)	C4L14_17305	CheA signal transduction histidine kinase
A222_03409	Osmosensitive K+ channel histidine kinase (EC:2.7.13.3)	C4L14_17330	universal stress protein C
A222_01768	P pilus assembly/Cpx signaling pathway, periplasmic inhibitor/zinc-resistance associated protein	C4L14_17465	LuxR family two component transcriptional regulator
A222_03091	PAS domain S-box	C4L14_17495	L-cystine-binding protein /diaminopimelate-binding protein
A222_03627	PAS domain S-box	C4L14_17670	hypothetical protein

A222_00543	PAS domain S-box	C4L14_17825	diguanylate cyclase (GGDEF)-like protein
A222_03487	PAS domain S-box	C4L14_17950	universal stress protein F
A222_03033	PAS domain S-box	C4L14_18110	isocitrate dehydrogenase kinase/phosphatase
A222_02812	PAS domain S-box	C4L14_18175	LytTR family two component transcriptional regulator
A222_02849	PAS domain S-box	C4L14_18180	two-component system LytT family sensor kinase
A222_04530	PAS domain S-box	C4L14_18200	diguanylate cyclase
A222_00774	PAS domain S-box	C4L14_18435	methyl-accepting chemotaxis protein
A222_00612	PAS domain S-box	C4L14_18455	EAL and modified HD-GYP domain-containing signal transduction protein
A222_03796	PAS domain S-box	C4L14_18495	hypothetical protein
A222_03050	PAS domain S-box	C4L14_18645	DNA-binding response OmpR family regulator
A222_04419	PAS domain S-box	C4L14_18650	signal transduction histidine kinase
A222_00855	PAS domain S-box	C4L14_18700	methyl-accepting chemotaxis sensory transducer with TarH sensor
A222_05689	PAS domain S-box (EC:2.7.13.3)	C4L14_18735	transcriptional regulator
A222_01019	PAS domain S-box (EC:2.7.13.3)	C4L14_18745	universal stress protein E
A222_00346	PAS domain S-box/diguanylate cyclase (GGDEF) domain	C4L14_18795	two-component system response regulator RstA
A222_04206	PAS domain S-box/diguanylate cyclase (GGDEF) domain	C4L14_18805	two-component system sensor histidine kinase RstB
A222_05143	PAS domain S-box/diguanylate cyclase (GGDEF) domain	C4L14_19110	LuxR family two component transcriptional regulator
A222_04192	PAS domain S-box/diguanylate cyclase (GGDEF) domain	C4L14_19205	EAL domain-containing protein (putative c-di-GMP-specific phosphodiesterase class I)
A222_00288	PAS domain S-box/diguanylate cyclase (GGDEF) domain	C4L14_19235	DNA-binding NarL/FixJ family response regulator
A222_02951	PAS domain S-box/diguanylate	C4L14_19265	CRP-like cAMP-binding protein

	cyclase (GGDEF) domain		
A222_05203	PAS domain S-box/diguanylate cyclase (GGDEF) domain	C4L14_19270	CRP-like cAMP-binding protein
A222_05647	PAS domain S-box/diguanylate cyclase (GGDEF) domain	C4L14_19315	hypothetical protein
A222_03863	PAS domain S-box/diguanylate cyclase (GGDEF) domain	C4L14_19390	EAL domain-containing protein (putative c-di-GMP-specific phosphodiesterase class I)
A222_00587	PAS domain S-box/diguanylate cyclase (GGDEF) domain	C4L14_19570	hypothetical protein
A222_04756	PAS domain S-box/diguanylate cyclase (GGDEF) domain	C4L14_20015	EAL domain-containing protein (putative c-di-GMP-specific phosphodiesterase class I)
A222_00293	PAS domain S-box/diguanylate cyclase (GGDEF) domain	C4L14_20045	EAL domain-containing protein (putative c-di-GMP-specific phosphodiesterase class I)
A222_01243	Phage shock protein A (IM30), suppresses sigma54-dependent transcription	C4L14_20110	two-component system sensor histidine kinase BaeS
A222_00625	Phage/conjugal plasmid C-4 type zinc finger protein, TraR family	C4L14_20115	two-component system response regulator BaeR
A222_05050	Phage/conjugal plasmid C-4 type zinc finger protein, TraR family	C4L14_20245	protein-tyrosine phosphatase
A222_05557	Phosphate regulon sensor kinase PhoR (EC:2.7.13.3)	C4L14_20280	LytTR family two component transcriptional regulator
A222_05556	Phosphate regulon transcriptional regulatory protein PhoB	C4L14_20285	two-component system LytT family sensor kinase
A222_00981	Phosphate starvation-inducible protein PhoH, predicted ATPase	C4L14_20395	serine/threonine-protein kinase HipA
A222_00345	Phosphoenolpyruvate-protein phosphotransferase (EC:2.7.3.9)	C4L14_20535	EAL domain-containing protein (putative c-di-GMP-specific phosphodiesterase class I)

A222_03429	Phosphohistidine phosphatase SixA (EC:3.1.3.-)	C4L14_20670	two-component system sensor histidine kinase RcsD
A222_04292	Positive regulator of sigma E activity	C4L14_20675	LuxR family two component transcriptional regulator
A222_01046	Predicted ATPase related to phosphate starvation-inducible protein PhoH	C4L14_20680	two-component system capsular synthesis sensor histidine kinase RcsC
A222_02498	Predicted protein-tyrosine phosphatise	C4L14_20760	two-component system sensor histidine kinase TctE
A222_00270	Predicted signal transduction protein	C4L14_20765	transcriptional regulator
A222_05542	Predicted signal transduction protein	C4L14_20830	two-component system chemotaxis response regulator CheV
A222_04300	Predicted signal transduction protein	C4L14_20965	PTS system IIA component (L-Asc family)
A222_00364	Predicted signal transduction protein	C4L14_21110	EAL domain-containing protein (putative c-di-GMP-specific phosphodiesterase class I)
A222_02833	Predicted signal transduction protein containing sensor and EAL domains	C4L14_21185	PAS domain S-box-containing protein/diguanylate cyclase (GGDEF)-like protein
A222_05459	Predicted signal transduction protein with a C-terminal ATPase domain (EC:2.7.13.3)	C4L14_21215	protein-tyrosine phosphatase
A222_01740	Predicted signal-transduction protein containing cAMP-binding and CBS domains	C4L14_21570	antitoxin ChpS
A222_00248	Predicted signal-transduction protein containing cAMP-binding and CBS domains	C4L14_21615	hypothetical protein
A222_00928	Predicted transmembrane sensor domain	C4L14_21640	chemotaxis protein CheZ
A222_01084	Protein tyrosine/serine phosphatise	C4L14_21650	two-component system chemotaxis response regulator CheB
A222_02754	Protein-tyrosine-phosphatase (EC:1.20.4.1)	C4L14_21655	chemotaxis protein methyltransferase CheR

A222_01989	Protein-tyrosine-phosphatase (EC:3.1.3.48)	C4L14_21845	methyl-accepting chemotaxis sensory transducer with TarH sensor
A222_04599	PTS IIA-like nitrogen-regulatory protein PtsN (EC:2.7.1.69)	C4L14_21875	CheA signal transduction histidine kinase
A222_00600	Putative Ser protein kinase	C4L14_21880	purine-binding chemotaxis protein CheW
A222_03107	Regulator of polyketide synthase expression	C4L14_21930	two-component system response regulator DcuR
A222_00775	Response regulator	C4L14_21935	two-component system sensor histidine kinase DcuS
A222_04628	Response regulator consisting of a CheY-like receiver domain and a Fis-type HTH domain	C4L14_21985	diguanylate cyclase (GGDEF)-like protein
A222_04528	Response regulator containing a CheY-like receiver domain and a GGDEF domain	C4L14_22005	LuxR family capsular biosynthesis transcriptional activator
A222_04957	Response regulator containing a CheY-like receiver domain and an HD-GYP domain	C4L14_22375	amino acid ABC transporter substrate-binding protein (PAAT family)
A222_02442	Response regulator containing a CheY-like receiver domain and an HD-GYP domain	C4L14_22405	diguanylate cyclase/phosphodiesterase
A222_00035	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	C4L14_22475	diguanylate cyclase/phosphodiesterase
A222_01920	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	C4L14_22705	PAS/PAC sensor hybrid histidine kinase
A222_01017	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	C4L14_22720	HPr-like nitrogen-regulatory protein NPr
A222_00613	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	C4L14_22725	UPF0042 nucleotide-binding protein

A222_03653	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	C4L14_22730	phosphotransferase IIA-like nitrogen-regulatory protein PtsN
A222_02078	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	C4L14_22800	acid stress-induced BolA-like protein IbaG/YrbA
A222_01259	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	C4L14_22840	two-component system sensor histidine kinase BasS
A222_02664	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	C4L14_22845	two-component system response regulator BasR
A222_00883	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	C4L14_23060	PTS system galactitol-specific EIIA component (Gat family)
A222_02426	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	C4L14_23320	two-component system response regulator PhoP
A222_01371	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	C4L14_23365	two-component system response regulator QseB
A222_01090	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	C4L14_23370	signal transduction histidine kinase
A222_03048	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	C4L14_23665	EAL domain-containing protein (putative c-di-GMP-specific phosphodiesterase class I)
A222_00889	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	C4L14_23705	methyl-accepting chemotaxis sensory transducer with TarH sensor

A222_03045	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	C4L14_23810	amino acid ABC transporter substrate-binding protein (PAAT family)
A222_03688	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	C4L14_24060	methyl-accepting chemotaxis sensory transducer with TarH sensor
A222_00870	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	C4L14_24075	diguanylate cyclase (GGDEF)-like protein
A222_02421	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	C4L14_24080	diguanylate cyclase (GGDEF)-like protein
A222_03568	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	C4L14_24085	diguanylate cyclase (GGDEF)-like protein
A222_04422	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	C4L14_24130	nucleotide-binding universal stress UspA family protein
A222_01199	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	C4L14_24240	guanosine-3',5'-bis(diphosphate) 3'-pyrophosphohydrolase
A222_03700	Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains		
A222_03081	Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains		
A222_04703	Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains		

A222_05717	Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains	
A222_05688	Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains	
A222_05358	Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains	
A222_03945	Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains	
A222_04896	Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains	
A222_03947	Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains	
A222_03595	Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains	
A222_05560	Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains	
A222_02096	Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains	
A222_05458	Response regulator of the LytR/AlgR family	
A222_03250	Response regulator with putative antiterminator output	

	domain	
A222_01600	Response regulator with putative antiterminator output domain	
A222_00184	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	
A222_00419	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	
A222_00418	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	
A222_04951	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	
A222_02354	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	
A222_04302	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	
A222_01769	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	
A222_03865	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	
A222_02560	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	

A222_04510	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	
A222_00473	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	
A222_03408	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	
A222_03230	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	
A222_01888	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	
A222_03887	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	
A222_00932	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	
A222_02323	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	
A222_04118	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	
A222_01782	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	

A222_05168	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	
A222_05393	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	
A222_05742	RNA polymerase-binding protein DksA	
A222_04892	RNA polymerase-binding protein DksA	
A222_02191	Serine phosphatase RsbU, regulator of sigma subunit	
A222_01617	Serine phosphatase RsbU, regulator of sigma subunit	
A222_00177	Serine phosphatase RsbU, regulator of sigma subunit	
A222_03368	Serine/threonine protein kinase (EC:2.7.11.-)	
A222_00076	Serine/threonine protein kinase (EC:2.7.11.1)	
A222_04956	Serine/threonine protein kinase involved in cell cycle control (EC:2.7.11.1)	
A222_03253	Serine/threonine protein phosphatise	
A222_00077	Serine/threonine protein phosphatase (EC:3.1.3.16)	
A222_03369	Serine/threonine protein phosphatase (EC:3.1.3.16)	
A222_00813	SH3 domain protein	
A222_02684	Sigma-54 interaction domain.	
A222_04734	Sigma54-dependent transcription regulator containing an AAA-type ATPase domain and a DNA-binding	

	domain	
A222_04509	Signal transduction histidine kinase	
A222_00989	Signal transduction histidine kinase	
A222_01887	Signal transduction histidine kinase	
A222_02559	Signal transduction histidine kinase	
A222_01767	Signal transduction histidine kinase	
A222_02355	Signal transduction histidine kinase	
A222_02322	Signal transduction histidine kinase	
A222_04117	Signal transduction histidine kinase	
A222_01783	Signal transduction histidine kinase	
A222_03654	Signal transduction histidine kinase	
A222_03434	Signal transduction histidine kinase	
A222_01921	Signal transduction histidine kinase	
A222_05166	Signal transduction histidine kinase	
A222_02154	Signal transduction histidine kinase	
A222_01496	Signal transduction histidine kinase	
A222_05036	Signal transduction histidine kinase	
A222_02430	Signal transduction histidine kinase	
A222_03946	Signal transduction histidine kinase (EC:2.7.13.3)	
A222_04629	Signal transduction histidine kinase (EC:2.7.13.3)	
A222_03231	Signal transduction histidine kinase (EC:2.7.13.3)	
A222_05392	Signal transduction histidine kinase (EC:2.7.13.3)	
A222_03864	Signal transduction histidine kinase	

	(EC:2.7.13.3)	
A222_04301	Signal transduction histidine kinase (EC:2.7.13.3)	
A222_00474	Signal transduction histidine kinase (EC:2.7.13.3)	
A222_04952	Signal transduction histidine kinase (EC:2.7.13.3)	
A222_03886	Signal transduction histidine kinase (EC:2.7.13.3)	
A222_04119	Signal transduction histidine kinase (EC:2.7.13.3)	
A222_05718	Signal transduction histidine kinase regulating C4-dicarboxylate transport system (EC:2.7.13.3)	
A222_05357	Signal transduction histidine kinase regulating C4-dicarboxylate transport system (EC:2.7.13.3)	
A222_03699	Signal transduction histidine kinase regulating C4-dicarboxylate transport system (EC:2.7.13.3)	
A222_01091	Signal transduction histidine kinase, nitrate/nitrite-specific (EC:2.7.13.3)	
A222_05312	Signal transduction histidine kinase, nitrogen specific (EC:2.7.13.3)	
A222_01959	SOS regulatory protein LexA (EC:3.4.21.88)	
A222_04196	Stress-induced morphogen (activity unknown)	
A222_03995	Sugar diacid utilization regulator	
A222_03376	Transcriptional regulator containing GAF, AAA-type	

	ATPase, and DNA binding domains	
A222_02345	Transcriptional regulator containing GAF, AAA-type ATPase, and DNA binding domains	
A222_03846	Transcriptional regulator containing PAS, AAA-type ATPase, and DNA-binding domains	
A222_03020	Transcriptional regulator containing PAS, AAA-type ATPase, and DNA-binding domains	
A222_02679	Transcriptional regulators containing an AAA-type ATPase domain and a DNA-binding domain	
A222_01033	Transcriptional regulators containing an AAA-type ATPase domain and a DNA-binding domain	
A222_03374	type VI secretion system FHA domain protein	
A222_00083	type VI secretion system FHA domain protein	
A222_00084	type VI secretion system FHA domain protein	
A222_00859	uncharacterized domain HDIG	
A222_01949	Universal stress protein UspA and related nucleotide-binding proteins	
A222_02145	Universal stress protein UspA and related nucleotide-binding proteins	
A222_01654	Universal stress protein UspA and related nucleotide-binding proteins	

A222_03283	Universal stress protein UspA and related nucleotide-binding proteins	
A222_05213	Universal stress protein UspA and related nucleotide-binding proteins	
A222_04480	Universal stress protein UspA and related nucleotide-binding proteins	
A222_03246	Universal stress protein UspA and related nucleotide-binding proteins	
A222_04456	Universal stress protein UspA and related nucleotide-binding proteins	
A222_00046	Yersinia virulence determinant (YopE)./Clostridial binary toxin A.	

4.13: Transcription genes

Transcription is the first step in gene expression, in which information from a gene is used to construct a functional product such as a protein, In *P. aeruginosa* N002 and *Enterobacter* sp.RC4 having a total of 486 and 356 transcription genes is described in the table 4.11.

Table 4.11: Transcription genes of *P. aeruginosa* N002 and *Enterobacter* sp. RC4

***P. aeruginosa* N002**		***Enterobater* sp. R4**	
Locus Tag	**Gene Product Name**	**Locus Tag**	**Gene Product Name**
A222_00032	putative choline sulfate-utilization transcription factor	C4L14_00030	LysR family transcriptional regulator
A222_00035	two component transcriptional regulator, LuxR family	C4L14_00220	GntR family transcriptional regulator
A222_00038	DNA-binding transcriptional regulator, LysR family	C4L14_00745	predicted DNA-binding transcriptional regulator YafY

A222_00054	Transcriptional regulator, contains XRE-family HTH domain	C4L14_00845	AlpA family transcriptional regulator
A222_00062	transcriptional regulator, LysR family	C4L14_00925	AsnC family transcriptional regulator
A222_00123	transcriptional regulator, HxlR family	C4L14_01110	DNA-binding transcriptional LysR family regulator
A222_00127	transcriptional regulator, GntR family	C4L14_01125	GntR family transcriptional regulator
A222_00128	DNA-binding transcriptional regulator, GntR family	C4L14_01200	GntR family transcriptional regulator
A222_00130	transcriptional regulator, LysR family	C4L14_01235	LysR family D-serine deaminase transcriptional activator
A222_00132	Predicted transcriptional regulator	C4L14_01290	LuxR family two component transcriptional regulator
A222_00135	phosphonoacetate hydrolase	C4L14_01315	DNA-binding transcriptional LysR family regulator
A222_00140	DNA-binding transcriptional regulator, LysR family	C4L14_01355	transcriptional regulator with PAS, ATPase and Fis domain
A222_00156	RNA polymerase sigma-70 factor, ECF subfamily	C4L14_01400	LacI family transcriptional regulator
A222_00160	LysR family transcriptional regulator, pca operon transcriptional activator	C4L14_01450	DNA-binding NarL/FixJ family response regulator
A222_00163	transcriptional regulator, IclR family	C4L14_01455	two-component system capsular synthesis response regulator RcsB
A222_00168	transcriptional regulator, LysR family	C4L14_01480	DNA-binding winged helix-turn-helix (wHTH) protein
A222_00171	transcriptional regulator, AraC family	C4L14_01545	transcriptional regulator GlxA family with amidase domain
A222_00175	transcriptional regulator, TetR family	C4L14_01570	DNA-binding transcriptional MerR regulator
A222_00180	Serine phosphatase RsbU, regulator of sigma subunit	C4L14_01605	two-component system response regulator FimZ (fimbrial Z protein)
A222_00189	transcriptional regulator, LysR family	C4L14_01610	DNA-binding winged helix-turn-helix (wHTH)

			protein
A222_00198	DNA-binding transcriptional regulator, LysR family	C4L14_01620	transcriptional regulator GlxA family with amidase domain
A222_00223	LysR family transcriptional regulator, malonate utilization transcriptional regulator	C4L14_01760	LysR family transcriptional regulator
A222_00224	DNA-binding transcriptional regulator, LysR family	C4L14_01770	DNA-binding NarL/FixJ family response regulator
A222_00231	transcriptional regulator, XRE family with cupin sensor	C4L14_01895	two-component system copper resistance phosphate regulon response regulator CusR
A222_00239	DNA-binding transcriptional regulator, LysR family	C4L14_01965	two-component system copper resistance phosphate regulon response regulator CusR
A222_00242	transcriptional regulator, IclR family	C4L14_02205	RNA polymerase RpoH-like sigma 32 subunit
A222_00249	transcriptional regulator, TetR family	C4L14_02300	LacI family transcriptional regulator
A222_00254	transcriptional regulator, AraC family	C4L14_02370	DeoR family transcriptional regulator
A222_00259	DNA-binding transcriptional regulator, MarR family	C4L14_02375	LuxR family maltose regulon positive regulatory protein
A222_00280	transcriptional regulator, GntR family	C4L14_02525	MarR family transcriptional regulator
A222_00284	DNA-binding transcriptional regulator, LysR family	C4L14_02530	LysR family transcriptional regulator
A222_00291	transcriptional regulator, ArsR family	C4L14_02545	RNA polymerase RpoS-like sigma 38 subunit
A222_00301	DNA-binding transcriptional regulator, LysR family	C4L14_02650	RpiR family transcriptional regulator
A222_00306	transcriptional regulator, TetR family	C4L14_02675	GTP pyrophosphokinase
A222_00319	AraC-type DNA-binding protein	C4L14_02770	LysR family glycine cleavage system transcriptional activator
A222_00385	transcriptional regulator, TetR family	C4L14_02910	LacI family transcriptional regulator

A222_00395	RNA polymerase, sigma 32 subunit, RpoH	C4L14_02920	GntR family transcriptional regulator
A222_00424	putative holliday junction resolvase	C4L14_02935	LacI family transcriptional regulator
A222_00425	putative transcriptional regulator	C4L14_02945	DNA-binding transcriptional LysR family regulator
A222_00436	transcriptional regulator, AraC family	C4L14_03015	DNA-binding transcriptional LysR family regulator
A222_00444	MarR family transcriptional regulator, repressor of the mexAB-oprM multidrug resistance operon	C4L14_03030	AsnC family transcriptional regulator
A222_00457	transcriptional regulator, TetR family	C4L14_03070	LysR family cys regulon transcriptional activator
A222_00467	DNA-binding transcriptional regulator, LysR family	C4L14_03125	DeoR family transcriptional regulator
A222_00476	cold-shock DNA-binding protein family	C4L14_03145	exoribonuclease-2
A222_00483	two-component system, OmpR family, catabolic regulation response regulator CreB	C4L14_03175	two-component system response regulator AdeR
A222_00493	RNA polymerase sigma-70 factor, ECF subfamily	C4L14_03240	psp operon transcriptional activator
A222_00496	transcriptional regulator, TetR family	C4L14_03245	phage shock protein A (PspA) family protein
A222_00498	transcriptional regulator, LysR family	C4L14_03255	phage shock protein C (PspC) family protein
A222_00500	DNA-binding transcriptional regulator, LysR family	C4L14_03320	LacI family transcriptional regulator
A222_00508	molybdate transport system regulatory protein	C4L14_03335	transcriptional regulator of aroF, aroG, tyrA and aromatic amino acid transport
A222_00512	DNA-binding transcriptional regulator, LysR family	C4L14_03600	GntR family transcriptional regulator
A222_00530	transcriptional regulator, AsnC family	C4L14_03755	AraC family transcriptional regulator
A222_00531	transcriptional regulator, AsnC family	C4L14_03805	two-component system response regulator QseB

A222_00532	transcriptional regulator, AsnC family	C4L14_03845	AraC-like DNA-binding protein
A222_00533	transcriptional regulator, AsnC family	C4L14_03850	RpiR family transcriptional regulator
A222_00547	DNA-binding transcriptional regulator, LysR family	C4L14_04045	LuxR family two component transcriptional regulator
A222_00555	transcriptional regulator, XRE family with cupin sensor	C4L14_04070	RNA polymerase RpoD-like sigma 70 subunit
A222_00594	Fic family protein	C4L14_04345	LuxR family two component transcriptional regulator
A222_00598	RNA polymerase, sigma 70 subunit, RpoD	C4L14_04400	DNA-binding PadR family transcriptional regulator
A222_00623	two component transcriptional regulator, LuxR family	C4L14_04495	RpiR family transcriptional regulator
A222_00703	transcription antitermination protein nusG	C4L14_04580	LysR family transcriptional regulator (chromosome initiation inhibitor)
A222_00711	DNA-directed RNA polymerase subunit beta'	C4L14_04680	IclR family transcriptional regulator
A222_00743	DNA-directed RNA polymerase subunit alpha	C4L14_04760	putative transcriptional regulator
A222_00754	transcriptional regulator, AraC family	C4L14_04765	putative Holliday junction resolvase
A222_00779	DNA-binding transcriptional regulator, LysR family	C4L14_04885	two-component system response regulator CitB
A222_00786	two component transcriptional regulator, LuxR family	C4L14_04920	LuxR family two component transcriptional regulator
A222_00792	DNA-binding transcriptional regulator, GntR family	C4L14_05055	RpiR family transcriptional regulator
A222_00793	transcriptional regulator, AraC family with amidase-like domain	C4L14_05090	DNA-binding GntR family transcriptional regulator
A222_00803	DNA-binding transcriptional regulator, LysR family	C4L14_05225	regulator of competence-specific genes
A222_00808	transcriptional regulator, BadM/Rrf2 family	C4L14_05560	FlgM family anti-sigma-28 factor
A222_00812	transcriptional regulator, GntR family	C4L14_05635	LysR family transcriptional regulator

A222_00820	transcriptional regulator, IclR family	C4L14_05775	TetR family transcriptional regulator
A222_00833	DNA-binding transcriptional regulator, LysR family	C4L14_05790	transcription-repair coupling factor
A222_00843	homoprotocatechuate degradation operon regulator, HpaR	C4L14_05815	N-acetylglucosamine kinase
A222_00846	transcriptional regulator, GntR family	C4L14_05870	two-component system response regulator PhoP
A222_00858	transcriptional regulator, AraC family	C4L14_05905	LuxR family two component transcriptional regulator
A222_00869	LysR family transcriptional regulator, regulator of gene expression of beta-lactamase	C4L14_05980	AraC-like DNA-binding protein
A222_00894	two component transcriptional regulator, LuxR family	C4L14_06085	RNA polymerase RpoE-like sigma-24 subunit
A222_00897	putative transcriptional regulator	C4L14_06120	RNAse III
A222_00904	transcriptional regulator, AraC family	C4L14_06150	RpiR family transcriptional regulator
A222_00917	transcriptional repressor NrdR	C4L14_06235	BglG family transcriptional antiterminator
A222_00923	NusB antitermination factor	C4L14_06265	BadM/Rrf2 family transcriptional regulator
A222_00943	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain	C4L14_06315	RpiR family transcriptional regulator
A222_00956	Transcriptional regulator of acetoin/glycerol metabolism	C4L14_06495	LacI family transcriptional regulator
A222_00982	DNA-binding transcriptional regulator, LysR family	C4L14_06555	LuxR family two component transcriptional regulator
A222_01004	transcriptional regulator, TetR family	C4L14_06565	DeoR family transcriptional regulator
A222_01013	transcriptional regulator, AsnC family	C4L14_06595	DNA-binding transcriptional LysR family regulator
A222_01031	two component transcriptional regulator, LuxR family	C4L14_06605	GntR family transcriptional regulator

A222_01047	Transcriptional regulator containing an AAA-type ATPase domain and a DNA-binding domain	C4L14_06615	GntR family transcriptional regulator/MocR family aminotransferase
A222_01052	transcriptional regulator, AraC family	C4L14_06625	DNA-binding transcriptional LysR family regulator
A222_01058	LuxR family transcriptional regulator, maltose regulon positive regulatory protein	C4L14_06655	DNA-binding transcriptional ArsR family regulator
A222_01081	RNA polymerase sigma-70 factor, ECF subfamily	C4L14_06665	MarR family transcriptional regulator
A222_01085	transcriptional regulator, LysR family	C4L14_06700	AraC-like DNA-binding protein
A222_01104	two component transcriptional regulator, LuxR family	C4L14_06740	predicted DNA-binding transcriptional regulator YafY
A222_01115	Predicted transcriptional regulator YheO, contains PAS and DNA-binding HTH domains	C4L14_06755	AraC family 4-hydroxyphenylacetate 3-monooxygenase operon regulatory protein
A222_01134	LysR family transcriptional regulator, glycine cleavage system transcriptional activator	C4L14_06795	homoprotocatechuate degradation regulator HpaR
A222_01149	AraC-type DNA-binding protein	C4L14_06835	MarR family transcriptional regulator
A222_01165	transcriptional regulator, BadM/Rrf2 family	C4L14_06840	TetR family transcriptional regulator
A222_01198	transcriptional regulator, AraC family with amidase-like domain	C4L14_06855	DNA-binding NarL/FixJ family response regulator
A222_01203	DNA-binding transcriptional regulator, LysR family	C4L14_06860	DNA-binding NarL/FixJ family response regulator
A222_01205	DNA-binding transcriptional regulator, LysR family	C4L14_07015	LysR family transcriptional regulator
A222_01225	transcriptional regulator, GntR family	C4L14_07035	Trp operon repressor
A222_01252	phage shock protein A (PspA) family protein	C4L14_07050	AraC family transcriptional regulator
A222_01262	transcriptional regulator, TetR family	C4L14_07060	two-component system aerobic respiration control protein ArcA

A222_01268	two component transcriptional regulator, LuxR family	C4L14_07145	LysR family transcriptional regulator
A222_01271	transcriptional regulator, LysR family	C4L14_07270	ATP-dependent helicase HepA
A222_01280	response regulator receiver modulated diguanylate cyclase	C4L14_07295	AraC family transcriptional regulator
A222_01283	transcriptional regulator, TetR family	C4L14_07320	SgrR family transcriptional regulator
A222_01293	Cd(II)/Pb(II)-responsive transcriptional regulator	C4L14_07365	LysR family transcriptional activator for leuABCD operon
A222_01304	transcriptional regulator, TetR family	C4L14_07380	LacI family transcriptional regulator
A222_01353	transcriptional regulator, LysR family	C4L14_07545	GntR family transcriptional regulator
A222_01361	RNA polymerase, sigma 38 subunit, RpoS	C4L14_07610	LuxR family two component transcriptional regulator
A222_01380	two component transcriptional regulator, LuxR family	C4L14_07830	CdaR family transcriptional regulator
A222_01385	DNA-binding transcriptional regulator, CsgD family	C4L14_07890	regulator of sigma E protease
A222_01390	DNA-binding transcriptional regulator, LysR family	C4L14_07955	Rof transcriptional antiterminator
A222_01397	LysR family transcriptional regulator, regulator for metE and meth	C4L14_08040	alkylphosphonate utilization operon protein PhnA
A222_01401	transcriptional regulator, DeoR family	C4L14_08065	GntR family transcriptional regulator
A222_01411	transcriptional regulator, TetR family	C4L14_08170	DNA-binding response OmpR family regulator
A222_01414	transcriptional regulator, AraC family	C4L14_08220	DNA-binding transcriptional LysR family regulator
A222_01420	transcriptional regulator, LysR family	C4L14_08250	MerR family redox-sensitive transcriptional activator SoxR
A222_01421	Barstar, RNAse (barnase) inhibitor	C4L14_08255	AraC family transcriptional regulator
A222_01423	transcriptional regulator, LacI family	C4L14_08270	RNA polymerase sigma-70 factor (ECF subfamily)

A222_01491	LuxR family transcriptional regulator, transcriptional activator of rhlAB and lasB	C4L14_08325	predicted DNA-binding protein (MmcQ/YjbR family)
A222_01510	DNA-binding transcriptional regulator, MarR family	C4L14_08395	SOS-response transcriptional repressor LexA
A222_01535	DNA-binding transcriptional regulator, LysR family	C4L14_08450	GntR family transcriptional regulator
A222_01545	transcriptional regulator, AraC family	C4L14_08500	MerR family transcriptional regulator
A222_01548	ATP-, maltotriose- and DNA-dependent transcriptional regulator MalT	C4L14_08600	LacI family transcriptional regulator
A222_01559	RNA polymerase sigma-70 factor, ECF subfamily	C4L14_08650	LuxR family two component transcriptional regulator
A222_01572	DNA-binding transcriptional regulator, LysR family	C4L14_08660	fimbrial protein FimY
A222_01579	NosR/NirI family transcriptional regulator, nitrous oxide reductase regulator	C4L14_08745	LacI family transcriptional regulator
A222_01589	transcriptional regulator, GntR family	C4L14_08785	predicted DNA-binding protein (MmcQ/YjbR family)
A222_01613	Two-component response regulator, AmiR/NasT family, consists of REC and RNA-binding antiterminator (ANTAR) domains	C4L14_08790	TetR family transcriptional regulator
A222_01624	anti-sigma-28 factor, FlgM family	C4L14_08835	AraC-like DNA-binding protein
A222_01629	Serine phosphatase RsbU, regulator of sigma subunit	C4L14_08965	TetR family transcriptional regulator
A222_01634	MarR family transcriptional regulator, transcriptional regulator for hemolysin	C4L14_09000	GntR family transcriptional regulator/MocR family aminotransferase
A222_01654	transcriptional regulator, LysR family	C4L14_09025	DNA-binding FrmR family transcriptional regulator
A222_01667	ATP-dependent helicase HepA	C4L14_09040	LacI family transcriptional regulator

A222_01696	RNA polymerase sigma-70 factor, ECF subfamily	C4L14_09050	DNA-binding transcriptional MocR family regulator
A222_01714	transcriptional regulator, AraC family	C4L14_09070	O(6)-alkylguanine repair protein YbaZ
A222_01718	cold-shock DNA-binding protein family	C4L14_09105	AsnC family transcriptional regulator
A222_01740	transcriptional regulator, GntR family	C4L14_09120	MarR-like DNA-binding transcriptional regulator SgrR of sgrS sRNA
A222_01765	DNA-binding transcriptional regulator, LysR family	C4L14_09275	NusB antitermination factor
A222_01770	transcriptional regulator, AraC family	C4L14_09290	transcriptional repressor NrdR
A222_01775	transcriptional regulator, AraC family	C4L14_09305	predicted DNA-binding transcriptional regulator YafY
A222_01786	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain	C4L14_09400	transcriptional antiterminator RfaH
A222_01793	condensin subunit ScpB	C4L14_09460	LysR family transcriptional regulator for metE and metH
A222_01799	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain	C4L14_09665	transcription termination factor Rho
A222_01807	transcriptional regulator, RpiR family	C4L14_09700	LysR family positive regulator for ilvC
A222_01817	transcriptional regulator, IclR family	C4L14_09750	DNA-binding transcriptional LysR family regulator
A222_01839	transcriptional regulator, LysR family	C4L14_09780	two-component system OmpR family response regulator
A222_01845	transcriptional regulator, TetR family	C4L14_09805	DeoR family transcriptional regulator
A222_01854	transcriptional regulator, LysR family	C4L14_09870	two-component system copper resistance phosphate regulon response regulator CusR
A222_01856	DNA-binding transcriptional regulator, LysR family	C4L14_09880	LysR family transcriptional regulator

A222_01887	transcriptional regulator, AraC family	C4L14_09905	predicted MarR family transcription regulator
A222_01904	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain	C4L14_09935	DNA-binding transcriptional LysR family regulator
A222_01923	Transcriptional regulator, contains XRE-family HTH domain	C4L14_09940	DNA-binding transcriptional LysR family regulator
A222_01933	two component transcriptional regulator, LuxR family	C4L14_09960	sigma-B regulation protein RsbU (phosphoserine phosphatase)
A222_01944	transcriptional regulator, TetR family	C4L14_10040	RNA polymerase sigma factor (sigma-70 family)
A222_01952	transcriptional regulator, AraC family	C4L14_10060	LacI family transcriptional regulator
A222_01972	SOS-response transcriptional repressor, LexA	C4L14_10155	two-component system response regulator CpxR
A222_01973	transcriptional regulator, TetR family	C4L14_10180	AraC family transcriptional regulator
A222_01977	transcription-repair coupling factor	C4L14_10225	LacI family transcriptional regulator
A222_02026	transcriptional regulator, TetR family	C4L14_10335	DeoR family transcriptional regulator
A222_02057	transcriptional regulator, TetR family	C4L14_10505	beta-glucoside kinase
A222_02058	DNA-binding transcriptional regulator, LysR family	C4L14_10785	DNA-binding transcriptional LysR family regulator
A222_02067	DNA-binding transcriptional regulator, LysR family	C4L14_10880	uncharacterized protein
A222_02071	transcriptional regulator, AraC family	C4L14_10885	transcription elongation factor GreB
A222_02091	two component transcriptional regulator, LuxR family	C4L14_10895	two-component system phosphate regulon response regulator OmpR
A222_02093	transcriptional regulator, GntR family	C4L14_11160	LysR family malonate utilization transcriptional regulator
A222_02094	RNA polymerase, sigma subunit, ECF family	C4L14_11205	predicted transcriptional regulator YheO
A222_02105	transcriptional regulator, TetR family	C4L14_11260	AraC-like DNA-binding protein

A222_02111	DNA-binding transcriptional regulator, LysR family	C4L14_11270	TetR family transcriptional regulator
A222_02113	transcriptional regulator, LysR family	C4L14_11350	DNA-binding NarL/FixJ family response regulator
A222_02131	transcription elongation factor GreB	C4L14_11370	LysR family transcriptional regulator of beta-lactamase
A222_02142	transcriptional regulator, MarR family	C4L14_11495	BadM/Rrf2 family transcriptional regulator
A222_02143	transcriptional regulator, LysR family	C4L14_11500	RNAse R
A222_02145	DNA-binding transcriptional regulator, LysR family	C4L14_11540	DeoR family transcriptional regulator
A222_02153	transcriptional regulator, LysR family	C4L14_11660	AsnC family transcriptional regulator
A222_02157	transcriptional regulator, LysR family	C4L14_11750	putative cold-shock DNA-binding protein
A222_02166	DNA-binding transcriptional regulator, MarR family	C4L14_11895	LysR family transcriptional regulator
A222_02183	RNA polymerase sigma-70 factor, ECF subfamily	C4L14_11915	cold shock CspA family protein
A222_02193	two-component system, OmpR family, copper resistance phosphate regulon response regulator CusR	C4L14_11950	predicted transcriptional regulator YheO
A222_02200	transcriptional regulator, GntR family	C4L14_12020	TetR family transcriptional regulator
A222_02204	Serine phosphatase RsbU, regulator of sigma subunit	C4L14_12215	DNA-binding transcriptional LysR family regulator
A222_02217	putative transcriptional regulator	C4L14_12240	regulator of nucleoside diphosphate kinase
A222_02222	DNA-binding transcriptional regulator, XRE-family HTH domain	C4L14_12270	cold shock protein E (CspE)
A222_02247	DNA-binding transcriptional regulator, LysR family	C4L14_12295	DNA-binding transcriptional LysR family regulator
A222_02279	transcriptional regulator, MerR family	C4L14_12470	N-acetylglucosamine repressor NagC
A222_02295	DNA-binding transcriptional regulator,	C4L14_12555	two-component system KDP operon response

	MerR family		regulator KdpE
A222_02300	transcriptional regulator, HxlR family	C4L14_12675	GntR family mannosyl-D-glycerate transport/metabolism transcriptional repressor
A222_02309	AraC-type DNA-binding protein	C4L14_12830	molybdate transport repressor ModE
A222_02317	transcriptional regulator, AraC family	C4L14_12880	GntR family histidine utilization transcriptional repressor
A222_02321	transcriptional regulator, BadM/Rrf2 family	C4L14_13020	TetR family transcriptional regulator
A222_02326	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain	C4L14_13135	DtxR family iron (metal) dependent repressor
A222_02332	transcriptional regulator, LysR family	C4L14_13175	GntR family transcriptional regulator
A222_02348	anaerobic nitric oxide reductase transcription regulator	C4L14_13200	LacI family transcriptional regulator
A222_02357	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain	C4L14_13290	DeoR family transcriptional regulator
A222_02393	cold-shock DNA-binding protein family	C4L14_13320	TetR family transcriptional regulator
A222_02416	DNA-binding transcriptional regulator, LysR family	C4L14_13335	ArsR family transcriptional regulator
A222_02429	transcriptional regulator, AraC family	C4L14_13410	DNA-binding response OmpR family regulator
A222_02431	two component transcriptional regulator, LuxR family	C4L14_13480	PaaX family transcriptional regulator
A222_02442	transcriptional regulator, AsnC family	C4L14_13575	AraC family transcriptional activator of tynA and feaB
A222_02466	AraC-type DNA-binding protein	C4L14_13625	GntR family transcriptional regulator
A222_02471	transcriptional regulator, LysR family	C4L14_14550	DNA-binding transcriptional LysR family regulator
A222_02476	transcriptional regulator, LysR family	C4L14_14580	LysR family tdc operon transcriptional activator

A222_02489	DNA-binding transcriptional regulator, LysR family	C4L14_14730	N-acetylglucosamine kinase
A222_02501	two-component system, OmpR family, copper resistance phosphate regulon response regulator CusR	C4L14_14745	two-component system phosphate regulon response regulator PhoB
A222_02506	transcriptional regulator, AraC family	C4L14_14770	DNA-binding winged helix-turn-helix (wHTH) protein
A222_02514	transcriptional regulator, AraC family	C4L14_14810	formate hydrogenlyase transcriptional activator
A222_02515	DNA-binding transcriptional regulator, LysR family	C4L14_14900	LacI family transcriptional regulator
A222_02528	DNA-binding transcriptional regulator, LysR family	C4L14_14930	anaerobic nitric oxide reductase transcription regulator
A222_02533	transcriptional regulator, LysR family	C4L14_14940	DeoR family transcriptional regulator
A222_02536	transcriptional regulator, AraC family	C4L14_14945	glucitol operon activator protein
A222_02537	transcriptional regulator, AraC family	C4L14_15060	AraC-like DNA-binding protein
A222_02541	transcriptional regulator, TetR family	C4L14_15205	DNA-binding transcriptional ArsR family regulator
A222_02546	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain	C4L14_15300	GntR family colanic acid and biofilm gene transcriptional regulator
A222_02556	DNA-binding transcriptional regulator, LysR family	C4L14_15315	LacI family transcriptional regulator
A222_02557	RNA polymerase sigma-70 factor, ECF subfamily	C4L14_15335	HxlR family transcriptional regulator
A222_02577	TyrR family helix-turn-helix domain-containing protein	C4L14_15345	HxlR family transcriptional regulator
A222_02579	transcriptional regulator, LysR family	C4L14_15410	TetR family transcriptional regulator
A222_02595	transcriptional regulator, LysR family	C4L14_15510	LuxR family transcriptional regulator of csgAB operon
A222_02602	RNA polymerase sigma-70 factor, ECF subfamily	C4L14_15640	LacI family transcriptional regulator

A222_02611	transcriptional regulator, LysR family	C4L14_15655	methionine repressor MetJ
A222_02650	RNA polymerase sigma-70 factor, ECF subfamily	C4L14_15750	LysR family hydrogen peroxide-inducible transcriptional activator
A222_02654	DNA-binding transcriptional regulator, LysR family	C4L14_15760	TetR family transcriptional regulator
A222_02661	two component transcriptional regulator, LuxR family	C4L14_16005	DNA-directed RNA polymerase subunit alpha
A222_02676	sigma-54 specific transcriptional regulator	C4L14_16020	MerR family Zn(II)-responsive transcriptional regulator of zntA
A222_02698	transcriptional regulator, AraC family	C4L14_16320	nucleoid protein H-NS
A222_02699	transcriptional regulator, AraC family	C4L14_16360	Lrp/AsnC family leucine-responsive transcriptional regulator
A222_02712	transcriptional regulator, LacI family	C4L14_16400	LuxR family two component transcriptional regulator
A222_02715	transcriptional regulator, LysR family	C4L14_16535	DNA-binding transcriptional MerR regulator
A222_02720	transcriptional regulator, XRE family with cupin sensor	C4L14_16720	transcriptional regulator GlxA family with amidase domain
A222_02733	GntR family transcriptional regulator	C4L14_16735	AraC-like DNA-binding protein
A222_02751	transcriptional regulator, AraC family	C4L14_16800	GntR family transcriptional regulator
A222_02755	transcriptional regulator, ArsR family	C4L14_16950	putative cold-shock DNA-binding protein
A222_02756	transcriptional regulator, AraC family	C4L14_16985	IclR family transcriptional regulator
A222_02759	MerR family transcriptional regulator, redox-sensitive transcriptional activator SoxR	C4L14_17105	RpiR family transcriptional regulator
A222_02762	transcriptional regulator, TetR family	C4L14_17155	YebC/PmpR family DNA-binding regulatory protein
A222_02765	transcriptional regulator, LysR family	C4L14_17465	LuxR family two component transcriptional regulator

A222_02773	transcriptional regulator, LacI family	C4L14_17475	LuxR family transcriptional regulator
A222_02774	transcriptional regulator, LysR family	C4L14_17505	RNA polymerase sigma-28 (SigD/FliA/WhiG) subunit
A222_02787	transcriptional regulator, AsnC family	C4L14_17585	AraC family transcriptional regulator
A222_02810	transcriptional regulator, propionate catabolism operon regulatory protein	C4L14_17600	DNA-binding transcriptional MocR family regulator
A222_02826	DNA-binding transcriptional regulator, LysR family	C4L14_17605	DNA-binding transcriptional LysR family regulator
A222_02836	transcriptional regulator, TetR family	C4L14_17640	DNA-binding transcriptional LysR family regulator
A222_02905	DNA-binding transcriptional regulator, LysR family	C4L14_17675	DNA-binding transcriptional LysR family regulator
A222_02907	transcriptional regulator, LysR family	C4L14_17695	DNA-binding protein StpA
A222_02914	DNA-binding transcriptional regulator, LysR family	C4L14_17790	regulator of sigma D
A222_02926	transcriptional regulator, AraC family	C4L14_17840	DNA-directed RNA polymerase subunit beta'
A222_02929	RNA polymerase sigma-70 factor, ECF subfamily	C4L14_17875	transcription antitermination protein nusG
A222_02940	transcriptional regulator, AsnC family	C4L14_18105	IclR family transcriptional regulator
A222_02946	DNA-binding transcriptional regulator, LysR family	C4L14_18175	LytTR family two component transcriptional regulator
A222_02965	DNA-binding transcriptional regulator, LysR family	C4L14_18235	TetR family transcriptional regulator
A222_02967	LysR family transcriptional regulator, cyn operon transcriptional activator	C4L14_18295	AlpA family transcriptional regulator
A222_02971	RNA polymerase sigma-70 factor, ECF subfamily	C4L14_18400	GntR family transcriptional regulator
A222_02991	DNA-binding transcriptional regulator, MocR family, contains an aminotransferase domain	C4L14_18490	LysR family transcriptional regulator

A222_02995	transcriptional regulator, AsnC family	C4L14_18545	transcriptional regulator of PTS gene
A222_03003	transcriptional regulator, TetR family	C4L14_18550	LysR family transcriptional regulator
A222_03007	DNA-binding transcriptional regulator, MerR family	C4L14_18580	XRE family transcriptional regulator
A222_03013	transcriptional regulator, IclR family	C4L14_18645	DNA-binding response OmpR family regulator
A222_03018	Transcriptional regulator containing PAS, AAA-type ATPase, and DNA-binding Fis domains	C4L14_18680	DNA-binding transcriptional LysR family regulator
A222_03025	LysR family transcriptional regulator, carnitine catabolism transcriptional activator	C4L14_18710	LysR family transcriptional regulator of abg operon
A222_03028	Predicted DNA-binding protein, MmcQ/YjbR family	C4L14_18795	two-component system response regulator RstA
A222_03044	two component transcriptional regulator, LuxR family	C4L14_18835	LacI family transcriptional regulator
A222_03047	two component transcriptional regulator, LuxR family	C4L14_18955	MarR family transcriptional regulator for hemolysin
A222_03064	transcriptional regulator, LysR family	C4L14_18995	TetR family transcriptional regulator
A222_03076	transcriptional regulator, LacI family	C4L14_19045	LacI family transcriptional regulator
A222_03106	purine catabolism regulatory protein	C4L14_19050	DNA-binding transcriptional LysR family regulator
A222_03109	RNA polymerase sigma-70 factor, ECF subfamily	C4L14_19110	LuxR family two component transcriptional regulator
A222_03127	LuxR family transcriptional regulator	C4L14_19200	DNA-binding transcriptional MerR regulator
A222_03142	Transcriptional regulator, contains XRE-family HTH domain	C4L14_19235	DNA-binding NarL/FixJ family response regulator
A222_03165	transcriptional regulator, TetR family	C4L14_19355	LysR family carnitine catabolism transcriptional activator
A222_03170	transcriptional regulator, LysR family	C4L14_19735	LacI family transcriptional regulator

A222_03176	DNA-binding transcriptional regulator, LysR family	C4L14_19950	MarR family transcriptional repressor of emrRAB
A222_03179	transcriptional regulator, AraC family with amidase-like domain	C4L14_20055	DNA-binding winged helix-turn-helix (wHTH) protein
A222_03195	transcriptional regulator, TetR family	C4L14_20115	two-component system response regulator BaeR
A222_03205	DNA-binding transcriptional regulator, LysR family	C4L14_20165	GntR family transcriptional regulator
A222_03233	two-component system, OmpR family, response regulator ParR	C4L14_20205	LacI family transcriptional regulator
A222_03253	response regulator receiver and ANTAR domain protein	C4L14_20225	DNA-binding transcriptional LysR family regulator
A222_03263	RNA polymerase sigma-70 factor, ECF subfamily	C4L14_20280	LytTR family two component transcriptional regulator
A222_03279	LuxR family transcriptional regulator, maltose regulon positive regulatory protein	C4L14_20290	DNA-binding transcriptional MerR regulator
A222_03280	LuxR family transcriptional regulator, maltose regulon positive regulatory protein	C4L14_20365	AraC-like DNA-binding protein
A222_03285	LysR family transcriptional regulator, cys regulon transcriptional activator	C4L14_20400	DNA-binding XRE family transcriptional regulator
A222_03301	transcriptional regulator, LysR family	C4L14_20420	LacI family transcriptional regulator
A222_03327	transcriptional regulator, AraC family	C4L14_20460	DNA-binding transcriptional LysR family regulator
A222_03378	sigma-54 dependent transcriptional regulator	C4L14_20580	DNA repair protein RadD
A222_03394	DNA-binding transcriptional regulator, MocR family, contains an aminotransferase domain	C4L14_20585	TetR family transcriptional regulator
A222_03395	DNA-binding transcriptional regulator, MarR family	C4L14_20675	LuxR family two component transcriptional regulator

A222_03412	two-component system, OmpR family, KDP operon response regulator KdpE	C4L14_20720	DNA-binding transcriptional LysR family regulator
A222_03419	transcriptional regulator, IclR family	C4L14_20765	transcriptional regulator
A222_03422	transcriptional regulator, GntR family	C4L14_20900	DNA-binding transcriptional LysR family regulator
A222_03430	transcriptional regulator, AraC family	C4L14_20970	LacI family transcriptional regulator
A222_03442	transcriptional regulator, HxlR family	C4L14_21030	RpiR family transcriptional regulator
A222_03446	transcriptional regulator, MarR family	C4L14_21065	LysR family transcriptional regulator
A222_03450	transcriptional regulator, AraC family	C4L14_21465	DNA-binding transcriptional LysR family regulator
A222_03482	aminoethylphosphonate catabolism associated LysR family transcriptional regulator	C4L14_21545	LysR family nitrogen assimilation transcriptional regulator
A222_03514	transcriptional regulator, TetR family	C4L14_21550	LysR family cys regulon transcriptional activator
A222_03529	transcriptional regulator, GntR family	C4L14_21735	DNA-binding winged helix-turn-helix (wHTH) protein
A222_03536	transcriptional regulator, GntR family	C4L14_21780	FlgM family anti-sigma-28 factor
A222_03554	transcriptional regulator, TetR family	C4L14_21850	RNA polymerase sigma factor for flagellar operon FliA
A222_03568	transcriptional regulator, DeoR family	C4L14_21905	DNA-binding transcriptional LysR family regulator
A222_03590	transcriptional regulator	C4L14_21930	two-component system response regulator DcuR
A222_03602	RNA polymerase, sigma 28 subunit, SigD/FliA/WhiG	C4L14_22005	LuxR family capsular biosynthesis transcriptional activator
A222_03620	two-component system, OmpR family, copper resistance phosphate regulon response regulator CusR	C4L14_22400	TetR family transcriptional regulator

A222_03627	LuxR family transcriptional regulator, quorum-sensing transcription factor LasR	C4L14_22415	DNA-binding protein Fis
A222_03634	transcriptional regulator	C4L14_22520	DNA-binding transcriptional LysR family regulator
A222_03643	DNA-binding transcriptional regulator, LysR family	C4L14_22545	ribonuclease inhibitor
A222_03653	transcriptional regulator, TetR family	C4L14_22560	ArgR family transcriptional regulator
A222_03657	transcriptional regulator, LysR family	C4L14_22740	RNA polymerase RpoN-/SigL-like sigma 54 subunit
A222_03659	two component transcriptional regulator, LuxR family	C4L14_22810	Nlp family transcriptional regulator
A222_03663	AraC-type DNA-binding protein	C4L14_22845	two-component system response regulator BasR
A222_03670	transcriptional regulator, HxlR family	C4L14_22855	transcription elongation factor GreA
A222_03675	RNA polymerase sigma-70 factor, ECF subfamily	C4L14_22910	NusA antitermination factor
A222_03679	transcriptional regulator, XRE family with cupin sensor	C4L14_23040	DeoR family transcriptional regulator
A222_03687	RNA polymerase, sigma subunit, ECF family	C4L14_23095	LysR family tdc operon transcriptional activator
A222_03691	DNA-binding response regulator, NarL/FixJ family, contains REC and HTH domains	C4L14_23135	DNA-binding transcriptional LysR family regulator
A222_03710	transcriptional regulator, LysR family	C4L14_23185	GntR family transcriptional regulator
A222_03724	transcriptional regulator, TetR family	C4L14_23240	DNA-binding transcriptional regulator LsrR (DeoR family)
A222_03727	DNA-binding transcriptional regulator, LysR family	C4L14_23320	two-component system response regulator PhoP
A222_03730	aminoethylphosphonate catabolism associated LysR family transcriptional regulator	C4L14_23330	DNA-binding winged helix-turn-helix (wHTH) protein
A222_03739	RNA polymerase sigma-70 factor, ECF subfamily	C4L14_23365	two-component system response regulator QseB

A222_03741	DNA-binding transcriptional regulator, FrmR family	C4L14_23375	BadM/Rrf2 family transcriptional regulator
A222_03749	transcriptional regulator, TetR family	C4L14_23390	AraC family transcriptional regulator
A222_03754	transcriptional regulator, MarR family	C4L14_23425	DNA-binding transcriptional LysR family regulator
A222_03756	transcriptional regulator, TetR family	C4L14_23450	TetR family transcriptional regulator
A222_03771	transcriptional regulator, GntR family	C4L14_23460	TetR family transcriptional regulator
A222_03777	DNA-binding transcriptional regulator, LysR family	C4L14_23470	TetR family transcriptional regulator
A222_03780	transcriptional regulator, AraC family	C4L14_23535	DNA-binding transcriptional LysR family regulator
A222_03805	transcriptional regulator, AraC family	C4L14_23635	DeoR family transcriptional regulator
A222_03811	transcriptional regulator, AraC family	C4L14_23685	DNA-binding transcriptional LysR family regulator
A222_03814	transcriptional regulator, TetR family	C4L14_23695	DNA-binding FrmR family transcriptional regulator
A222_03838	transcriptional regulator, LysR family	C4L14_23715	DNA-binding transcriptional LysR family regulator
A222_03843	arginine utilization regulatory protein	C4L14_23800	LysR family glycine cleavage system transcriptional activator
A222_03855	LysR family transcriptional regulator, glycine cleavage system transcriptional activator	C4L14_23865	XRE family transcriptional regulator
A222_03857	transcriptional regulator, AraC family	C4L14_23885	AraC family transcriptional activator FtrA
A222_03860	two-component system, OmpR family, response regulator PhoP	C4L14_23950	MarR family multiple antibiotic resistance transcriptional regulator
A222_03880	cold-shock DNA-binding protein family	C4L14_24040	DeoR family transcriptional regulator
A222_03882	two-component system, OmpR family, response regulator RstA	C4L14_24050	DNA-binding transcriptional LysR family regulator
A222_03891	transcriptional regulator, LysR family	C4L14_24115	AlpA family transcriptional regulator

A222_03894	transcriptional regulator, GntR family	C4L14_24140	DNA-binding transcriptional ArsR family regulator
A222_03895	DNA-binding transcriptional regulator, LysR family	C4L14_24150	RNA polymerase sigma (SigZ) subunit
A222_03898	DNA-binding transcriptional regulator, LysR family	C4L14_24240	guanosine-3',5'-bis(diphosphate) 3'-pyrophosphohydrolase
A222_03900	LuxR family transcriptional regulatory, chaperone HchA-associated	C4L14_24245	DNA-directed RNA polymerase subunit omega
A222_03908	transcriptional regulator, LysR family	C4L14_24270	AlpA family transcriptional regulator
A222_03929	transcriptional regulator, AraC family	C4L14_24340	TetR family transcriptional regulator
A222_03971	transcriptional regulator, LysR family	C4L14_24530	GntR family transcriptional regulator
A222_03989	transcriptional regulator, CdaR family	C4L14_24565	mannitol repressor MtlR
A222_04011	Transcriptional regulator, contains XRE-family HTH domain	C4L14_24610	DNA-binding transcriptional LysR family regulator
A222_04027	transcriptional regulator, IclR family	C4L14_24685	IclR family transcriptional regulator
A222_04040	transcriptional regulator	C4L14_24715	AraC family transcriptional regulator
A222_04074	DNA-binding regulatory protein, YebC/PmpR family	C4L14_24755	transcriptional regulator GlxA family with amidase domain
A222_04077	cold-shock DNA-binding protein family	C4L14_24785	putative transcriptional regulator
A222_04098	DNA-binding transcriptional regulator, MarR family	C4L14_24795	putative cold-shock DNA-binding protein
A222_04101	transcriptional antiterminator, Rof	C4L14_24820	LacI family transcriptional regulator
A222_04107	GTP pyrophosphokinase		
A222_04112	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain		
A222_04153	transcriptional regulator, AraC family with amidase-like domain		

A222_04171	DNA-binding transcriptional regulator, LysR family	
A222_04172	transcriptional regulator, LysR family	
A222_04175	transcriptional regulator of aroF, aroG, tyrA and aromatic amino acid transport	
A222_04184	transcriptional regulator, AraC family	
A222_04210	transcriptional regulator, TetR family	
A222_04219	transcriptional regulator, AraC family	
A222_04222	transcriptional regulator, TetR family	
A222_04235	DNA-binding transcriptional regulator, LysR family	
A222_04236	DNA-binding transcriptional regulator, LysR family	
A222_04252	Superfamily II DNA or RNA helicase, SNF2 family	
A222_04254	transcriptional regulator, GntR family	
A222_04261	transcriptional regulator, AraC family	
A222_04269	DNA-binding transcriptional regulator, LysR family	
A222_04273	transcriptional regulator, AraC family	
A222_04283	RNAse III	
A222_04291	RNA polymerase, sigma-24 subunit, RpoE	
A222_04298	two-component system, OmpR family, response regulator TctD	
A222_04306	transcriptional regulator, AraC family	
A222_04315	DNA-binding transcriptional regulator, LysR family	
A222_04336	DNA-binding transcriptional regulator, LysR family	

A222_04343	AraC-type DNA-binding protein	
A222_04344	DNA-binding transcriptional regulator, LysR family	
A222_04373	RNA polymerase sigma-70 factor, ECF subfamily	
A222_04394	transcriptional regulator, AraC family	
A222_04402	two component transcriptional regulator, LuxR family	
A222_04450	DNA-binding transcriptional regulator, IclR family	
A222_04463	transcriptional regulator, ArsR family	
A222_04473	LysR family transcriptional regulator, chromosome initiation inhibitor	
A222_04491	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain	
A222_04506	methylated-DNA-protein-cysteine methyltransferase related protein	
A222_04510	response regulator receiver modulated diguanylate cyclase	
A222_04552	transcriptional regulator, AraC family with amidase-like domain	
A222_04580	RNA polymerase, sigma 54 subunit, RpoN/SigL	
A222_04611	two-component system, response regulator RegA	
A222_04617	transcriptional regulator, XRE family with cupin sensor	
A222_04626	transcriptional regulator, AsnC family	
A222_04700	transcriptional regulatory protein RtcR	
A222_04786	DNA-binding transcriptional regulator, MerR family	

A222_04841	Transcriptional regulator of competence genes, TfoX/Sxy family	
A222_04889	NusA antitermination factor	
A222_04900	transcription elongation factor GreA	
A222_04915	transcriptional regulator, GntR family	
A222_04922	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain	
A222_04924	Cu(I)-responsive transcriptional regulator	
A222_04931	transcriptional regulator, AsnC family	
A222_04934	transcriptional regulator, AraC family	
A222_04952	DNA-binding transcriptional regulator, CsgD family	
A222_04980	transcriptional regulator, TetR family	
A222_04993	response regulator receiver modulated diguanylate cyclase	
A222_05003	DNA-binding protein Fis	
A222_05028	Predicted transcriptional regulator YdeE, contains AraC-type DNA-binding domain	
A222_05035	two-component system, OmpR family, copper resistance phosphate regulon response regulator CusR	
A222_05040	transcriptional regulator, TetR family	
A222_05046	RNA polymerase sigma-70 factor, ECF subfamily	
A222_05053	DNA-binding transcriptional regulator, LysR family	
A222_05057	transcriptional regulator, GntR family	
A222_05065	DNA-binding transcriptional regulator, LysR family	

A222_05088	RNAse R	
A222_05137	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain	
A222_05138	transcriptional regulator, TetR family	
A222_05143	DNA-binding transcriptional regulator, LysR family	
A222_05184	transcriptional regulator, LysR family	
A222_05187	transcriptional regulator, AraC family	
A222_05217	transcriptional regulator, TetR family	
A222_05243	DNA-binding transcriptional regulator, LysR family	
A222_05262	transcriptional regulator, GntR family	
A222_05273	DNA-binding transcriptional regulator, MerR family	
A222_05317	DNA-binding transcriptional regulator, MarR family	
A222_05340	DNA-binding transcriptional regulator, LysR family	
A222_05351	transcriptional regulator, LysR family	
A222_05362	Two-component response regulator OmpR	
A222_05363	uncharacterized protein	
A222_05381	DNA-binding transcriptional regulator, LysR family	
A222_05404	transcription termination factor Rho	
A222_05422	regulator of sigma D	
A222_05428	two component transcriptional regulator, LytTR family /Two-component response regulator AlgR	
A222_05439	regulator of nucleoside diphosphate kinase	

A222_05449	transcriptional regulator, GntR family	
A222_05459	DNA-binding transcriptional regulator, LysR family	
A222_05474	transcriptional regulator, AsnC family	
A222_05489	transcriptional regulator, AraC family	
A222_05502	DNA-directed RNA polymerase subunit omega	
A222_05503	guanosine-3',5'-bis(diphosphate) 3'-pyrophosphohydrolase	
A222_05507	transcriptional regulator, AraC family	
A222_05509	transcriptional regulator, LysR family	
A222_05521	transcriptional regulator, GntR family	
A222_05525	two component transcriptional regulator, winged helix family	
A222_05545	transcriptional regulator, TetR family	
A222_05551	transcriptional regulator, AraC family with amidase-like domain	
A222_05553	DNA-binding transcriptional regulator, LysR family	
A222_05560	transcriptional regulator, AraC family with amidase-like domain	
A222_05574	putative transcriptional regulator	
A222_05600	DNA-binding transcriptional regulator, LysR family	
A222_05603	transcriptional regulator, GntR family	
A222_05609	DNA-binding transcriptional regulator, LysR family	
A222_05611	transcriptional regulator, RpiR family	
A222_05680	transcriptional regulator, RpiR family	

A222_05699	transcriptional regulator, GntR family	
A222_05724	transcriptional regulator, DeoR family	

4.14: Gene expression analysis

P. aeruginosa N002 and *Enterobacter* sp. RC4 bacteria genome only four genes *fadB, xylX*, *mdlC* and *xylL* are selected for validation of function in crude oil stress condition. The qRT-PCR analysis of control against crude oil treated condition confirms that the respective genes are involved in crude oil degradation (Table 4.12).

Table 4.12: Validation of crude oil degrading genes of *P. aeruginosa* N002 and *Enterobacter* sp. RC4 through qRT-PCR

Locus tag	Primer for qRT-PCR		Expression Status	Gene name	Protein name or gens function	Log2 Fold Change	
***P. aeruginosa* N002 / *Enterobacter* sp. RC4**	**5'-3' sequence**	**Tm (˚C)**				**RNA-seq**	**qRT-PCR**
A222_01965/ C4L14_09380	GAGAACCCGAAGGTCAAGC	63.8	Up-regulated	*fadB*	Fatty acid oxidation complex subunit alpha	12	0.05
	GATGGAGATGGTGGAGGTGT	63.9					
A222_02507	CAAGCTGCTCAAGGTCAAGG	65	Down-regulated	*xylX*	Toluate 1,2-dioxygenase alpha subunit	-3.43	-2.26
	GAACAGGAAGCCACGGTAGG	65.8					
A222_05052	TGAAGTGGAGCCACGAAC	62.7	Down-regulated	*mdlC*	Benzoylformate decarboxylase	-4.13	-1.05
	GTAGGGAATCGACAGGTACAC	59.9					
A222_02510	CGAAGCCCTACCAGCACTAC	63.7	Down-regulated	*xylL*	Cis-1,2-dihydroxycyclohexa-3,4-diene carboxylate dehydrogenase	-4.85	-3.77
	CAACCGACGAGACGTTGAC	64.5					
RpoD is Bacterial housekeeping genes for transcription	GGGCGAAGAAGGAAATGGTC	66.8					
	CAGGTGGCGTAGGTGGAGAA	67.4					

4.15: Validation of improvement of soil functional characteristics under small field condition

Soil sample phycochemical property is analysis before applying bacterial culture there property and after applying bacterial culture. The Urease activity was measured by recording the absorbance of release of ammonium nitrate from 1.0 g of soil, Phosphatase activity of soil was determined by measuring the release of p-nitrophenol from p-nitrophenyl phosphate and Dehydrogenase activity was determined by measuring the iodo-nitro tetrazolium formamide released from 2,3,5-Triphenyl chloride (TTC) which are showing in the table 4.13.

Table 4.13: Soil improvement and functional characteristics of *P. aeruginosa* N002 and *Enterobacter* sp. RC4

Soil treated with *P. aeruginosa* N002	**Control**	**After treatment**
Soil moisture content	13.78	14.68
Soil Respiration (gmL^2)	0.81	1.07
Soil evaporation ($gCO^2\ m^{-2}\ h^{-1}$)	13.6	13.5
Soil Cation Exchange Capacity (CEC)	127.3	124.41
Soil pH	5.03	6.05
Bulk density (gm/cm^3)	1.05	1.11
Urease (NH_4+-N mg per g dry soil per 3 hrs)	11.52	10.47
Phosphatase (Mole of PNP released per g dry soil per hrs)	1.17	0.72
Dehadrogenase (µg TPF g^{-1} dry soil $24h^{-1}$)	36.11	12.41
Soil treated with *Enterobacter* sp. RC4	**Control**	**After treatment**
Soil moisture content	13.78	15.26
Soil Respiration (gmL2)	0.43	0.9
Soil evaporation (gCO2 m-2 h-1)	15.3	12.8
Soil Cation Exchange Capacity (CEC)	127.3	122.11

Soil pH	5.03	6.95
Bulk density	1.05	1.89
Urease (NH_4+-N mg per g dry soil per 3 hrs)	11.52	10.41
Phosphatase (Mole of PNP released per g dry soil per hrs)	1.17	0.72
Dehadrogenase (µg TPF g^{-1} dry soil $24h^{-1}$)	36.11	33.81

CHAPTER - 5

Discussion

Crude oil contamination in soil and water bodies is a major threat to the pristine ecological condition and is reported at regular intervals across the globe. Contamination occurs during exploration, abandonment of drill sites, accidental spillage from production units, refining processes and transportation (Hollinger et al., 1997; Sarma et al., 2004; Ulric et al., 2008; Adipah et al., 2019). Crude oil are considered the main energy source and materials for different industries (Varjaniand Upasani, 2016) and those industries generate huge quantities of oily sludge containing various hydrocarbons and other recalcitrant compounds which may lead to severe environmentalpollution due to its wide distribution, persistence andtoxic nature (Hu et al., 2013 and Fuentes et al., 2015).Crude oil is mainly composed of a complex mixture of n-alkanes, aromatic hydrocarbons and polar fractions with hetero-atoms of nitrogen, sulfur, oxygen, and asphaltenes (Ekanemet al., 2011) due to itsnon-degradability, the contaminants remain in the environment for long time and affect the physical, chemical and biochemical behavior of the contaminated soil/water (Head et al., 2006). Further, because of xenobiotc nature of crude oil and oil products, their difficult degradation results continued environmental threats which require utmost attention for remediation (Johnsen et al., 2007). Remediation of crude oil contaminated environment employs different processes like physical, chemical and biological. Physical and chemical

processes are costly, not eco-friendly and requires extensive site restoration. In contrast to these, bioremediation processes are being preferred presently due to their cost effectiveness, less stress to environment and complete recovery of contaminated sites (Lovley et al., 2003; Vila et al., 2001). In bioremediation, free living or immobilized bacteria having high crude oil degradation ability are used individually or in consortium. Addition of crude oil degrading fungi or algae can also be done. In many cases bioremediation is coupled with phyto-remediation employing plant species highly tolerant to crude oil and other hydrocarbon contaminants. However, bioremediation efficiency is limited due to complex and toxic nature of crude oil and intricacies associated with adaptation and survival of microorganisms under adverse crude oil contaminated environment (Ward et al., 1978; Lovley et al., 2003; Kleinsteuber et al., 2006). However, in the recent past, screening of different crude oil utilizing bacteria from different ecological niche and their utilization in clean-up of oil contaminated environment have been reported (Ward et al., 1978; Sarma et al., 2004). Several research work have reported the crude oil degradation potential of different bacterial strains belonging to the genus *Lysinibacillus*, *Brevibacillus*, *Bacillus*, *Paenibacillus*, *Stenotrophomonas*, *Alcaligenes*, *Delftia*, *Achromobacter* and *Pseudomonas* (Abe et al., 2003; Margesin et al., 2003; Hamamura et al., 2006; Hazen et al., 2010; Yenn et al., 2014; Sarma et al., 2013; Sarma et al., 2014).

Present study was focused on the genome characters of two bacteria i.e., *P. aeruginosa* N002 and *Enterobacter* sp. RC4 and their role in restoration of oil degraded soils from two oil contaminated fields of Duliajan and Galeky of Assam in North-East India.

5.1: The whole genome sequencing

Whole genome sequence revealed that both the bacteria are having single circular chromosome. In genome N002 having total sequence length 65,37,648 bp contains 5764 annotated genes with 100% total genome coverage and genome RC4 having total sequence length 5,029,294 bp contain 4,985annotated genes with 100% total genome coverage. The assembled genome *P. aeruginosa* N002 contains 11,038 open reading frames (ORFs) and encodes 5629 protein coding genes (CDS), 135 RNA genes, 63 tRNA genes, 12 rRNA genes and G+C content present is 62.36 another genome *Enterobacter* sp. RC4 contains 10,021 open reading frames (ORFs) and encodes 4895 protein coding genes (CDS), 168 RNA genes, 74tRNA genes, 8rRNA genes and G+C content present is 54.77 (Table 4.1). On the basis of cluster of orthologous groups (COGs) circular chromosome map showing 26 different category CDS of metabolic function (Figure 4.2.A and 4.2.B).

5.2: COG analysis of *P. aeruginosa* N002 and *Enterobacter* sp. RC4 withoil degrading bacteria

A one-sample t-test was used to evaluate possible significant differences of the gene abundances of each COG categories between *P. aeruginosa* N002 and other hydrocarbon degrading genomes in the IMG genome database and *Enterobacter* sp. RC4 and other hydrocarbon degrading genomes in the IMG genome database (Table 4.2).

The strain N002 showed significant ($P< 0.001$) higher abundance of genes related to aminoacid transport and metabolism (E, 9.335%), cell motility (N, 2.92%), inorganic ion transport and metabolism (P, 5.69), intracellular trafficking, secretion and vesicular transport (U, 3.24%), posttranslational modification, protein turnover, chaperones (O, 3.8%), signal transduction mechanisms (T, 6.56%), and also the genes with unknown functions (S, 10.04%) compared to that of hydrocarbon degrading bacteria (Binnewies et al., 2006;

Callaghan et al., 2012) . Moreover, abundances of genes belonging to chromatin structure and dynamics (B), RNA processing and modification (A) categories in N002 were similar compared to other hydrocarbon degrading strains. In contrary to these, abundance of genes related to carbohydrate transport andmetabolism (G, 4.31%), cell cycle control, cell division, chromosome partitioning (D, 0.67%), chromatin structure and dynamics, coenzyme transport and metabolisms (H), defense mechanisms (V), energy production and conversion (C), general functions predictions (R), lipid transport andmetabolism(I), nucleotide transportmetabolism(F), replication, recombination and repair (L), secondary metabolites biosynthesis, transport and catabolism (Q), translation, ribosomal structure and biogenesis (J), are significantly lower ($P< 0.001$). Another strain RC4 showed significant ($P< 0.001$) higher abundance of genes related to amino acid transport and metabolism (E, 8.8%), cell motility (N, 1.88%), inorganic ion transport and metabolism(P, 5.008), intracellular trafficking, secretion and vesicular transport (U, 2.711%), posttranslational modification, protein turnover, chaperones (O, 3.59%), signal transduction mechanisms (T, 5.36%), and also the genes with unknown functions (S, 5.24%) compared to that of hydrocarbon degrading bacteria (Myers et al. 2000; Hacke et al. 2000; Binnewies et al. 2006; Callaghan et al. 2012). Moreover, abundances of genes belonging to chromatin structure and dynamics (B), RNA processing and modification (A) categories in RC4 were similar compared to other hydrocarbon degrading strains. In contrary to these, abundance of genes related to carbohydrate transport andmetabolism (G, 9.52%), cell cycle control, cell division, chromosome partitioning (D, 1.03%), chromatin structure and dynamics, coenzyme transport and metabolisms (H), defense mechanisms (V), energy production and conversion (C), general functions predictions (R), lipid transport and metabolism (I), nucleotide

transportmetabolism (F), replication, recombination and repair (L), secondary metabolites biosynthesis, transport and catabolism (Q), translation, ribosomal structure and biogenesis (J), are significantly lower ($P< 0.001$).

4.3: COG analysis *P. aeruginosa* N002 and *Enterobacter* sp. RC4 within same species

A one-sample t-test was used to evaluate possible significant differences of the gene abundances of each COG categories between *P. aeruginosa* N002 and other same species genomes in the IMG genome database and *Enterobacter* sp. RC4 and other same species genomes in the IMG genome database (Table 4.3).

The *P. aeruginosa* N002 and other same species genomes in the IMG genome database results showed that the abundances of amino acid transport and metabolism (E, 9.35%), carbohydrate transport and metabolism (G, 4.31%), cell motility (N, 2.92%), coenzyme transport and metabolism (H, 4.03%), energy production and conversion (C, 6.3%), functions unknown (S, 10.04%), general function prediction (11.62%) lipid transport and metabolism (I, 4.43%), nucleotide transport and metabolism (F, 2.04%) post translation modification, protein turnover, chaperons (O, 3.8%), RNA processing andmodification (A, 0.06%), secondary metabolites biosynthesis, transport and catabolism (Q, 3.09%), signal transduction mechanisms (T, 6.56%) were appreciably higher than the average values. However, the genes responsible for cell cycle regulation, cell division, chromosome partitioning (D, 0.67%), defense mechanisms (V, 1.41), inorganic ion transport andmetabolism (P, 5.69%), intracellular trafficking, secretion, and vesicular transport (U, 3.24%), replication, recombination and repair (L, 2.42%), transcription (K, 9.27), translation, ribosomal structure and biogenesis (J, 3.8%) were found comparatively ($P < 0.001$) lower than the average levels (Table 4.3).

The*Enterobacter* sp. RC4 and other same species genomes in the IMG genome database results showed that the abundances of amino acid transport and metabolism (E, 8.8%), carbohydrate transport and metabolism (G, 9.52%), cell motility (N, 3.95%), coenzyme transport and metabolism (H, 4.9%), energy production and conversion (C, 5.4%), functions unknown (S, 5.24%), general function prediction (R 7.05%) lipid transport and metabolism (I, 3.16%), nucleotide transport and metabolism (F, 2.316%) post translation modification, protein turnover, chaperons (O, 4.049%), RNA processing andmodification (A, 0.036%), secondary metabolites biosynthesis, transport and catabolism (Q, 2.086%), signal transduction mechanisms (T, 5.531%) were appreciably higher than the average values. However, the genes responsible for cell cycle regulation, cell division, chromosome partitioning (D, 1.017%), defense mechanisms(V, 2.34), inorganic ion transport andmetabolism (P, 6.22%), intracellular trafficking, secretion, and vesicular transport (U, 2.18%), replication, recombination and repair (L, 3.338%), transcription (K, 8.48), translation, ribosomal structure and biogenesis (J,5.83%) were found comparatively ($P < 0.001$) lower than the average levels (Table 4.3).

Analysis of the abundance of the COG categories of *Polymorphumgilvum* SL003B-26A1T, *Acinetobacterbaylyi* ADP1, *Acinetobacter lwoffii* SH145, *Alcanivorax borkumensis* SK2, *Bacillus thuringiensis* Bt407, *Burkholderia cepacia* 383, *Geobacillus thermodenitrificans* NG80-2, *Gordonia bronchialis* DSM 43247, *P. aeruginosa* PAO1, *Pseudomonas fluorescens* Pf-5, *Rhodococcus jostii* RHA1, *Desulfococcus oleovorans* Hxd3, *Desulfatibacillum alkenivorans* AK-01, *Marino bacteralgicola* DG893, *Mycobacterium bovis* AF2122/97, *Nocardia farcinica* IFM 10152, *Paracoccus denitrificans* PD1222 and

Xylella fastidiosa 9a5c compared with *P. aeruginosa* N002 and *Enterobacter* sp. RC4 revealed varied similarities among these hydrocarbon degrading strains (Table 4.3).

The toxicity and scarce availability of oil components as carbonsources could be the driving forces for these bacteria to evolve sensingand response systems to avoid damage by hydrocarbons and pursueco-metabolic pathways. In the genomes of N002, RC4, SL003B-26A1T, SK2, NG80-2 and AK-01, the abundances of protein categories responsible for carbohydrate transport and metabolism are lower than the average value of other hydrocarbon degrading bacteria.

This is in accord with the low carbohydrate availability in the environments where these strains were isolated. Fatty acids are important intermediate products in the alkanedegradation pathway and lipid transport and metabolism aretherefore important for the further degradation of alkanes.

5.4: The genomic island and Insertion sequences of *P. aeruginosa* N002 and *Enterobacter* sp. RC4

The process by which the content and organization of genetic information of a species changes over time is known asgenome evolution. This process includes four forms of changes: point mutations and gene conversions, rearrangements, deletions, and insertions offoreign DNA. Geneloss and acquisition are genomic changes that can rapidly andradically alter the life-style of a bacterium in“quantum leaps” (Groisman, 1996). Genomic islands are clusters of genes within a bacterial genome that appear to have been acquired by a mechanisms ofhorizontal geneflux include mobile genetic elements such asconjugative plasmids, bacteriophages, transposons, insertionelements and genomic islands, as well as the mechanism of recombination of foreign DNA into host DNA (Milkman el al.,

1999;). They were first observed in pathogenic bacteria and were designated pathogenicity islands due to the fact that many of them encoded genes for production of toxins or other pathogenicity factors. It is now recognized that genomic islands are also present in many nonpathogenic bacteria, and can encode a wide variety of traits.These latter mechanisms seem to be the primary forcesby which bacteria genetically adapt to novel environments andby which bacterial populations diverge and form separate, evolutionary distinct species. Acquisition of foreign genes is obviously coupled with gene loss because genome growth is notunlimited. The balance between selective gene acquisition andsecondarily imposed gene loss implies that addition of a foreign gene increases the probability of loss of some residentfunction of lower selective value (Lawrence et al., 1999; Ochman et al., 2000). *P.aeruginosa* N002 having total 19 IS sequence which are found in four IS family *i.e.*, IS3, IS3 ssgr IS407, IS91, IS5. And also 7 IS sequence found in GI-1, GI-9 GI-19, GI-22. Another bacteria *Enterobacter* sp. RC4 having total 25 IS sequence which are found in seventeen IS Family *i.e.* IS4_ssgr_IS4, ISL3, IS3_ssgr_IS407, IS5_ssgr_IS903, IS5_ssgr_IS5, IS3_ssgr_IS3, IS1, ISNCY_ssgr_IS1202, IS481, IS3_ssgr_IS150, ISAs1, Tn3, IS3_ssgr_IS2, IS630, IS3_ssgr_IS51, IS110_ssgr_IS1111 and IS110 (Table 4.4). All the GIs of N002 and RC4 are predicted by SeqWord Sniffer and Island Viewer methods and localization of predicted GIs is shown in figure 4.3.A and 4.3.B.The genome *Enterobacter* sp. RC4 havingtotal 68 genomic island (GI) sequences which cover 6,44,897 bp or 12.82 % of thewhole chromosome and the genome sequence of *P. aeruginosa* N002 is having a total of 40 genomic island (GI) sequence which combination can cover 19,81,504 bp or 30.3% of thewhole chromosome (Table 4.5). Identification and analysis of horizontally transferred genomic islands revealed identical genes available in other known hydrocarbon utilizing

bacteria such as *Polymorphum gilvum* SL003B-26A1 (Nie et al., 2012), *Alcanivorax borkumensis, Oleispira antarctica* etc. The GIs having probable horizontal origins might have a significant role in adapting the bacteria to different abiotic stress, besides conferring antimicrobial resistance and secondary metabolite production. Thismight have been acquired fromother organisms or evolved after transfer in response to the changed environmental condition (crude oil contamination) in its vicinity. GIs associated with adaptation and environmental interest has had substantial impact on bacterial evolution.

5.5. The two components system of *P. aeruginosa* N002 and *Enterobacter* sp. RC4

Two-component systems (TCSs) are the largest family of multi-step signal transduction pathways and its convert chemical or physical stimuli into biological responses such as changes in gene expression. TCSs are serves as a basic stimulus response coupling mechanism to allow organisms to sense and respond to changes in many different environmental conditions (Stock et al., 2006). Two component systems typically consist of a membrane bound histidine kinase that senses a specific environmental stimulus and a corresponding response regulator that mediates the cellular response, mostly through differential expression of target genes (Mascher et al., 2006). Although two-component signaling systems are found in all domains of life, they are most common in bacteria, particularly in Gram-negative and cyanobacteria (Capra et al., 2012). The genome *P. aeruginosa* N002 and *Enterobacter* sp. RC4 having total of 213 and 164 TCSs genes (Table 4.6) and the two component system pathway of N002 and RC4 genome are showing in the figure (Figure 4.4.A to Figure 4.4.R) where we found the genes EnvZ/OmpR for osmolarity sensing, CitA/CitB for supposed to regulate the expression of the genes for citrate fermentation in response to external citrate under anaerobic conditions, DcuS/DcuR for

controlling gene expression in response to C4-Dicarboxylates, LuxR regulatory family is particularly involved in communicate through the synthesis and binding of molecular signals in Quorum-Sensing mechanisms, CheA/CheY for chemotaxis, NtrC/GlnL/NtrY response regulator is the transcriptional activator for nitrogen-regulated promoters, and DesR/DesK for thermo sensing, that genes involved in signal transduction and TCS signifies the strain's ability to regulate cell metabolism to a wide variety of environment (Cai et al., 2002; Rowsell et al., 1995; Aguilar et al., 2001).

5.6: Stress response genes of *P. aeruginosa* N002 and *Enterobacter* sp. RC4

Many genes are involved in bacteria to adapted themselves in different environment condition, the genome *P. aeruginosa* N002 having 10 genes and *Enterobacter* sp. RC4 having 7 genes under the environmental adaptation category plant pathogen interaction pathway genes which are showen in Table 4.7 and Figure 4.5.

P. aeruginosa N002 and *Enterobacter* sp. RC4 having a total 40 and 89 xenobiotic degradation genes (Table 4.8) which are involved in diffrent xenobiotic degradation pathway such as chloroalkane & chloroalkene degradation, naphthalene degradation, benzoate degradation, aminobenzoate degradation, benzoate degradation, xylene degradation, nitrotoluene degradation, styrene degradation, ethylbenzene degradation, dioxin degradation, caprolactam degradation, metabolism of xenobiotics by cytochrome P450, atrazine degradation, fluorobenzoate degradation, vanilate degradation, toluene degradation and naphthalene degradation pathways (Figure 4.6.A to Figure 4.6.N). For aromatic compounds e.g., benzoate and xylene degrading genes include, benzoate 1,2-dioxygenase (A222_02521-02522), 2-polyprenylphenol hydroxylase and related flavodoxinoxido reductases (A222_02523) and dehydrogenases (A222_02524). Aminobenzoate degrading genes include

2-polyprenylphenol hydroxylase and related flavodoxinoxido reductases (A222_02525), acylphosphatases (A222_04091) and alkaline phosphatase (A222_01670). 2-polyprenylphenol hydroxylase (A222_02523, C4L14_13510), 3-hydroxyacyl-CoA dehydrogenase (A222_03417), acetyl-CoA acetyl transferases (A222_02033), benzoate 1,2-dioxygenase (A222_02521) for benzoate degradation. 2-haloalkanoic acid dehalogenase (A222_04245) and alcohol dehydrogenase (A222_03895) for chloroalkane and chloroalkenedegradation.2-haloalkanoic acid dehalogenase (A222_04245) and catechol 1, 2-dioxygenase (A222_02532) for chlorocyclohexane and chlorobenzene degradation. S-(hydroxymethyl) glutathione dehydrogenase/ class III alcohol dehydrogenase (A222_01345) is responsible for naphthalene degradation. Toluene degradation is traced back to catechol 1,2-dioxygenase (A222_02532), dienelactone hydrolase and related enzymes (A222_02328) and succinate dehydrogenase locus (A222_03463-A222_03466) whereas 2-polyprenylphenol hydroxylase (A222_02523, C4L14_15230) and benzoate 1,2-dioxygenase (A222_02521-A222_ 02522) for xylene degradation. Caprolactamdegradationismainly carried out by the gene enoyl-CoA hydratase/carnithineracemase (A222_01532, A222_03208, A222_03288, C4L14_09380). The genes acyl CoA: acetate/3-ketoacid CoA transferase, alpha subunit (A222_00224-A222_00225) and AsptRNAAsn/ Glu-tRNAGlnamidotransferase A subunit and related amidases (A222_04347, A222_04470, A222_00803) along with homogentisate 1,2-dioxygenase locus enzymes (A222_03016- A222_03018) are involved in styrene degradation. Benzoate 1,2-dioxygenase (A222_ 02521-A222_02522) is involved in fluorobenzoate degradation. Eventhough, alkane 1-monooxygenase (EC 1.14.15.3) (alkB, A222_02440)is a part of the core genome, it is located in the region of GI-10. Several putative regulatory genes were found in the genome of N002, related to catabolism of

aromatic compounds. PcaR transcription regulator (A222_00160) was found with β-ketoadipate pathway, which is required for chemotactic response to aromatic compounds (e.g., benzoate degrading pathways). Several LysR-type transcriptional regulators (LTRRs) related to various compounds were identified, such aschloroalkane (A222_01344), chloroalkene (A222_05631), for benzoate degradation (A222_00231), for catechol degradation (A222_02529). GntR family transcriptional regulators found in the strains are associated with aromatic compound degradation, including putative GntR regulator, vanR (A222_05088) related to the regulation of vanilate catabolism. Several AraC-type transcriptional regulators related to various compounds were identified, for benzoate and fluorobenzoate (A222_02520), aminobenzoate (A222_02528). Xenobiotic response element (XRE) transcriptional regulator (A222_00223, C4L14_13570) and IclR-type regulator (A222_03015) both were identified in link with styrene degradation pathways.

The sequence genome *P. aeruginosa* N002 and *Enterobacter* sp. RC4 having total 77 and 102 number of Cell motility and chemotaxis genes (Table 4.9), in the genome (C4L14_21875, C4L14_17305, C4L14_17305, C4L14_21875, C4L14_17305, C4L14_21875, A222_03599, A222_00186) locus found CheA/CheY genes for flagella assembly plays an important role in cell motility and chemotaxis, which could also help bacteria move torelatively close niches and attach to the oil-water interface where the utilization of hydrocarbons as carbon source takes place. A total of 344 Signal transduction genes from *P. aeruginosa* N002 and 240 Signal transduction genes from *Enterobacter* sp. RC4 genomes are founed to involved in diffrent function of Signal transduction mechanism (Table 4.10) and 486 genes in *P. aeruginosa* N002 and 356 numbers of genes in *Enterobacter* sp. RC4 found for transcription function for the genome (Table 4.11).

5.7: Functional genes ameliorates environmental soil stress of *P. aeruginosa* and *Enterobacter* sp. RC4

The mechanism of petroleum hydrocarbon degradation is initially brought about by an oxidative process catalyzed by oxygenases and peroxidases. Degradation pathways convert the organic pollutants into intermediates of the tricarboxylic acid (TCA) cycle. When TCA cycle is down-regulated upon oxygen and nutrient depletion, glyoxylate cycle is an alternative pathway in generating energy for the survival of microorganisms. To validate the crude oil degraded gene, *fadB* (Fatty acid oxidation complex subunit alpha), *xylX* (Toluate1,2-dioxygenase alpha subunit), *mdlC* (Benzoylformate decarboxylase) and *xylL* (Cis-1,2-dihydroxycyclohexa-3,4-diene carboxylate dehydrogenase) differentially expressed genes (DEGs) were selected for quantitative real-time PCR (qRT-PCR) analysis. From KEGG pathway database it was observed that gene *mdlC* is involved in the xenobiotic degradation pathway such as aminobenzoate degradation where it converts alpha-Oxo benzeneacetic acid to benzaldehyde and 4-Hydroxyphenylglyoxylate is converted to 4-Hydroxybenzaldehyde. Gene *xylL* is involved in the xenobiotic pathway biodegradation i.e., degradation of aromatic compounds such as benzoate degradation, where it converts (1R,6S)-1,6-Dihydroxycyclohexa-2,4-diene-1-carboxylate to catechol; fluorobenzoate degradation where it converts 3-Fluorocyclohexadiene-cis,cis-1,2-diol-1-carboxylate to 3-Fluorocatechol; 5-Fluorocyclohexadiene-cis,cis-1,2-diol-1-carboxylate and 4-Fluorocyclohexadiene-cis, cis-1,2-diol-1-carboxylate to 4-Fluorocatechol; xylene degradation where it converts cis-1,2-Dihydroxy-4-methylcyclohexa-3,5-diene-1-carboxylate to 4-Methylcatechol, 1,2-Dihydroxy-6-methylcyclohexa-3,5-dienecarboxylate and 1,6-Dihydroxy-5-methylcyclohexa-2,4-dienecarboxylate to 2,3-Dihydroxytoluene. Gene *xylX* is

involved in the xenobiotic pathway biodegradation where benzoate is converted to (1R,6S)-1,6-Dihydroxycyclohexa-2,4-diene-1-carboxylate in the benzoate degradation pathway; 2-Fluorobenzoate is converted to 2-Fluorocyclohexadiene-cis,cis-1,2-diol-1-carboxylate and 6-Fluorocyclohexadiene-cis,cis-1,2-diol-1-carboxylate, 3-Fluorobenzoate is converted to 3-Fluorocyclohexadiene-cis,cis-1,2-diol-1-carboxylate and 5-Fluorocyclohexadiene-cis,cis-1,2-diol-1-carboxylate, 4-Fluorobenzoate is converted to 4-Fluorocyclohexadiene-cis,cis-1,2-diol-1-carboxylate in the fluorobenzoate degradation pathway; and xylene degradation pathway where Toluate, o-Toluate and m-Methlybenzoate are converted to cis-1,2-Dihydroxy-4-methylcyclohexa-3,5-diene-1-carboxylate, 1,2-Dihydroxy-6-methlycyclohexa-3,5-dienecarboxylate and 1,6-Dihydroxy-5-methylcyclohexa-2,4-dienecarboxylate respectively. Gene *FadB* involved long chain fatty acid degradation in this process, here fadB help to transported across the cell membrane via a transport and acyl-activation mechanism of outer membrane protein (*fadL*) which converts enoyl-CoA to 3-ketoacyl-CoA via 3-hydroxylacyl-CoA through hydration and oxidation. Among the selected genes from the RNA-seq data, *fadB* is up-regulated and *mdlC*, *xylC*and *xylX*are down-regulated. The relative expression levels of the target genes were calculated according to the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). Expression was normalized using the reference gene, *rpoD* reported to be one of the top most stable reference gene (Gomes et al., 2018). A comparison in the normalized gene expression ($\log_2$ fold change) of the selected strains genes *fadB, mdlC*, *xylC* and *xylX* from RNA-seq and qRT-PCR data respectively was made which are as follows: *fadB* (12 and -0.05); *mdlC* (-4.13 and -1.05); *xylL* (-4.85 and -3.77); *xylX* (-3.43 and -2.26). It can be observed that genes *viz.* gene *fadB* was neither considered up-regulated nor down-regulated showing a fold change expression of -0.05; *mdlC*, *xylL* and *xylX* were down-

regulated by 1.05 fold, 3.77 fold and 2.26 fold respectively in response to crude oil treatment when compared with their respected controls (Table 4.12). The qRT-PCR results showed that the expression patterns of the above genes were basically concordant with that of the RNA-seq result. Therefore, we can conclude that overall the qRT-PCR data is consistent with the RNA-seq data, thereby confirming reliability of RNA-seq data. Although the observed fold changes in their expression detected by sequencing did not exactly match those detected by qRT-PCR, which may reflect differences in the sensitivity and specificity between qRT-PCR and high-throughput sequencing technology, the detected expression patterns were mostly consistent for all the selected genes confirming the reliability of the RNA-seq results (Liu et al., 2017 and Yu et al., 2017). From the findings of the selected genes, *fadB* (Fatty acid oxidation complex subunit alpha), *xylX* (Toluate 1, 2-dioxygenase alpha subunit), *mdlC* (Benzoylformate decarboxylase) and *xylL* (Cis-1, 2-dihydroxycyclohexa-3,4-diene carboxylate dehydrogenase) are validated through qRT-PCR showing similar DEG pattern in both approaches are the potential candidate genes for carrying genomics approaches for sustainable enhancement of the crude oil degradation efficiency of selected strain. From the findings of the present study, the genes *viz.mdlC, xylL, xylX and fadB* exhibits similar DEG pattern in both approaches where they showed a decrease in the level of gene expression (downregulation) in response to crude oil treatment. Knocking out of these downregulatory genes may not be necessa as it may affect the degradation pathway of the other existing genes thereby inhibiting the degradation.

5.8: Improvement of soil functional characteristics

Crude oil contaminated soil are improved by using microbial consortium. Such contaminants have an effect on the soil biological properties of soil by modifying the

population of specific microflora and thus affect soil enzymatic activities (Caravaca and Roldan, 2003; Wyszkowski and Wyszkowska 2005; Wyszkowski and Wyszkowska, 2009). Besides, the soil chemistry is equally important in developing a biodegradation potential for contaminated soil (Rogers *et al*., 1993). Soil chemical properties of crude oil contaminated are improved, where soil *p*H improved 5.03 to 6.05 in N002 and 6.95 in RC4, Soil moisture improved 13.78 to 14.68 in N002 and 15.26 in RC4.Microbial enzymatic activity of microbes in contaminated soil was found improvesuch as Dehadrogenase (µg TPF g^{-1} dry soil $24h^{-1}$) activity improved 36.11 to 34.41 in N002 and 33.81 in RC4 (Table 4.13). Before bioremediation significantly least quantity of dehydrogenase, phosphatase and urease activity were estimated in contaminated soil compared with normal soil. Dehydrogenase activity in soil has been used to monitor microbial activity as an index for the total oxidative activity (Alef, 1995). The presence of crude oil in soil alters the population size of beneficial microbes and their enzymatic activities in soil thereby changing the biological nature of crude oil contaminated soil. In addition, it has been realized that microbial population size and soil enzyme activity are considered to be good sign of soil physical condition (Anderson and Domsch, 1990).

CHAPTER - 6

Summary and Conclusions

Crude oil contamination is a worldwide problem, which is growing more serious with economic development and effects are long lasting and remediation is challenging. The history of Oil and Gas Industry in India dates back to 1889 when India's first oil well was drilled near Digboi town of Assam. The Digboi refinery was commissioned in the year 1901 and is currently the oldest operating refinery in India. Since then oil exploration/production and other activities are in operation in the entire eight North-Eastern states of India. Among the North-Eastern states, Assam especially upper Assam produces high amount of crude oil. Exploration of oil, production and contamination of environment is an inevitable process, which deteriorates the environment to a major extent. Crude oil contaminated soil have a serious impact on the environment and human health worldwide. Petroleum hydrocarbons contaminate the environment during transportation, abandonment of drill sites, accidental spillage from production units, coastal oil refining, storage tank failure, flooding, maintenance, offshore oil production, shipping activities and accidental spillage. Crude oil is mainly composed of a complex mixture of n-alkanes, aromatic hydrocarbons and polar fractions with hetero-atoms of nitrogen, sulfur, oxygen, and asphaltenes. Crude oil and petroleum waste derivatives are composed of hydrocarbons and considered pollutants, thereby difficult to treat and remove from the environment. Petroleum oil is an important

strategic resource for which all countries compete fiercely. Indeed, anthropogenic activity is reliant on oil to meet its energy demands, which causes the petroleum industry to flourish. Petroleum hydrocarbons are considered the main energy source and materials for different industries. Petroleum hydrocarbons are a large group of chemicals which causes harmful effects and biodegradation resistance, environmental deterioration and bioaccumulation potential. Petroleum industries generate huge quantities of oily sludge containing various hydrocarbons and other recalcitrant compounds which may lead to severe environmental pollution due to its wide distribution, persistence and toxic nature. It has been estimated that around 8×10^4 to 1×10^7 tons of petroleum hydrocarbons will be globally released per year. In India, more than 28,000 tons of oil containing sludge is generated annually by oil refineries.

The present research work reports the whole genome sequence analysis of *Pseudomonas aeruginosa* strain N002 and *Enterobacter* sp. strain RC4 isolated from crude oil contaminated soil of Duliajan and Gelakey of Assam in North-East India having high crude oil degradation ability. The whole genome of the strain N002 was sequenced by shotgun sequencing using Ion Torrent platform and another strain RC4 was sequenced by shotgun sequencing using Illumina platform and complete sequence analysis of both the bacteria were done. *P. aeruginosa* N002 has single circular chromosome having total sequence length 65,37,648 contains 5764 annotated genes with 100% total genome coverage and G+C content of 62.36% and also it contains 11,038 open reading frames (ORFs) and encodes 5629 protein coding genes (CDS), 135 RNA genes, 63 tRNA genes and 12 rRNA genes and another strain Enterobactersp RC4 has single circular chromosome having total sequence length 5,029,294 contains 4,984 annotated genes with 100% total genome coverage

and G+C content of 54.77% and also it contains 10,021open reading frames (ORFs) and encodes 4,898 protein coding genes (CDS), 186 RNA genes, 74 tRNA genes and 8 rRNA genes. It was found that the strain N002 and RC4 revealed versatility for degradation, emulsification and metabolizing of crude oil. Analysis of cluster of orthologous group (COG) revealed that N002 and RC4 has significantly higher gene abundance for cell motility, lipid transport and metabolism, intracellular trafficking, secretion and vesicular transport, secondary metabolite biosynthesis, transport and catabolism, signal transduction mechanism and transcription than average levels found in other genome sequences of the same bacterial species. However, lower gene abundance for carbohydrate transport and metabolism, replication, recombination and repair, translation, ribosomal structure, biogenesis was observed in N002 and RC4 strain than average levels of other bacterial species.

Crude oil contaminated soil generally is an extremely hostile habitat where various xenobiotic hydrocarbon compounds and heavy metals are persistent. These compounds have toxic effect on microbial activity of the soil. The N002 and RC4 genome analysis revealed the underlying mechanism and degradation pathways for crude oil and some of its major components, besides throwing light upon different bacterial adaptive responses in such harsh environment. Several genes involved in environment stress sensing and response, signal transduction, cell defenses etc. have also been reported. The N002 and RC4 genome carries several laterally acquired regions for survival in hostile environments. Presence of stress-response genes and transporter equipped genome gives N002 and RC4 an advantage for survival in crude oil polluted soil. Genes involved in crude oil degradation pathway and functional genomics characterization of these genes can be further carried out to enhance the

degradation efficiency of microbes at the crude oil contaminated sites. Further by selecting these genes and genome, genetically modified bacteria can be developed through genetic engineering or gene editing technology which is most efficient for bioremediation of crude oil contaminated areas. In this regard, the study is highly significant in contributing bioremediation of crude oil polluted environments and ecological restoration.

References

Abe, H.H., Mori, T.O. (2003). Genetics of polycyclic aromatic hydrocarbon metabolism in diverse aerobic bacteria. *Biosci, Biotechnol Biochem*, 67(2): 225–243.

Aguilar, P.S., Hernandez-Arriaga, A.M., Cybulski, L.E. (2001). Molecular basis of thermosensing: a two-component signal transduction thermometer in *Bacillus subtilis*, *EMBO J*. 20: 1681.

Aguilar, P.S., Hernandez-Arriaga, A.M., Cybulski, L.E., Erazo, A.C., de Mendoza, D. (2001). Molecular basis of thermosensing: a two-component signal transduction thermometer in *Bacillus subtilis*. *EMBO J*, 20: 1681-1697.

Aguilera, F., Méndez, J., Pásaroa, E., Laffona, B. (2010). Review on the effects of exposure to spilled oils on human health. *J. Appl. Toxicol*, 30: 291-301.

Ahmed, N., Dobrindt, U., Hacker, J., Hasnain, S.E. (2008). Genomic fluidity and pathogenic bacteria: applications in diagnostics, epidemiology, and intervention. *Nat Rev Microbiol*. 6: 387–394.

Al-Dous, E.K., George, B., Al-Jaber, M.Y., Wang, H., Salameh, Y.M., Al-Azwani, E.K. and Srinivasa, C. (2011).*De novo* genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). *Nature Biotechnol*, 29: 521–527.

Alef, K. (1995). Dehydrogenase activity. In: Alef, K., Nannipieri, P. (Eds.), Methods in Applied Soil Microbiology and Biochemistry. *Academic Press, New York*. 228–229.

Alikhan, N., Petty, N.K., Ben Zakour, N.L., Beatson. S.A. (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics*, 12: 402.

Allen, T.E., Price, N.D., Joyce, A.R., Palsson, B.O. (2006). Long-range periodic patterns in microbial genomes indicate significant multi-scale chromosomal organization. *PLoS Comput. Biol*, 2: e2.

Al-Mailem, D.M., Sorkhoh, N.A., Al-Awadhi, H., Eliyas, M., Radwan, S.S., (2010). Biodegradation of crude oil and pure hydrocarbons by extreme halophilic archaea from hypersaline coasts of the Arabian Gulf. *Extremophiles,* 14: 321-328.

An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, 475: 348–352. Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B. (2009). Real-time DNA sequencing from single polymerase molecules. *Science*, 323, 133–138.

Anderson, T.H., Domsch, K.H. (1990). Application of ecophysiological quotients (qCO2 and qD) on microbial biomass from soils of different cropping histories. *Soil Biol. Biochem. Bioche*, 25: 393-395.

Annweiler, E., Richnow, H.H, Aetntranikian, G., Hebenbrock, S., Garms, C., Franke, S., Franke,W., Michaelis,W. (2000). Naphthalene degradation and incorporation of naphthalene derived carbon into biomass by the thermophilic *Bacillus thermoleovorans*. *Appl. Environ. Microbiol*. 66:518.

Atlas, R. M. (1981). Microbial degradation of petroleum hydrocarbons: an environmental perspective. *Microbiol Rev*, 45(1): 180–209.

Balba, M. T., Al-Awadhi, N., Al-Daher, R. (1998). Bioremediation of oil contaminated soil: Microbiological methods for feasibility assessment and field evaluation. *J. Microbiol. Meth.* 32: 155-164.

Balba, M. T., Al-Awadhi, N., Al-Daher, R. (1998). Bioremediation of oil contaminated soil: Microbiological methods for feasibility assessment and field evaluation. *J. Microbiol. Meth,* 32: 155-164.

Barnett, M.J., Fisher, R.F., Jones, T., Komp, C., Abola, A.P., Barloy-Hubler, F., Bowser, L., Capela, D., Galibert, F., Gouzy, J. (2001). Nucleotide sequence and predicted functions of the entire *Sinorhizobium. meliloti* pSymA mega plasmid. *Proc. Natl. Acad. Sci. USA,* 8:9883–9888.

Bhatnagar, S., Kumari, R. (2013). Bioremediation: A sustainable tool for environmental management - A review. *Ann. Rev. Res. Biol,* 3: 974-993.

Bihari, Z. (2013). Current Trends in Bioremediation and Biodegradation: Next-Generation Sequencing. *J Bioremed Biodeg*, 4: e138.

Binnewies, T.T., Motro, Y., Hallin, P.F., Lund, O., Dunn, D., La, T., Hampson, D.J., Bellgard, M., Wassenaar, T.M., Ussery, D.W. (2006). Ten years of bacterial genome sequencing: comparative-genomics-based discoveries, *Funct Integr Genom,* 6: 165-181.

Blandin, G., Ozier-Kalogeropoulos, O., Wincker, P., Artiguenave, F. and Dujon, B. (2000). Genomic exploration of the hemiascomycetous yeasts. 16. Candida tropicalis. *FEBS Lett*, 487: 91-94.

Bonetta, L. (2006). Genome sequencing in the fast lane. *Nat. Methods*, 3: 141–147.

Bouchez, M., Blanchet, D., Vandecasteele, J.P. (1995). Degradation of polycyclic aromatic hydrocarbons by pure strains and by defined strain associations: inhibition phenomena and cometabolism. *Appl. Microbiol. Biotechnol*. 43:156-164.

Bridges, E.M., Oldeman, L.R. (1999). Global assessment of human-induced soil degradation. *Arid Soil Res Rehabil,* 13: 319–325.

Cai, S.J., Inouye, M. (2002). EnvZ–OmpR interaction and osmoregulation in Escherichia coli, *J Biol Chem*, 277: 24-55.

Cai, S.J., Inouye, M. (2002). EnvZ–OmpR interaction and osmoregulation in *Escherichia coli*. *J Biol Chem*, 277: 24155-24171.

Callaghan, A.V., Morris, B.E., Pereira, I.A., McInerney, M.J., Austin, R.N., Groves, J.T., Kukor, J.J., Suflita, J.M., Young, L.Y., Zylstra, G.J., Wawrik, B. (2012). The genome sequence of *Desulfatibacillus alkenivorans* AK-01: a blueprint for anaerobic alkane oxidation, *Environ Microbiol*, 14: 101-117.

Capra, E.J., Laub, M.T. (2012). Evolution of two-component signal transduction systems. *Annu Rev Microbiol*. 66: 325–47.

Carlton, J.M., Angiuoli, S.V., Suh, B.B., Kooij, T.W., Pertea, M., Silva, J.C., Ermolaeva, M.D., Allen, J.E., (2002).Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoeliiyoelii*. *Nature*, 419: 512-519.

Casjens, S. (2003). Prophages and bacterial genomics: what have we learned so far? *Mol Microbiol.* **49**: 277-300.

Cerniglia, C.E. (1984). Microbial metabolism of polycyclic aromatic hydrocarbons. *Adv. Appl. Microbiol.* 30:31-71.

Cliften, P.F., Hillier, L.W., Fulton, L., Graves, T., Miner, T., Gish, W.R., Waterston, R.H., Johnston, M., (2001). Surveying *Saccharomyces* genomes to identify functional elements by comparative DNA sequence analysis. *Genome Res*, 11: 1175-86.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E. 3rd, Tekaia, F., Badcock, K., Basham, D. (1998). Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature,* 396: 190–198.

Das, D., Baruah, R., Sarma Roy, A., Singh, A.K., Deka Boruah, H.P., Kalita, J., Bora, T.C., (2015). Complete genome sequence analysis of *Pseudomonas aeruginosa* N002 reveals its genetic adaptation for crude oil degradation. *Genomics,* 105(3): 182-90.

De Maayer, P., Chan, W.Y., Rubagotti, E., Venter, S. N., Toth, I. K., Birch, P. R. J., Coutinho, T.A. (2014). Analysis of the Pantoea Ananatis Pan-Genome reveals factors underlying its ability to colonize and interact with plant, insect and vertebrate hosts. *BMC Genomics*, 15:404.

Deka Boruah, H.P., Kumar, B.S.D. (2002). Biological activity of Secondary metabolites produced by *P. fluorescens* strain RRLJ 008. *Folia Microbiologica,* 47(4): 359-363.

Deka Boruah, H.P., Rabha, B.K., Saikia, N., Kumar, B.S.D. (2003). Fluorescent *Pseudomonas influences* palisade mesophyll development and spatial root development in *Phaseolus vulgaris. Pl Soil*, 256: 291-301.

Deng, W., Li, Y., Vallance, B.A., Finlay, B.B. (2001). Locus of enterocyte effacement from Citrobacter rodentium: sequence analysis and evidence for horizontal transfer among attaching and effacing pathogens. *Infect Immun*, 69: 6323-6335.

Diggle, P.S., Whiteley, M. (2020). Microbe profile: Pseudomonas aeruginosa: opportunistic pathogen and lab rat. *Microbiology*, 166(1): 30–33

Dobrindt, U., Hochhut, B., Hentschel, U., Hacker, J. (2004). Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol,* 2: 414-424.

Drukewitz, S.H., von Reumont, B. M. (2019). The significance of comparative genomics in modern evolutionary venomics. *Front Ecol Evol*, 7: 163.

Eiler, A., Heinrich, F., Bertilsson, S. (2012). Coherent dynamics and association networks among lake bacterioplankton taxa. *ISME J6*, 2:330-342.

Fleischmann, R., Adams, M., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae*. *Science*, 269:496–512.

Frankel, G., Phillips, A.D., Rosenshine, I., Dougan, G., Kaper, J.B., Knutton, S. (1998). Enteropathogenic and enterohaemorrhagic Escherichia coli: more subversive elements. *Mol Microbiol*, 30: 911-921.

Fraser, C. M., Fleischmann, R.D. (1997). Strategies for whole microbial genome sequencing and analysis. *Electrophoresis*, 18:1207–1216.

Gao, H., Wang, Y., Liu, X., Yan, T., Wu, L., Alm, E., Arkin, A., Thompson, D.K., Zhou, J. (2004). Global transcriptome analysis of the heat shock response of Shewanella oneidensis. *J Bacteriol*, 186: 7796-803.

Gilbert, C., Cordaux, R. (2013). Horizontal transfer and evolution of prokaryote transposable elements in eukaryotes. *Genome Biol Evol*, 5: 822-832.

Glick, B.R. (2003). Phytoremediation: synergistic use of plants and bacteria to clean up the environment. *Biotechnol Adv*, 21: 383-393.

Glick, B.R. (2012). Plant Growth-Promoting Bacteria: mechanisms and applications. *Scientifica*, v.2012: 15-35.

Golyshin, P. N., Martins Dos Santos V.A., Kaiser, O., Ferrer, M., Sabirova, Y.S., Lünsdorf, H., Chernikova, T.N., Golyshina, O.V., Yakimov, M.M., Puhler, A., Timmis, K.N.

(2003). Genome sequence completed of Alcanivorax borkumensis, a hydrocarbon-degrading bacterium that plays a global role in oil removal from marine systems. *J Biotechnol*, 106: 215-20.

Gomes, A.E.I., Stuchi, L.P., Siqueira, N.M.G., Henrique, J.B.,Vicentini, R., Ribeiro, M.L., Darrieux, M., Ferraz, L.F.C. (2018). Selection and validation of reference genes for gene expression studies in Klebsiella pneumoniae using reverse transcription quantitative real-time PCR. *Sci Rep*, 8: 9001.

Gresham, D., Dunham M.J., Botstein, D. (2008). Comparing whole genomes using DNA microarrays, *Nat Rev Genet*, 9(4):291-302.

Groisman, E. A., Ochman, H. (1996). Pathogenicity islands: bacterial evolution in quantum leaps. *Cell*, 87: 791–794.

Guzman, P.E., Romeu, A., Garcia-Vallve, S. (2008). Completely sequenced genomes of pathogenic bacteria: A review. *Enferm. Infecc. Microbiol. Clin*, 26:88–89.

Hacker, J., Blum-Oehler, G., Hochhut, B., Dobrindt, U. (2003). The molecular basis of infectious diseases: pathogenicity islands and other mobile genetic elements. A review. *Acta Microbiol Immunol Hung* 50: 321-330.

Hacker, J., Blum-Oehler, G., Muhldorfer, I., Tschape, H. (1997). Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol Microbiol*, 23: 1089-1097.

Hacker, J., Kaper, J.B. (2000). Pathogenecity islands and the evolution of microbes. *Ann Rev Microbiol*, 54: 641–679.

Hamamura, N., Olson, S.H., Ward, D.M., Inskeep, W.P. (2006). Microbial population dynamics associated with crude-oil biodegradation in diverse soils. *Appl Environ Microbiol*, 72: 6316–6324.

Harayama, S., Kishira, H., Kasai, Y. and Shutsubo, K. (1999) Petroleum biodegradation in marine environments. *J. Mol. Microbiol. Biotechnol*, 1: 63-70.

Hazen, T.C., Dubinsky, E.A., DeSantis, T.Z. (2010). Deep-sea oil plume enriches indigenous oil degrading bacteria. *Science*, 330: 306–308.

Head, I.M., Jones, D.M., Larter, S.R. (2003). Biological activity in the deep subsurface and the origin of heavy oil. *Nature*, 426: 344-360.

Head, I.M., Jones, D.M., Roling, W.F.M. (2006). Marine microorganisms make a meal of oil. *Nat. Rev. Microbiol.* 4: 173–182.

Hollinger, C. (1997). Contaminated environments in the subsurface and bioremediation: organic contaminants. *FEMS Microbiol Rev*, 20: 517–523.

Hu, G., Li, J., Zeng, G. (2013). Recent development in the treatment of oily sludge from petroleum industry: a review. *J Hazard Mater*. 261: 470–90.

Jacob, F., Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol*, 3:318–356.

Jeukens, J., Boyle, B., Bianconi, I., Kukavica-Ibrulj, I., Tümmler, B., Bragonzi, A., Levesque, R. C. (2013). Complete genome sequence of persistent cystic fibrosis isolate *Pseudomonas aeruginosa* strain RP73. *Genome Announc*, 1: e00568.

Jie, C., Jing-zhang, C., Man-zhi, T., Zi-tong, G. (2002). Soil degradation: a global problem endangering sustainable development. *J Geographical Sci*, 12: 243-252.

Johnsen, A.R., Schmidt, S., Hybholt, T.K., Henriksen, S., Jacobsen, C.S., Andersen, O. (2007). Strong impact on the polycyclic aromatic hydrocarbon (PAH)-degrading community of a PAH-polluted soil but marginal effect on PAH degradation when priming with bioremediated soil dominated by Mycobacteria. *Appl Environ Microbiol*, 73: 1474–1480.

Jores, J., Rumer, L., and Wieler, L.H. (2004) Impact of the locus of enterocyte effacement pathogenicity island on the evolution of pathogenic Escherichia coli. *Int J Med Microbiol*, 294: 103-113.

Juhas, M., van der Meer, J., Gaillard, M., Harding, R.M., Hood, D.W., Crook, D.W. (2009). Genomic islands: Tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol. Rev*, 33:376–393.

Kersters, K., Ludwig, W., Vancanneyt, M., De-Vos, P.D., Gillis, M., Schleifer, K.H. (1996) Recent changes in the classification of the pseudomonads: an overview. *Syst Appl Microbiol*, 19(4): 465–477.

Kimura, M. A. (1980). Simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol*, 16: 111–120.

Kleinsteuber, S., Riis, V., Fetzer, I., Harms, H., Müller, S. (2006). Population dynamics within a microbial consortium during growth on diesel fuel in saline environments. *Appl Environ Microbiol*, 72: 3531.

Kube, M., Chernikova, T.N., Al-Ramahi, Y., Beloqui, A., Lopez-Cortez, N., Guazzaroni, M. E,, Heipieper, H. J., Klages, S., Kotsyurbenko, O. R., Langer, I. (2013). Genome sequence and functional genomic analysis of the oil-degrading bacterium *Oleispira Antarctica*. *Nat Commun*, 4: 2156-2172.

Kuiper, I., Lagendijk, E., Bloemberg, G. Lugtenberg, B. (2004). Rhizoremediation: A beneficial plant-microbe interaction. *Mol. Plant-Microbe Interact*. 17: 6-15.

Kumar, A., Bisht, B. S., Joshi., V. D., Dhewa, T. (2011). Review on bioremediation of polluted environment: A management tool. *Int. J. Environ. Sci*, 1: 1079-1093.

Lal, B., Khanna, S. (1996). Degradation of crude oil by *Acinetbacter calcoaceticus* and *Alcaligenes odorans. J. Appl. Microbiol,* 81:355-362.

Land, M., Hauser, L., Jun, S.R., Nookaew, I., Leuze, M.R., Ahn, T.H., Karpinets, T., Lund, O., Kora, G., Wassenaar, T., Poudel, S., Ussery, D.W. (2015). Insights from 20 years of bacterial genome sequencing. *Funct Integr Genomics*, 15: 141–161.

Lapierre, P., Gogarten, J. P. (2009). Estimating the size of the bacterial Pan-Genome. *Trends Genet*, 25, 107-110.

Larsson, P., Oyston, P.C., Chain, P., Chu, M.C., Duffield, M., Fuxelius, H.H., Garcia, E., alltorp, G., Johansson, D., Isherwood, K.E., Karp, P.D., Larsson, E., Liu, Y., Michell, S., Prior, J., Prior, R., Malfatti, S., Sjostedt, A., Svensson, K., Thompson, N., Vergez, L., Wagg, J.K., Wren, B.W., Lindler, L.E., Andersson, S.G., Forsman, M., Titball, R.W. (2005). The complete genome sequence of *Francisella tularensis*, the causative agent of tularemia. *Nat Genet*, 37: 153-159.

Laub, M.T., Goulian, M. (2007).Specificity in two-component signal transduction pathways. *Ann Rev Gene*, 41: 121–145.

Lawrence, J. G., Hendrickson, H. (2005). Genome Evolution in Bacteria: Order beneath Chaos. *Curr. Opin. Microbiol*, 8(5): 572-578.

Lawrence, J. G., Roth, J. R. (1999). Genomic flux: genome evolution bygene loss and acquisition, *ASM Press,Washington, D.C*., 263–279.

Lee, S. H., Lee, W. S., Lee, C. H., Kim, J. G. (2008). Degradation of phenanthrene and pyrene in rhizosphere of grasses and legumes. *J. Hazard. Mater*,153: 892-898.

Len, A. P., Edward, M. R. (2003). Comparative genomic tools and databases: providing insights into the human genome. *J Clin Invest*. 111(8): 1099–1106.

Lerat, E., Daubin, V., Moran, N.A. (2003). From gene trees to organismal phylogeny in prokaryotes: The case of the Gamma Proteobacteria. *PLoS Biol*, 1(1): E19.

Lim, M. W., Von Lau, E., Poh, P. E. (2016). A comprehensive guide of remediation technologies for oil contaminated soil- present works and future directions. *Mar Pollut Bull*, 109 (1): 14-45.

Liu, W., Luo, Y., Teng, Y., Li, Z., Ma, L.Q. (2010). Bioremediation of oily sludge-contaminated soil by biostimulating indigenous microbes. *Environ. Geochem. Health*, 32: 23-29.

Liu, X., Shen, B., Du, P., Wang, N., Wang, J., Li, J., Sun, A. (2017). Transcriptomic analysis of the response of Pseudomonas fluorescens to epigallocatechin gallate by RNA-seq. *PLoS One*, 12: e0177938.

Livak, K.J., Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCt Method. *Methods* 25: 402-408.

Loman, N., Pallen, M. (2015). Twenty years of bacterial genome sequencing. *Nat Rev Microbiol*, 13: 787–794.

Lovley, D. R. (2003). Cleaning up with genomics: applying molecular biology to bioremediation. *Nat Rev Microbiol*, 1: 35-44.

Lovley, D.R. (2003). Cleaning up with genomics: applying molecular biology to bioremediation. *Nat Rev Microbiol*, 1: 35–44.

Lugtenberg, B.J.J., De Weger, L.A., Bennett, J.W. (1991). Microbial stimulation of plant growth and protection from disease. *Curr Opin Biotech,* 2: 457-464.

Ma, J, Zhai, G. (2012). Microbial Bioremediation in Omics era: Opportunities and Challenges. *J Bioremed Biodeg*, 3:e120.

Maddocks, S.E., Oyston, P.C. (2008). Structure and function of the LysR type transcriptional regulator (LTTR) family proteins. *Microbiology*, 154: 3609-3627.

Mahillon, J., Chandler, M. (1998). Insertion sequences. *Microbiol Mol Biol Rev*, 62: 725-774.

Mahillon, J., Leonard, C., Chandler, M. (1999). IS elements as constituents of bacterial genomes. *Res Microbiol*,150: 675-687.

Mardis, E.R. (2008). Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet*, 9: 387–402.

Margesin, R., Labbé, D., Schinner, F.W., Greer, C. (2003). Hydrocarbon-degrading microbial populations in contaminated and pristine alpine soils. *J Appl Environ Microbiol*, 69: 3085-3097.

Margesin, R., Schinner, F. (2001).Biodegradation and bioremediation of hydrocarbons in extreme environments.*Appl. Microbiol Biotechnol*, 56: 650-663.

Mascher, T., Helmann, J.D., Unden, G. (2006). Stimulus perception in bacterial signal-transducing histidine kinases, *MMBR*. 70 (4): 910–38.

Mattison, K., Oropeza, R., Byers, N., Kenney, L.J. (2002). A phosphorylation site mutant of OmpR reveals different binding conformations at ompF and ompC. *J Mol Biol,* 315: 497-511.

Maxam, A.M., Gilbert, W. (1977). A new method for sequencing DNA. *PNAS.*74: 560-564.

McClelland, M., Florea, L., Sanderson, K., Clifton, S.W., Parkhill, J., Churcher, C., Dougan, G., Wilson, R.K., Miller, W. (2000).Comparison of the *Escherichia coli* K-12 genome with sampled genomes of a Klebsiella pneumoniae and three salmonella enteric serovars, *Typhimurium*, *Typhi*and *Paratyphi*. *Nucleic Acids Res*, 28: 4974-86.

McGuire, A.M., Hughes, J.D., Church, G.M. (2000).Conservation of DNA regulatory motifs and discovery of new motifs in microbial genomes. *Genome Res,* 10: c744-757.

Medini, D., Donati, C., Tettelin, H., Masignani, V., Rappuoli, R. (2005). The microbial Pan-Genome. *Curr Opin Genet Dev*, 15(6): 589-594.

Milkman, R. (1999). Gene transfer in *Escherichia coli*. *ASM Press,Washington, D.C*., 291–310.

Mira, A., Martín-Cuadrado, A. B., D'Auria, G., Rodríguez-Valera, F. (2010). The bacterial Pan-Genome: A new paradigm in microbiology. *Int. Microbiol*, 13: 45-57.

Moran, N.A. (2002). Microbial minimalism: genome reduction in bacterial pathogens. *Cell*, 108: 583–586.

Morelli, I.S., Del Panno, M. T., De Antoni, G. L., Painceira, M. T. (2005). Laboratory study on the bioremediation of petrochemical sludge-contaminated soil. *Intl Biodeterior. Biodegrad*. 55:271-278.

Mougous, J.D., Cuff, M.E., Raunser, S., Shen, A., Zhou, M., Gifford, C. A., Goodman, A. L., Joachimiak, G., Ordoñez, C.L., Lory, S., Walz, T., Joachimiak, A., Mekalanos, J. J. (2006). A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. *Science*, 312: 1526-1541.

Muffler, A., Bettermann, S., Haushalter, M., Hörlein, A., Neveling, U., Schramm, M., Sorgenfrei, O. (2002). Genomewide transcription profiling of Corynebacterium

glutamicum after heat shock and during growth on acetate and glucose. *J Biotechnol*, 98, 255-68.

Muzzi, A.; Masignani, V.; Rappuoli, R. (2007). The Pan-Genome: Towards a knowledge based discovery of novel targets for vaccines and antibacterials. *Drug Discov Today*, 12: 429-439.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 403: 853–858.

Nataro, J.P., Kaper, J.B. (1998). Diarrheagenic Escherichia coli. *Clin Microbiol Rev*,11: 142-201.

Neilson, A.H., Allard, A.S. (1998). Microbial metabolism of PAHs and heterarenes. (In) The hand book of environmental chemistry ed.A.H. (Neilson) vol. 3, part J, Springer-Verlag, Berlin, Heiderberg. Pp.1-64.

Nichols, T.D., Wolf, D.C., Rogers, H.B., Beyrouty C.A., Reynolds. C.M. (1996). Rhizosphere microbial populations in contaminated soils. *Water Air and Soil Poll*. 95:165-178.

Nie, Y., Tang, Y.Q., Li, Y., Chi, C.Q., Cai, M., Wu, X.L. (2012). The genome sequence of *Polymorphum gilvum* SL003B-26A1T reveals its genetic basis for crude oil degradation and adaptation to the saline soil, PLoS ONE, 7: e31261.

Nwaoguikpe, R. N. (2011). The effect of crude oil spill on the ascorbic acid content of some selected vegetable species: *Spinacea oleraceae, Solanum melongena* and *Talinum triangulare* in an oil polluted soil. *Pakistan J. Nutr*, 10: 274-281.

Ochman, H., Lawrence, J.G., Groisman, E.A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature*, 405: 299–304.

Oeltjen, J.C., Malley, T.M., Muzny, D.M., Miller, W., Gibbs, R.A., Belmont, J.W. (1997). Large-scale comparative sequence analysis of the human and murine Bruton's tyrosine kinase loci reveals conserved regulatory domains. *Genome Res,* 7: 315-29.

Okoh, A. I. (2006). Biodegradation alternative in the cleanup of petroleum hydrocarbon pollutants. *Biotechnol. Mol. Biol. Rev*. 1: 38-50.

Olajire AA, Essien JP (2014) Aerobic Degradation of Petroleum Components by Microbial Consortia. *J Pet Environ Biotechnol* 5: 1000195.

Onwurah, I.N.E. (2007) Crude oil spills in the environment effects and some innovative clean up biotechnologies. *Int J Environ Res,* 1: 94–104.

Ou, H.Y., Chen, L.L., Lonnen, J., Chaudhuri, R.R., Thani, A.B., Smith, R., Garton, N.J., Hinton, J., Pallen, M., Barer, M.R., Rajakumar, K. (2006). A novel strategy for the identification of genomic islands by comparative analysis of the contents and contexts of tRNA sites in closely related bacteria. *Nucleic Acids Res*, 34: e3.

Oudot, J., Merlin, F. X., Pinvidic, P. (1998). Weathering rates of oil components in a bioremediation experiment in estuarine sediments. *Mar Environ Res*, 45(2): 113–125.

Palleroni, N.J. (1993). Pseudomonas classification.*Antonie Van Leeuwenhoek*, 64(3-4): 231-251.

Palleroni, N.J. (2010). The *Pseudomonas* story. *Environ Microbiol,* 12(6): 1377–1383.

Palleroni, N.J., Ballard, R.W., Ralston, E., Doudoroff, M. (1972). Deoxyribonucleic acid homologies among some Pseudomonas species. *J Bacteriol*, 110(1): 1–11.

Palleroni, N.J., Kunisawa, R., Contopoulou, R., Doudoroff, M. (1973).Nucleic acid homologies in the genus Pseudomonas. *Int J Syst Bacteriol*, 23(4): 333–339.

Parkhill, J., Sebaihia, M., Preston, A., Murphy, L.D., Thomson, N., Harris, D.E., Holden, M.T., Churcher, C.M., Bentley, S.D., Mungall, K.L., Cerdeno-Tarraga, A.M., Temple, L., James, K., Harris, B., Quail, M.A., Achtman, M., Atkin, R., Baker, S., Basham, D., Bason, N., Cherevach, I., Chillingworth, T., Collins, M., Cronin, A., Davis, P., Doggett, J., Feltwell, T., Goble, A., Hamlin, N., Hauser, H., Holroyd, S., Jagels, K., Leather, S., Moule, S., Norberczak, H., O'Neil, S., Ormond, D., Price, C., Rabbinowitsch, E., Rutter, S., Sanders, M., Saunders, D., Seeger, K., Sharp, S., Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Unwin, L., Whitehead, S., Barrell, B.G., Maskell, D.J. (2003). Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica. *Nat Genet* 35: 32-40.

Parkhill, J., Wren, B.W., Thomson, N.R., Titball, R.W., Holden, M.T., Prentice, M.B., Sebaihia, M., James, K.D., Churcher, C., Mungall, K.L., Baker, S., Basham, D., Bentley, S.D., Brooks, K., Cerdeno-Tarraga, A.M., Chillingworth, T., Cronin, A., Davies, R.M., Davis, P., Dougan, G., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Karlyshev, A.V., Leather, S., Moule, S., Oyston, P.C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S., Barrell, B.G. (2001). Genome sequence of Yersinia pestis, the causative agent of plague. *Nature* 413: 523-527.

Paudyn, K., Rutter, A., Kerry Rowe, R., Poland, J. (2008). Remediation of hydrocarbon contaminated soils in the Canadian Arctic by land farming. *Cold Regions Sci. Technol*, 53: 102-114.

Pennacchio, L.A., Rubin, E.M. (2001). Genomic strategies to identify mammalian regulatory sequences. *Nat Rev Genet*, 2: 100-109.

Perrin, A. (2002). Comparative genomics identifies the genetic islands that distinguish *Neisseria meningitidis*, the agent of cerebrospinal meningitis, from other *Neisseria* species. *Infect Immunol,* 70: 7063-7072.

Petrosino, J.F., Xiang, Q., Karpathy, S.E., Jiang, H., Yerrapragada, S., Liu, Y., Gioia, J., Hemphill, L., Gonzalez, A., Raghavan, T.M., Uzman, A., Fox, G.E., Highlander, S., Reichard, M., Morton, R.J., Clinkenbeard, K.D., Weinstock, G.M. (2006). Chromosome rearrangement and diversification of *Francisella tularensis* revealed by the type B (OSU18) genome sequence. *J Bacteriol*, 188: 6977-6985.

Pinyakong, O., Habe, H. and Omori, T.J. (2003). The unique aromatic catabolic genes in sphingomonads degrading polycyclic aromatic hydrocarbons (PAHs). *J Gen Appl Microbiol* 49:1-19.

Pradhan, S., Conrad, J., Paterek, J., Srivastava, V. (1998). Potential of phytoremediation for treatment of PAHs in soil at MGP sites. *J. Soil Contam*, 7: 467-480.

Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P., Gu, Y. (2012). A tale of three next generation sequencing platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, 13: 341.

Rabus, R., Kube, M., Heider, J., Beck, A., Heitmann, K., Widdel, F., Reinhardt, R. (2005). The genome sequence of an anaerobic aromaticdegrading denitrifying bacterium, strain EbN1. *Arch Microbiol*, 183:27-36.

Rasko, D., Altherr, M. R., Han, C. S., Ravel, J. (2005). Genomics of the *Bacillus cereus* group of organisms. *FEMS Microbiol. Rev*, 29(2): 303-329.

Rogers, J.A., Tedaldi, D.J., Kavanaugh, M.C. (1993). A screening protocol for bioremediation of contaminated soil. *Env Progress,*12: 147-156.

Rothberg, J.M., Hinz, W., Rearick, T.M., Schultz, J., Mileski, W., Davey, M., Leamon, J.H., Johnson, K., Milgrew, M.J., Edwards, M. (2011).

Rowsell, E.H., Smith, J.M., Wolfe, A., Taylor, B.L. (1995). CheA, CheW, and CheY are required for chemotaxis to oxygen and sugars of the phosphotransferase system in Escherichia coli, *J Bacteriol.* 177: 6-11.

Rowsell, E.H., Smith, J.M., Wolfe, A., Taylor, B.L. (1995). *CheA, CheW,* and *CheY* are required for chemotaxis to oxygen and sugars of the phosphotransferase system in *Escherichia coli*. *J Bacteriol*, 177: 6011-6025.

Roy, A.S., Yenn, R., Singh, A.K., Deka Boruah, H.P., Saikia, N., Deka, M. (2013). Bioremediation of crude oil contaminated tea plantation soil using two *Pseudomonas aeruginosa* strains AS 03 and NA 108. *Afr. J. Biotechnol,*12:2600-2610.

Sanders, J. W.E, Sanders, C.C. (1997). Enterobacter spp.: Pathogens poised to flourish at the turn of the century. *Clin Microbiol Rev,* 10: 220–241.

Sanger, F., Coulson, A.R. (1975). Rapid Method for Determining Sequences in DNA by Primed Synthesis with DNA-Polymerase. *J Mol Biol,* 94: 441-448.

Sanger, F., Nicklen, S., Coulson, A.R., (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci*, 74:5463–5467.

Sarma, A.R., Baruah, R., Borah, M., Singh, A.K., Deka Boruah, H. P., Saikia, N., Deka, M., Dutta, N., Bora, T.C. (2014). Bioremediation potential of native hydrocarbon degrading bacterial strains in crude oil contaminated soil under microcosm study. *Int J Biodeterior Biodegrad*, 94: 79–89.

Sarma, A.R., Baruah, R., Gogoi, D., Borah, M., Singh, A.K., Deka Boruah, H.P. (2013). Draft genome sequence of *Pseudomonasaeruginosa* Strain N002, isolated from crude oil-contaminated soil of Geleky, Assam, India. *Genome Annnoucement*, e00104–00112.

Sarma, A.R., Yenn, R., Singh, A.K., Deka Boruah, H.P., Saikia, N., Deka, M. (2013). Bioremediation of crude oil contaminated tea plantation soil using two *Pseudomonas aeruginosa* strains AS 03 and NA 108. *Afr J Biotechnol*, 12(19): 2600–2610.

Sarma, P.M., Bhattacharya, D., Krishnan, S., Lal, B. (2004).Degradation of polycyclic aromatic hydrocarbon by a newly discovered enteric bacterium, Leclercia adecarboxylata. *J Appl Environ Microbiol,* 70: 3163-3179.

Schut, G. J., Brehm, S. D., Datta, S., Adams, M.W.W. (2003). Whole-genome DNA microarray analysis of a hyperthermophile and an archaeon: Pyrococcus furiosus grown on carbohydrates or peptides. *J Bacteriol*, 185: 3935-47.

Seshadri, R., Adrian, L., Fouts, D.E., Eisen, J.A., Phillippy, A.M., Methe, B.A., Ward, N.L., Nelson, W.C., Deboy, R.T., Khouri, H.M., Kolonay, J.F., Dodson, R.J., Daugherty, S.C., Brinkac, L.M., Sullivan, S.A., Madupu, R., Nelson, K.E., Kang, K.H., Impraim, M., Tran, K., Robinson, J.M., Forberger, H.A., Fraser, C.M., Zinder, S.H., Heidelberg, J.F. (2005). Genome sequence of the PCE-dechlorinating bacterium Dehalococcoides ethenogenes. *Science*, 307:105-8.

Sharifi, M., Sadeghi, Y., Akharpour, M. (2007). Germination and growth of six plant species on contaminated soil with spent oil. *Int. J. Environ. Sci. Technol*, 4:463-470.

Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D., Church, G.M. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. *Science,* 309:1728–1732.

Siguier, P., Filee, J., Chandler, M. (2006a). Insertion sequences in prokaryotic genomes. *Curr Opin Microbiol*, 9: 526-531.

Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., Chandler, M. (2006b). ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res*, 34: D32-36.

Sims, G.K. (2013). Current Trends in Bioremediation and Biodegradation: Stable Isotope Probing. *J Bioremed Biodeg*, 4: e134.

Sivashankari, S., Shanmughavel, P. (2007). Comparative genomics - A perspective *Bioinformation*. 1(9): 376–378.

Snape, I., Riddle, M. J., Stark, J. S., Cole, C. M., King, C. K., Duquesne, S., Gore, D. B. (2001). Management and remediation of contaminated sites at Casey Station, Antarctica. *Polar Rec*, 37: 199-214.

Snipen, L., Almøy, T., Ussery, D. W. (2009) Microbial comparative Pan-Genomics using binomial ixture models. *BMC Genomics*, 10: 385.

Stanier, R.Y., Palleroni, N.J., Doudoroff, M. (1966). The aerobic pseudomonads a taxonomic study. *Microbiol*, 43(2): 159–271.

Stock, A.M., Robinson, V.L., Goudreau, P.N. (2006). Two-component signal transduction. *Annu Rev Biochem*, 69 (1): 183–215.

Stomp, A.M., Han, K.H., Gordon, M.P. (1993).Genetic improvement of tree species for remediation of hazardous wastes. *In Vitro Cell Dev Biol Pl*, 29: 227-232.

Stover, C., Pham, X.Q., Erwin, A.L., Mizoguchi, S. D., Warrener, P., Hickey, M. J., Brinkman, F.S. L., Hufnagle, W. O., Kowalik, D. J., Lagrou, M., Garber, R. L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L. L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R., Smith, K., Spencer, D., Wong, G.K.S., Wu, Z., Paulsen, I. T., Reizer, J., Saier, M. H., Hancock, R.E.W., Lory, S., Olson, M. V. (2000). Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature*, 406: 959–964.

Su, Z., Ning, B., Fang, H., Hong, H., Perkins, R., Tong, W., Shi, L. (2011). Next-generation sequencing and its applications in molecular diagnostics. *Expert Rev. Mol. Diagn*, 11: 333–343.

Suresh, B., Ravishankar, G. A. (2004). Phytoremediation- A Novel and Promising Approach for Environmental Clean-up. *Crit Rev Biotechnol*, 24 (3): 97-124.

Swamy, C.T., Gayathri, D., Devaraja, T.N., Bandekar, M., D'Souza, S.E., Meena, R.M., Ramaiah, N. (2016). Plant growth promoting potential and phylogenetic characteristics of a lichenized nitrogen fixing bacterium *Enterobacter cloacae*. *J Basic Microbiol*. 56(12): 1369-1379.

Syvanen, M. (2012). Evolutionary implications of horizontal gene transfer. *Annu Rev Genet*, 46:341–58.

Szurmant, H., Hoch, J.A. (2010). Interaction fidelity in two-component signaling. *Curr Opi Microbiol*, 13: 190–197.

Tamura,K., Stecher, G., Peterson, D., Filipski, A., Kumar, s. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol*, 30(12): 2725–2729.

Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., Angiuoli, S. V, Crabtree, J., Jones, A. L., Durkin, A.S. (2005). Genome Analysis of Multiple Pathogenic Isolates of Streptococcus Agalactiae: Implications for the Microbial "PanGenome". *Proc Natl Acad Sci USA.*, 102,13950.

Tettelin, H., Riley, D., Cattuto, C., Medini, D. (2008). Comparative genomics: the bacterial pan-genome. *Curr Opin Microbiol*, 11: 472–477.

Tiedje, J. M. (2002). Shewanella-the environmentally versatile genome. *Nat Biotechnol*, 20: 1093-4.

Titball, R.W., Petrosino, J.F. (2007). *Francisella tularensis* genomics and proteomics. *Ann N Y Acad Sci*, 1105: 98-121.

Trapp, S., Köhler, A., Larsen, L., Zambrano, K., Karlson, U. (2001). Phytotoxicity of fresh and weathered diesel and gasoline to willow and poplar trees. *J. Soils Sediments*, 1: 71-76.

Trofimov, S. Y., Rozanova, M. S. (2003). Transformation of soil properties under the impact of oil pollution. *Eurasian Soil Sci*, 36: 82-87.

Ulric, W., Rehm, H. J., Reed, G. (2008).Contaminant soil areas, different countries and contaminant monitoring of contaminants, in Environmental Process II, *Soil Decontamination Biotechnol*, 11: 5–42.

Van Hamme, J., Singh, A., Ward, O. (2003). Recent advances in petroleum microbiology. *Microbiol. Mol. Biol. Rev*, 67: 503-549.

Van Hamme, J.D., Singh, A., Ward, O.P. (2003). Recent advances in petroleum microbiology. *Microbiol Mol Biol Rev*, 67: 503–549.

Van Loon, L.C., Bakker, P.A., Pieterse, C.M. (1998). Systemic resistance induced by rhizosphere bacteria. *Anl Rev Phytopathol,* 36: 453-483.

Vega-Jarquin, C., Dendooven, L., Magana-Plaza, I., Thalasso, F., Ramos-Valdivia, A. (2001). Biotransformation of n-hexadecane by cell suspension cultures of Cinchona robusta and Dioscorea composita. *Environ. Toxicol. Chem.*, 20: 2670-2675.

Vidali, M. (2001) Bioremediation. An overview. *Pure Appl. Chem*, 73: 1163-1172.

Vila, J., Lopez, Z., Sabate, J., Minguillón, C., Solanas, A.M., Grifoll, M. (2001). Identification of a novel metabolite in the degradation of pyrene by *Mycobacterium* sp. Strain AP1: action of isolates on two and three-ring polycyclic aromatic hydrocarbons. *J Appl Environ Microbiol,* 67: 5497-5511.

Vogel, T.M., Simonet, P., Jansson, J.K., Hirsch, P.R., Tiedje, J.M. (2009). Terra Genome: a consortium for the sequencing of a soil metagenome. *Nat Rev Microbiol,* 7: 252-261.

Wang, Z., Gerstein, M., Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, 10:57–63

Ward, D.M., Brock, T.D. (1978). Hydrocarbon biodegradation in hyper saline environments, *Appl Environ Microbiol,* 35: 353.

Wei, L., Liu, Y., Dubchak, I., John, S., Park, J. (2002). Comparative genomics approaches to study organism similarities and differences, *J. Biomed. Inform*, 35:142–150

White, P., Wolf, D., Thoma, G., Reynolds, C. (2006). Phytoremediation of alkylated polycyclic aromatic hydrocarbons in a crude oil-contaminated soil. *Water Air Soil Pollut.* 169: 207-220.

Whiteman, R.H.K., Scheffer, R.J., Strobel, (1987). Factors influencing root formation in dicots by Agrobacterium rhizogenes. *Can J Bot*, 66: 642-644.

Willenbrock, H., Hallin, P.F., Wassenaar, T.M., Ussery, D.W. (2007). Characterization of probiotic *Escherichia coli* isolates with a novel pan-genome microarray, *Genome Biol*, 8, R267.

Wu,W., Yu, F., Zong, Z. (2018). *Enterobacter sichuanensis* sp. nov., recovered from human urine. *Int J Syst Evol Microbiol*, 68(12) : 3922–3927

Wuichet, K., Cantwell, B.J., Zhulin, I.B. (2010). Evolution and phyletic distribution of two-component signal transduction systems. *Curr Opi Microb*, 13: 219–225.

Wyszkowski, M., Wyszkowska, J. (2005). Effect of enzymatic activity of diesel oil contaminated soil on the chemical composition of oat (*Avena sativa* L.) and maize (*Zea mays* L.) *Plant Soil Environ*. 51:360–367.

Wyszkowski, M., Ziolkowska, A. (2009). Effect of compost, bentonite and calcium oxide on content of some macro elements in plants from soil contaminated by petrol and diesel oil. *J Elementol*, 14(2): 405–418.

Xu, M., Fujita, D., Hanagata, N. (2009). Perspectives and challenges of emerging single-molecule DNA sequencing technologies. *Small*, 5: 2638–2649.

Xu, X., Liu, W., Tian, S., Wang, W., Qi, Q., Jiang, P., Gao, X., Li, F., Li, H., Yu, H. (2018). Petroleum hydrocarbon-degrading bacteria for the remediation of oil pollution under aerobic conditions: A perspective analysis. *Front Microbiol*. 9: 2885.

Yenn, R., Bora, M., Deka Boruah, H.P., et al. (2014). Phytoremediation of abandoned crude oilcontaminated drill sites of Assam with the aid of a hydrocarbon-degrading bacterial formulation. *Int J Phys*, 16(9): 909-925.

Yu, L., Ma, J., Niu, Z., Bai, X., Lei, W., Shao, X., Chen, N., Zhou, F., Wan, D. (2017). Tissue-Specific transcriptome analysis reveals multiple responses to salt stress in Populus euphratica seedlings. *Genes*, 8: E372.

List of Publications

1. Dhrubajyoti Das, Reshita Baruah, Abhijit Sarma Roy, Anil Kumar Singh, Hari Prasanna Deka Boruah, Jatin Kalita and Tarun Chandra Bora (2015). Complete genome sequence analysis of *Pseudomonas aeruginosa* N002 reveals its genetic adaptation for crude oil degradation. *Genomics*, 105: 182–190.
2. Dhrubajyoti Das, Channakeshavaiah Chikkaputtaiah, Anil Kumar Singh, Hari Prasanna Deka Boruah, Chitta Ranjan Deb. Draft genome sequence of crude oil degrading *Enterobacter* sp. RC4 isolated from Duliajan region of Assam, India. *J Biotechnol.* (Under Review).
3. Dhrubajyoti Das, Gabriella T Mawlong, Yogita N. Sarki, Channakeshavaiah Chikkaputtaiah, Hari Prasanna Deka Boruah, Chitta Ranjan Deb. Transcriptome-wide analysis of crude oil degrading strains from oil contaminated areas of North-East India. *Gene.* (Under Review).

List of Papers Presented in Conference/Seminars

1. Das, D., Singh, A.K., Deb, C.R and Deka Boruah, H.P. 2017. Genomics of *Pseudomonas aeruginosa* N002 strain and its genetic adaptation in oil pollutedenvironment**.** In:National Seminar on'Advances in Biological Science Research', Department of Botany, Nagaland University, Lumami, Nagaland, February 28-March 01, 2017.
2. Das, D. Chikkaputtaiah, C., Singh, A.K., Deb, C.R. and DekaBoruah, H.P. 2018. Genomics of *Enterobacter* sp. RC4 strain and its genetic adaptation in oil polluted environment. In:The Frontiers in Chemical Biology',MRSI-North East Chapter in CSIR-NEIST, Jorhat, June 26-28, 2018.

****