BOLU ABANT IZZET BAYSAL UNIVERSITY
INSTITUTE OF GRADUATE STUDIES
DEPARTMENT OF BIOLOGY

GENOME MAPPING OF PHOSPHATE SOLUBILIZING BACTERIA USING BIOINFORMATIC ANALYSIS

MASTER OF SCIENCE

UĞUR ÇABUK

BOLU, AUGUST 2020

APPROVAL OF THE THESIS

GENOME MAPPING OF PHOSPHATE SOLUBILIZING BACTERIA USING BIOINFORMATIC ANALYSIS submitted by **Uğur ÇABUK** and defended before the below named jury in partial fulfillment of the requirements for the degree of **Master of Science** in **Department of Biology, Institute of Graduate Studies of Bolu Abant Izzet Baysal University in 25.08.2020** by

Examining Committee Members	**Signature**
Supervisor Prof. Dr. Nusret ZENCIRCI Bolu Abant Izzet Baysal University	.....
Co-Supervisor Assist. Prof. Dr. Ercan Selçuk ÜNLÜ Bolu Abant İzzet Baysal University	.....
Member Prof. Dr. Hakan ÖZKAN Çukurova University	.....
Member Assoc. Prof. Dr. Özlem ÖZBEK Hitit University	.....
Member Assist. Prof. Dr. Murat TELLI Bolu Abant İzzet Baysal University	.....

Prof. Dr. Osman GÖRÜR ……

Director of Institute of Graduate Studies

To my dear family,
Celal ÇABUK and Cemile ÇABUK

DECLARATION

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Uğur ÇABUK

_____

ABSTRACT

GENOME MAPPING OF PHOSPHATE SOLUBILIZING BACTERIA USING BIOINFORMATIC ANALYSIS
MSC THESIS
UĞUR ÇABUK
BOLU ABANT İZZET BAYSAL UNIVERSITY INSTITUTE OF GRADUATE STUDIES
DEPARTMENT OF BIOLOGY
(SUPERVISOR: PROF. DR. NUSRET ZENCİRCİ)
(CO-SUPERVISOR: ASSIST. PROF. DR. ERCAN SELÇUK ÜNLÜ)
BOLU, AUGUST 2020

Pseudomonas species is known as a plant growth-promoting bacteria as well as a good phosphate solubilizer. *Pseudomonas* species is commonly used in the studies, not only in the biotechnological and agricultural fields, but also in the field of medicine. In this study, the whole genome sequencing of the *Pseudomonas* species isolated from the root of the einkorn wheat (*Triticum monococcum spp. monococcum*) in Seben/Bolu was performed using the next generation sequencing (Illumina NextSeq 550) with a total of 2,188,517 paired-end reads, were totally 655,555,100 bases, in 76x depth. Paired-end reads were assembled using *de novo* assembly approach and was created the draft genome of *Pseudomonas spp. ESU1531* with a 252 contig with a GC content of 61.03% and a length of 6,520,191. Annotation of all the genes in the genome was done by PROKKA. According to the PROKKA result, the coding genes in the genome was predicted as 5652 and 4547 of them were functionally assigned. 1 rRNA and 66 tRNA were found in the genome. In addition, phylogenetic of 16s rRNA and multi-locus analyses were done and *Pseudomonas spp. ESU1531* taxonomically classified. The whole genome sequencing and analysis of *Pseudomonas spp. ESU1531* will benefit many different studies. In particular, it will benefit studies for the use of bio fertilizers and/or biocontrol agents in the biotechnological and agricultural fields; on the other hand, it will contribute to future studies in understanding the phosphate solubilizing mechanism of the *Pseudomonas* species. Moreover, *Pseudomonas* spp. *ESU1531* will also be a very important reference in genomic and/or metagenomic studies in the Seben region in the future.

KEYWORDS: *Pseudomonas*, Phosphate solubilizing bacteria, Next generation sequencing, Whole genome sequencing, Phylogenetic Analysis

ÖZET

FOSFAT ÇÖZÜCÜ BAKTERİNİN BİYOENFORMATİK ANALİZLERLE GENOM HARİTASININ ÇIKARILMASI
BİYOLOJİ BÖLÜMÜ
UĞUR ÇABUK
BOLU ABANT İZZET BAYSAL ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİYOLOJİ ANABİLİM DALI
(TEZ DANIŞMANI: PROF. DR. NUSRET ZENCIRCI)
(İKİNCİ DANIŞMAN: DR. ÖĞR. ÜYESİ ERCAN SELÇUK ÜNLÜ)
BOLU, AĞUSTOS - 2020

Pseudomonas türü bitki büyümesini teşvik edici bakteriler olarak bilinmesinin yansıra iyi bir fosfat çözücü olarak da tanımlanmaktadır. *Pseudomonas* türü sadece biyoteknolojik ve tarımsal alanda değil aynı zamanda tıp alanında da yaygın bir şekilde kullanılmaktadır. Bu çalışmada, Seben bölgesinde yetişen Iza buğdayı (*Triticum monococcum spp. monococcum*) bitkisinin kökünden izole edilmiş ve fosfat çözünürlük aktivitesi doğrulanmış *Pseudomonas* türünün tüm genom dizilemesi yeni nesil dizileme (Illumina NextSeq 550) kullanılarak yapılmış ve 76x derinliğinde çift-uç okumalı toplamda 2,188,517 okuma ile 655,555,100 baz elde edilmiştir. Sekans okumaları, *de novo* birleştirme yaklaşımıyla bir araya getirilerek GC içeriği 61.03% ve uzunluğu 6,520,191 olan 252 contig ile *Pseudomonas* spp. *ESU1531* taslak genomu oluşturulmuştur. Tüm genlerin tanımlanması PROKKA ile yapılmıştır. PROKKA sonuçlarına göre, 5652 kodlayan gen tahmin edilmiş ve bunların 4547'si fonksiyonel olarak atanmıştır. Genomda, 1 adet 16s rRNA ve 66 adet tRNA bulunmuştur. Ek olarak, 16s rRNA filogenetik ve çoklu lokus analizi yapılarak izole edilen bakteri taksonomik olarak sınıflandırılmıştır. *Pseudomonas* spp. *ESU1531*'in tüm genom analizi, birçok farklı çalışmaya yarar sağlayabileceği düşünülmektedir. Özellikle, biyoteknolojik ve tarımsal alanda biyogübre veya biyokontrol ajanı olarak kullanılmasına yönelik yapılan çalışmalara, diğer yandan *Pseudomonas* türünün fosfat çözme mekanizmasının anlaşılmasına katkı sağlayacaktır. Bununla birlikte, *Pseudomonas* spp. *ESU1531*, ileride yapılacak Seben bölgesindeki genomik veya metagenomik çalışmalarda da oldukça önemli bir referans olacaktır.

ANAHTAR KELİMELER: *Pseudomonas*, Fosfat çözücü bakteriler, Yeni nesil dizileme, Tüm genom dizileme, Filogenetik Analiz

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

LIST OF ABBREVIATIONS AND SYMBOLS

Al : Alluminum
ATP : Adenosine Triphosphate
BLAST : Basic Local Alignment Search Tool
Ca : Calcium
COG : Cluster of Orthologous Group
ddNTPs : Dideoxy-nucleotides
DNA : Deoxyriboneucleic acid
Fe : Iron
gyrB : DNA gyrase subunit B
ML-based : Maximum likelihood-based
MLSA : Multi-locus Sequence Analysis
NCBI : National Centre for Biotechnology Information
NGS : Next Generation Sequencing
ORF : Open Reading Frame
P : Phosphorus
PCR : Polymerase Chain Reaction
PGPB : Plant Growth-Promoting Bacteria
PROKKA : Rapid Prokaryotic Genome Annotation
PSB : Phosphate Solubilizing Bacteria
RNA : Ribonucleic acid
rpoB : RNA polymerase subunit beta
rpoD : RNA polymerase, sigma 70 (sigma D)
rRNA : Ribosomal RNA
SGS : Second Generation Sequencing
TGS : Third Generation Sequencing

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to:

My main advisor, **Prof. Dr. Nusret ZENCİRCİ**, who showed me how I should be an academic person in the scientific area.

My co-advisor, **Assist. Prof. Dr. Ercan Selçuk ÜNLÜ**, who is the first person to encourage, support and believe me when I wanted to start a master program in bioinformatics. You gave me the opportunity that is going to be a very important basis for my future academic career. I am very grateful for your advice in science and academic career.

My thesis exam committee members: **Prof. Dr Hakan ÖZKAN**, **Assoc. Prof. Dr. Özlem ÖZBEK**, and **Assist. Prof. Dr. Murat TELLİ** for the comments and feedback in the dissertation.

My inspiration, **Dr. Ozan Gönensin BOZDAĞ**, for his mentorship and valuable advice on genome analyses and all contributions. You showed me what and how to do and gained me wide perspectives about the evolution and bioinformatics. I am very glad to meet you and I am very grateful for your kindness and time for me all the time.

Current members of the ESU-Lab: **Ömer Can ÜNÜVAR** for friendship, coffee-breaks in the front of institute. Moreover, you taught me molecular techniques in the laboratory and showed me what we do pre-bioinformatics. **Dr. Gülgez Gökçe YILDIZ** for scientific activities and great dinner. **Yunus ŞAHİN** for discussions about bioinformatics and programming languages.

Members of Zencirci's Lab: **Ferdi AĞIL, Çisem DOĞAN, Mehmet ÖRGEÇ** for friendship, coffee breaks, and lunch times.

My friends that I spent time together during my master study: **İsmail Ahmet SÜRÜCÜ**, my housemate, for spending many hours during computer games and our unforgettable conversations about everything, **Burak BULUT** and **Esma**

DEMİR for friendship and coffee breaks, and my dearest friends, **Begüm Serra, Deniz İdil, Borja, Zeynep** and **Han**, who are in Sweden, and **Deniz AKBULUT**, who is in Germany, for their valuable friendships.

Finally, biggest thanks to my family: **Celal ÇABUK**, **Cemile ÇABUK**, and my sisters, **Hülya** and **Zeliha**, for supporting all the time and believing and encouraging me to keep going in the way of academic career. I could not make it without your supports. I hope, I can make it much more with your supports.

I thank you all from the bottom of my heart!

Uğur ÇABUK

Bolu, August 2020

1. INTRODUCTION

1.1 Importance of Phosphorus for Plant

Phosphorus (P) is one of the most important macronutrient for plant as it plays a vital role in plant growth and development (Alori et al., 2017). It is referred as major nutrient for plant because it is a non-renewable and finite source around the world. P is taken into as orthophosphates $H_2PO_4^-$ and $H2PO4^{-2}$, which is in inorganic form, by plants through root hairs, root tips, and the outermost layers of root cells. It may either be stored in the root or transported to be used for metabolic and cellular activities from root to other parts of plant. While P stored in the root contributes growth and development of lateral roots and fibrous rootlets, P is also effective on flowering, fruiting, seed production, and maturation (Brady and Weil, 2017; Khan and Zaidi, 2009). Therefore, the lack of P may cause seed size, and seed number (Arif et al., 2017).

As P is found in the structure of small and macromolecules, it plays a key role in many metabolic and cellular activities. In terms of existing in their structure, it is found in

- ATP, which is required for energy transfer, substance transport,
- Phospholipids, which is an important compounds for cell membrane,
- DNA, which is genetic material carries genetic information about all the process in organism,
- RNA, which is polymeric molecule and play a role in the gene expression, coding,
- Genes and chromosomes.

Consequently, it plays a key role in metabolic processes that includes transfer of information for generations, signal transduction, photosynthesis, and respiration in plant (Arif et al., 2017; Brady and Weil, 2017; Sharma et al., 2013).

1.1.1 Phosphorus Cycle and Phosphate Availability in Soil

Phosphorus is found in the forms of both organic and inorganic P in soil. While source of organic phosphorus occurs from death and decay of plant, animals, and microorganisms, inorganic phosphorus happens from sedimentary rock. Organic P, component of humus, is mineralized by bacteria so that it can be taken up by plant (Fig. 1.1). Besides, soil microorganisms, microbes, play a role in immobilization to fix it in soil (Brady and Weil, 2017; Khan and Zaidi, 2009).

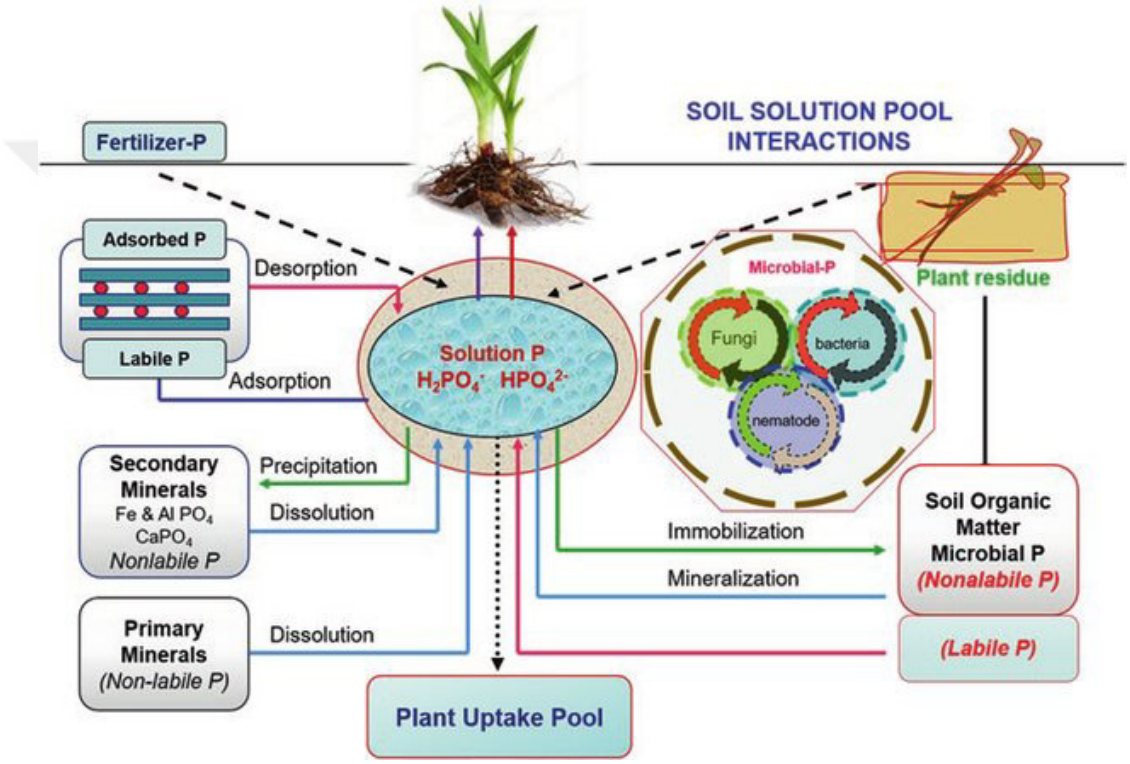

Figure 1.1. Phosphorus Cycle in Soil (Arif et al., 2017)

In soil, total content of phosphorus is 0.05% (w/w) is found as inorganic and organic form. However, only 0.01% of total P is available for plant. Plants require, at least, 30 μmol.L^{-1} P for optimal productivity, however, it was shown that only 1 μmol.L^{-1} P is available in various soil samples due to the solubility issues (Adhya et al., 2015; Sharma et al., 2013). Even though some fertilizers can increase the soluble P content, , P molecules can be turned into insoluble forms due to readily fixed with other metallic elements (Brady and Weil, 2017).

Basically, insoluble P is divided into two as organic and inorganic P (as called Pi). Pi is mainly occurred by apatite mineral, is found in sedimentary rock, and is represented by apatite, hydroxyapatite, and oxyapatite in soil. On the other hand, since P is formed chemical bounds with iron (Fe), aluminum (Al), and calcium (Ca), it transforms to unavailable form for plant uptake and fixed in soil. As these phosphate compounds are dissolved that depend on different pH ranges in soil, plant cannot effectively uptake them (Adhya et al., 2015; Brady and Weil, 2017).

Organic P compounds are generally found as inositol, phosphate mono-esters, phosphate di-esters, and phospholipids in soil. Inositol constitutes ~60% of total organic P in soil. The amount of inositol phosphate depends on type of soil, such as alkaline, acidic. Other compounds including nucleic acids, phospholipids may constitute up to 1-2% of total organic P in soil (Alori et al., 2017; Brady and Weil, 2017; Rodríguez and Fraga, 1999).

To overcome this limiting P deficiency condition, excessive chemical fertilizers are applied to the soil to meet the demand for P-requirement of plant. However, excess of application of P fertilizer is not sustainable for agriculture since it reduces the productivity and disrupts soil fertility and environment. It is demonstrated that over application of P fertilizers cause forcing immobilization of P and eutrophication in ecosystem (Alori et al., 2017; Sharma et al., 2013).

1.2 Plant Growth-Promoting Bacteria (PGPB)

Soil plays host for many living organism which live together and some of them interact symbiotic relationship with plants, the others, however, can be shown pathogenic interaction with plants (Biswas et al., 2016). Thus, PGPB can be divided into three groups which are symbiotic, asymbiotic, and pathogenic interaction with plants. While bacteria in soil also directly promote growth and development of plant by providing nutrient, nitrogen, phosphorus, potassium, take up by plants, they can also indirectly enhance plant growth by preventing phyto-pathogenic organisms through antibiosis, induction of systemic resistance, and competitive exclusion (Gouda et al., 2018; Khan and Zaidi, 2009) .

Bacteria living in rhizosphere are referred as plant growth-promoting rhizobacteria which enhance to plant growth and development or are referred as plant growth-promoting bacteria in the terms of all bacteria in soil in the context of symbiotic relationship with plants. The most widely known bacteria that have a symbiotic relationship with plants are rhizobia. On the other hand, there are free living bacteria in the soil and colonized into the root of plant (Ramakrishna et al., 2019).

Plant Growth-Promoting Rhizobacteria (PGPB) is one of the most important agents for crop productivity as well as ecosystem. They not only enhance plant growth and development via phosphate solubilization, nitrogen fixation, production of indole-3-acetic acid, siderophore, ACC deaminase, but also protects the plants against elements of biotic stress such as pathogenic bacteria, viruses , and insects (Gouda et al., 2018; Khan and Zaidi, 2009; Ramakrishna et al., 2019).

Besides, many studies have shown that there are some bacteria that have some common trait with each other on different plants. On the other hands, there are some that are specific host interaction with specific plant as well as depend on types and region of soil (Khan and Zaidi, 2009; Ramakrishna et al., 2019; Sharma et al., 2013).

1.2.1 Phosphate Solubilizing Bacteria

Phosphate solubilizing bacteria (PSB) are included in PGPB. As aforementioned, many PGPB have many skills that benefit the plant. However, some bacteria are commonly known as PSB since they solve higher phosphate than the others. Generally, these are *Pseudomonas, Bacillus, Rhizobium, Burkholderia, Achromobacter, Agrobacterium, Micrococcus, Aerobacter, Flavobacterium* and *Erwinia* as bacterial genera. Some species has been shown in Table 1.1.

Table 1.1. Some Bacteria known as P solubilizers

Genera	Species
Acetobacter	*Acetobacter sp., A. liquefaciens, A. diazotropicus,*
Azospirillum	*Azospirillum brasilense, A. irakense*
Bacillus	*Bacillus subtilis, B. pumilis*
Burkholderia	*Burkholderia cepecia. B. sp.*
Enterobacter	*Enterobacter aerogenes, E. agglomerans, E. cloacae*
Rhizobium	*Rhizobium sp., R. meliloti, R. loti*
Pseudomonas	*Pseudomonas cepacia, P. putida, P. chlororaphis, P. striata, P. syringae, P. fluorescens, P.alcaligenes,*

1.2.1 Mechanism of Phosphate Solubilization in PSB

In soil, PSB can solubilize inorganic and organic P. The main mechanism to solubilize phosphorus;

1. Production and releasing of organic acid, i.e. acetate, citrate, gluconate, ketogluconate, lactatate, etc.
2. Release of extracellular enzymes like phosphatases, phytases,
3. Release of P during organic matter decomposition.

There are some main enzymes that are nonspecific phosphatases, phytases, and phosphanatases and C-P lyases play key roles in organic phosphorus solubilization in phosphate solubilizing bacteria. Nonspecific enzymes phosphatases play a role in dephosphorylation by breaking phosphoester or phosphoanhydrid bonds of organic compound in soil. Phosphatase activity in soil depends on organic matter, P content, pH, and temperature conditions in soil. Phytases can release P from organic compounds P and directly contribute to take P by plant. Also, C-P lyases perform releasing of P from organic compounds P by breaking C-P bonds in the organic matter. Some genes responsible for organic phosphorous solubilization are *phoD, appA, phnX*, and *phnJ* (Liang et al., 2020; Shrivastava et al., 2018).

On the other hand, mineral phosphate solubilization (MPS) is also performed by PSB. Main organic acids are involved in the MPS are produced by some bacterial species are gluconic acid, malic acid, oxalic acid, citric acid, and formic acid. Glucose dehydrogenase (GDH) is an enzyme that is responsible for the production of gluconic acid is produced by gcd. Moreover, pyrroquinoline quinone (PQQ) is a

cofactor that is required for GDH activity through the phosphate solubilization process. Many studies have focused on PQQ and GDH in the MPS in PBS. Besides, PQQ plays role in the production of antimicrobial substances and utilizing of systemic plant defenses. Therefore, PQQ is not only important for phosphate solubilization process but also for plants' metabolic activities (Meyer et al., 2011; Shrivastava et al., 2018).

Even though there is little knowledge available about the genetic background of the other organic acids in terms of inorganic phosphate mineralization, among the genes mentioned above, *gcd* gene has predominantly found (Liang et al., 2020).

1.3 *Pseudomonas*

Pseudomonas is known as the largest genus within *Gammaproteobacteria*. *Pseudomonas* genus contains more than 220 species with a total of 528 completely finished genomes and 8578 draft genomes that are found in the *Pseudomonas* genome database (Gomila et al., 2015; Lalucat et al., 2020; Winsor et al., 2016).

They can be found in various environmental conditions and diverse interactions with host and environment (Fig. 1.2). They can be pathogenic, free-living and plant-beneficial. Therefore, they are considered as good models comprehensive scientific studies. Most known and studied species is *Pseudomonas aeuriginosa* strain which is a pathogenic bacterium for humans. On the other hand, they are also well-known plant growth-promoting bacteria such as *Pseudomonas fluorescens, Pseudomonas putida, Pseudomonas striata,* etc. (Shrivastava et al., 2018; Silby et al., 2011). They can colonize in the root or leaf of plants. Therefore, they are directly involved in plant growth.

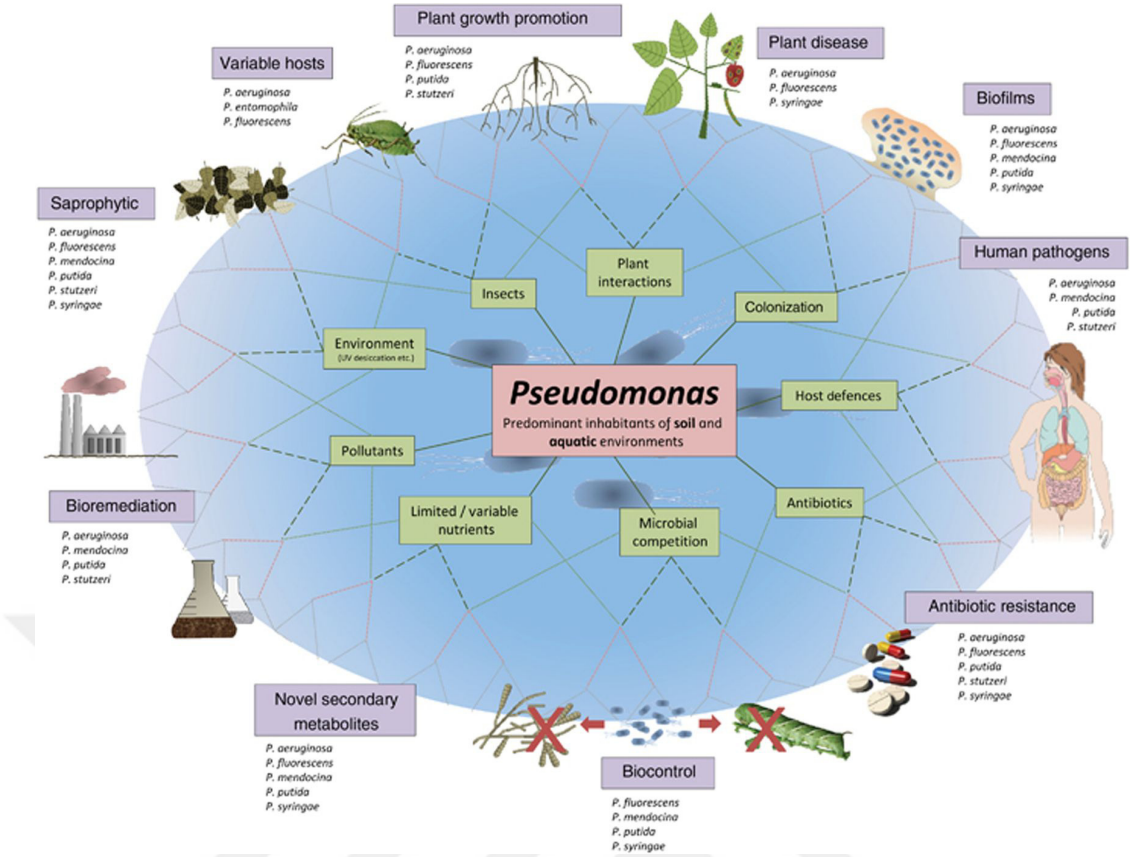

Figure 1.2. Diversity of *Pseudomonas* on the environment (Silby et al., 2011)

On the other hand, as shown in Figure 1.2, since they are also rather effective in bioremediation, it is important for biotechnological studies. For instance, *Pseudomonas putida* is known as plant growth-promoting bacteria as well as it is capable of bioremediation of toxic substances in wastes (Silby et al., 2011). These multi-functional roles of *Pseudomonas* species make them valuable for biotechnological applications as biocontrol agents and bioinoculants.

Since there are many and diverse species in *Pseudomonas,* many phylogenetic studies recently have been done. Taxonomy of some *Pseudomonas* species have been reclassified in the most of these studies in time (Gomila et al., 2015; Hesse et al., 2018; Lalucat et al., 2020). Since genome size of *Pseudomonas* species varies even among strains, phylogenetic studies have been always complex and crucial. Therefore, the other analyzes following 16s rRNA phylogenetic analysis have been also performed as shown in Fig. 1.3.

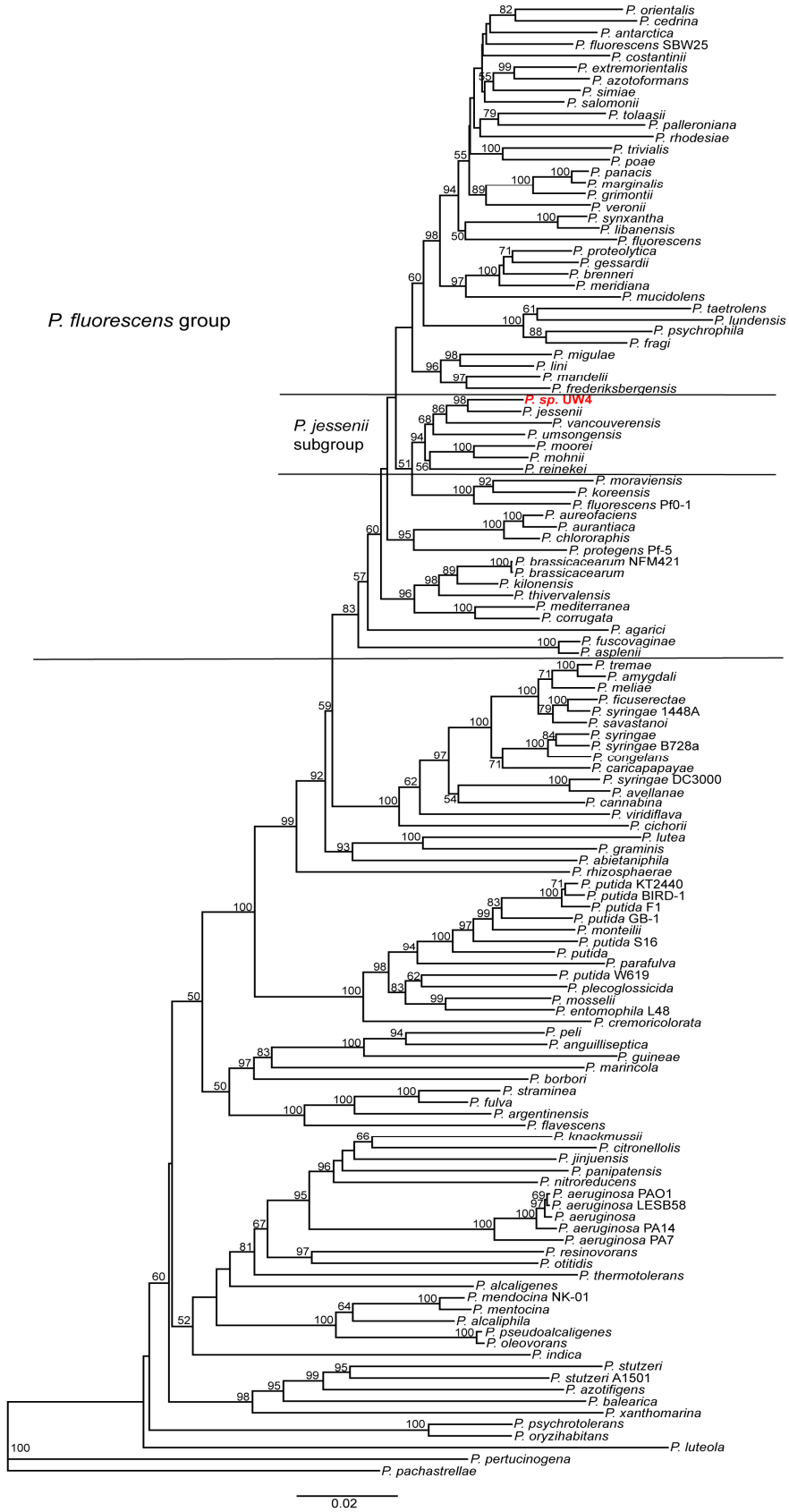

Figure 1.3. Phylogenetic studies of *Pseudomonas* species based on the housekeeping genes (16s rRNA, *gyrB*, *rpoB*, *rpoD*) (Duan et al., 2013)

Pseudomonas sp. *UW4* is a well-known plant growth-promoting bacteria that has been studied deeply (Duan et al., 2013). The further study of this type of *Pseudomonas* species has become a necessity within the scope of benefit for the agricultural and biotechnological studies.

1.4 Sequencing Technologies

Since Watson and his colleagues discovered structure of DNA in 1953 (Watson and Crick, 1953), DNA studies have been accelerated. As DNA contains genetic information for living organisms, researchers have been focused on DNA sequences. Therefore, the first sequencing was achieved by Sanger and his colleagues since they produced the first complete protein-coding gene sequence, the coat protein of bacteriophage MS2, in 1972 (Heather and Chain, 2016). It was called as first generation sequencing that was a beginning of sequencing technologies in the history. Subsequently, second and third generation sequencing were developed, respectively.

1.4.1 First Generation Sequencing

First generation sequencing is divided into two as Sanger and Maxam-Gilbert sequencing. However, Maxam-Gilbert method has been uncommon due to using toxic and radioactive chemicals.

Sanger Sequencing is an enzymatic sequencing known as the chain termination or the dideoxynucleotide method. Basically, it is required denatured template of DNA, primer, DNA polymerase, and dideoxy-nucleotides (dNTPs) to perform reaction. Each base is marked with dNTPs that are used for elongation of nucleotide. DNA polymerase provides to be included the dNTPs into the DNA template. The end of process, DNA fragments occurs depends on their lengths. These DNA fragments are separated based on their lengths using gel electrophoresis and visualized by X-ray or UV light (Kchouk et al., 2017). Illustration is shown in Figure 1.4.

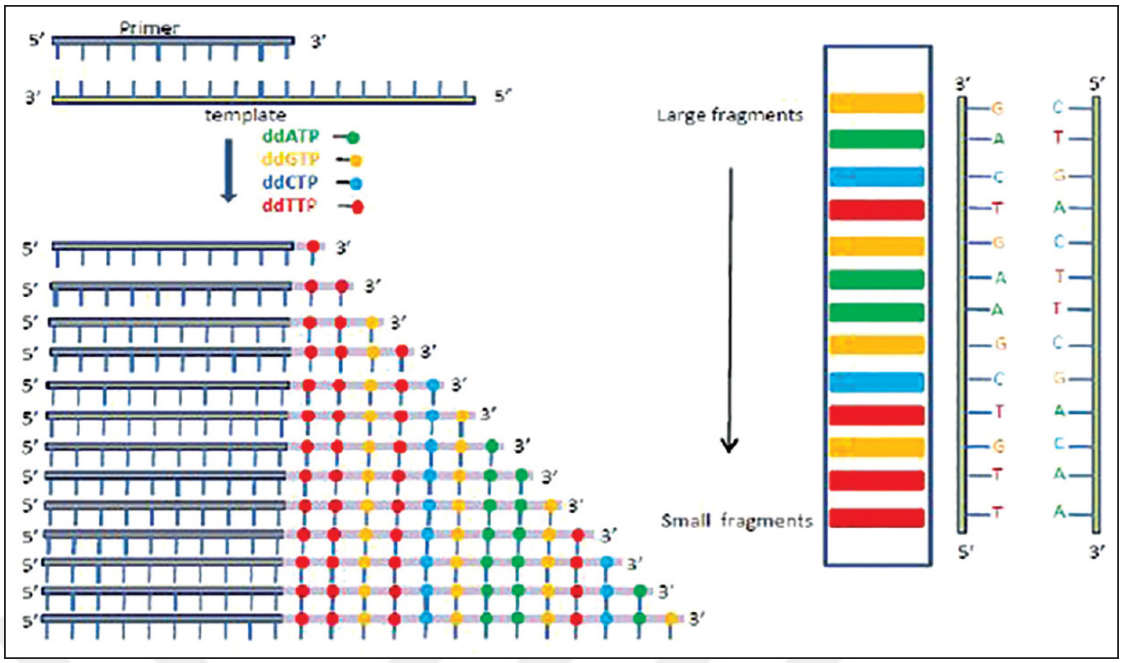

Figure 1.4. Illustration of Sanger Sequencing (Aliyu, 2014)

1.4.2 Second Generation Sequencing

Although Sanger Sequencing is a robust method that gives high accuracy, it has some limitations such as being time-consuming and expensive. Therefore, new sequencing methods, called as second-generation sequencing (SGS) or next generation sequencing (NGS), were developed after 2005. General features of second generation sequencing are

- Faster than first generation sequencing,
- Cost-effective compared to Sanger sequencing
- There is no need gel electrophoresis to detect output.

NGS systems are based on short read sequencing approach that is separated into two groups based on ligation- or synthesis-based techniques. In NGS area, Illumina platform is dominant compared to other platforms such as Roche/454 sequencing, ABI/SOLiD sequencing.

Sequencing-by-synthesis (SBS) has been firstly applied by Roche/454 that was used pyrosequencing technique that has been showed in Figure 1.5. Pyrosequencing technique gives high error rates in homopolymer region, particularly (Heather and Chain, 2016; Kchouk et al., 2017).

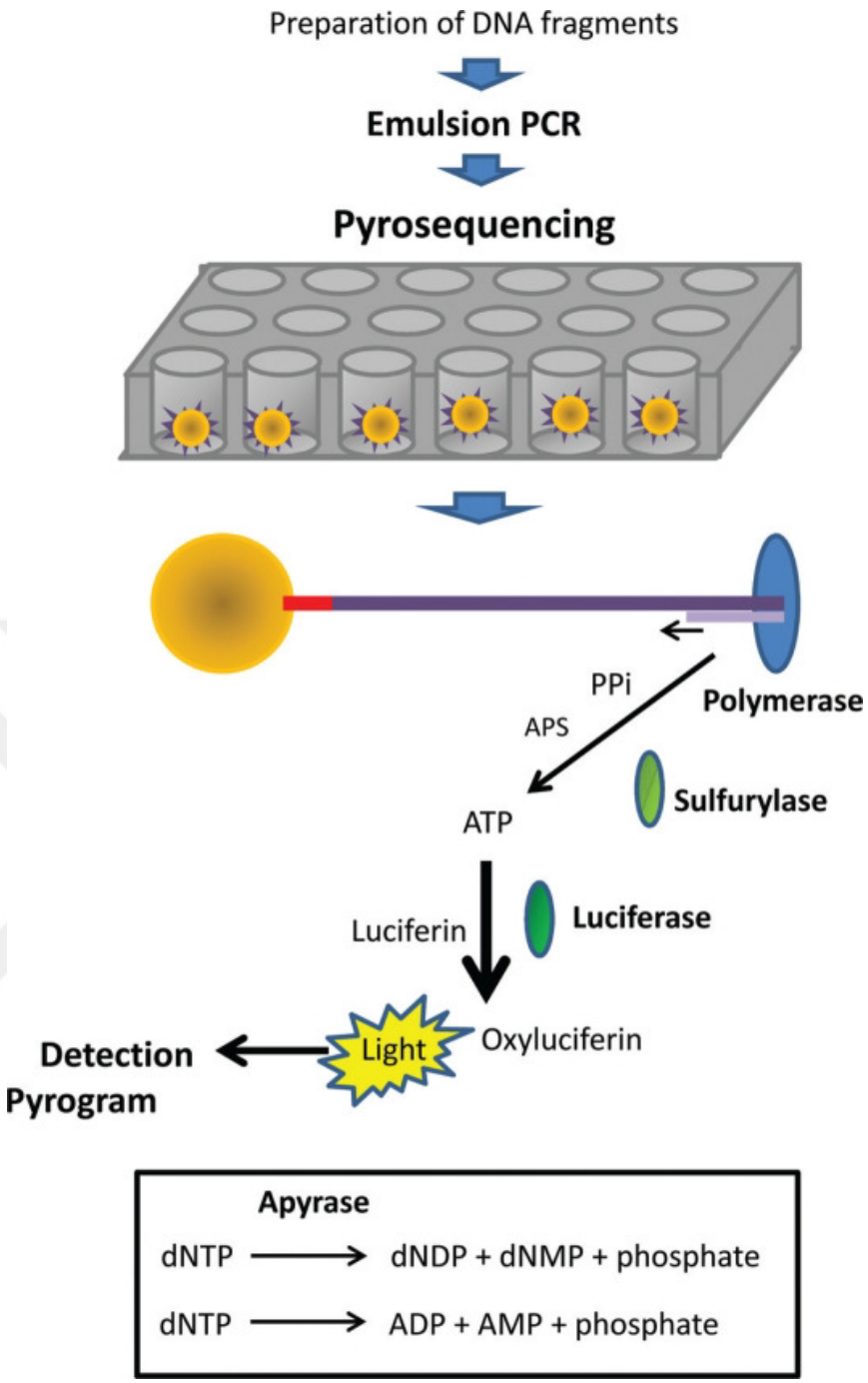

Figure 1.5. Steps of Pyrosequencing technique (Siqueira et al., 2012)

Illumina platform uses another approach that is based on reversible terminator using bridge PCR in Figure 1.6. It is similar to Sanger sequencing by ending the reaction when a reversible terminator nucleotide is incorporated onto the DNA template (Siqueira et al., 2012).

Firstly, the intact DNA material is fragmented and special adaptors are added. When the fragments transfer into a solid surface, which is called flow cell containing 8 independent channels, cluster generation start to amplify the template using bridge PCR that forms a bridge from one primer to another primer, allowing it to synthesize its complementary during annealing.. This provides that each of the amplified fragments contains approximately one million copies. Then, primer and modified nucleotide are added. In this method, four reversible modified nucleotides are added to the reaction that blocks 3'-OH. Due to the blocker, polymerase can add only one nucleotide at each time. After the image taken at each cycle, the modified nucleotide is removed chemically and the next cycle is initiated (Masoudi-Nejad et al., 2013).

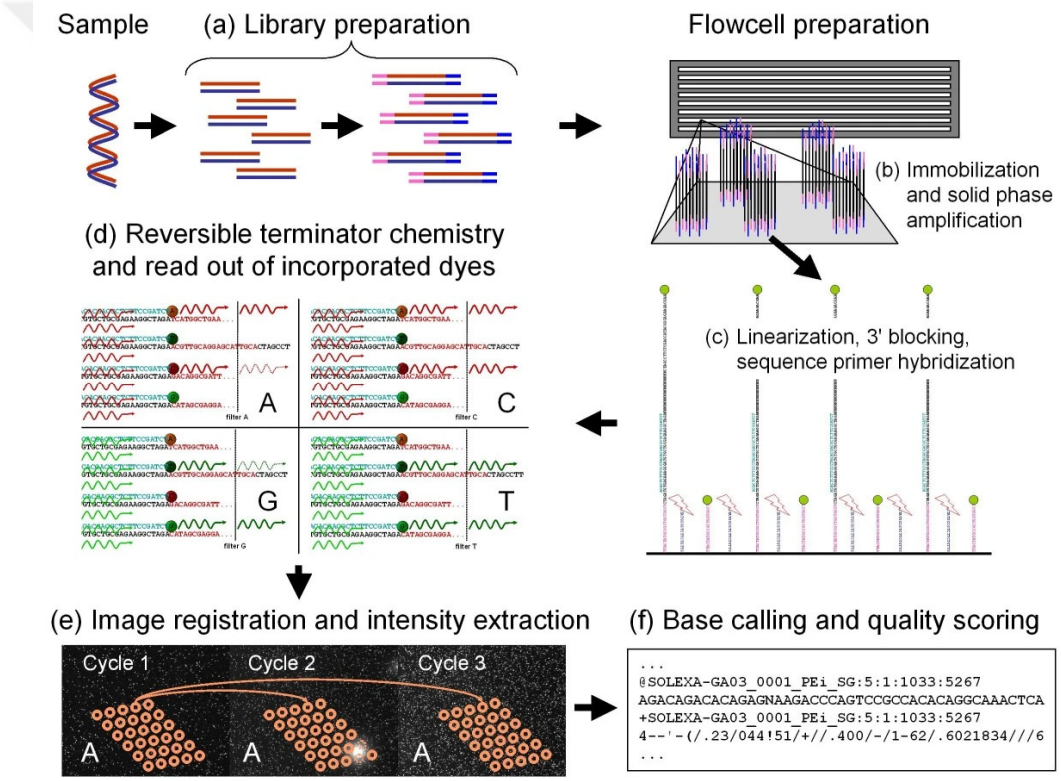

Figure 1.6. Illumina Protocol (Kircher et al., 2011)

In SGS, ABI/SOLiD developed a different approach which is based on DNA ligation technique in Figure 1.7. In this technique, beads and emulsion PCR is used as in pyrosequencing. DNA fragments are ligated to their adapters, oligonucleotides, are going to be attached to beads and then amplified by emulsion PCR. Then, these clonally amplified products are embedded on glass plate with adding several of

fluorescent labels. DNA fragments are deduced according to their colors emitted(Ambardar et al., 2016; Kchouk et al., 2017; Siqueira et al., 2012).

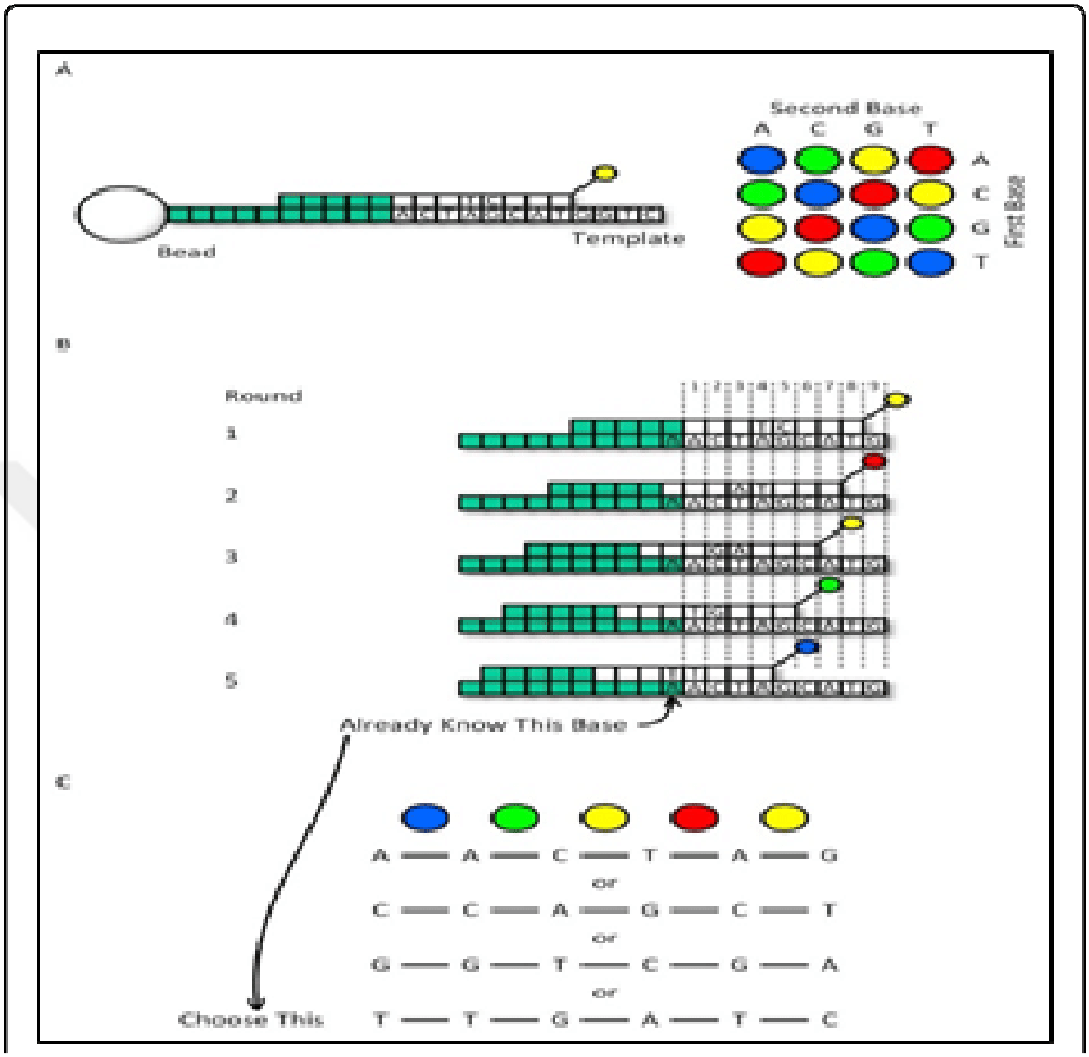

Figure 1.7. DNA ligation technique of ABI/SOLID (Mardis, 2008)

1.4.3 Third-Generation Sequencing

Due to the limitations in the second generation sequencing platforms such as short read lengths, it has become the need to a new generation, third generation sequencing (TGS) (Ambardar et al., 2016). First, it has been presented single molecule real time (SMRT) sequencing by Pacific Bioscience. The following this, Oxford Nanopore technologies and Helicos™ and the others developed TGS methods. Especially, they provide longer reads than SGS, but also more error rate of

detection of nucleotide which of in homopolymer regions, particularly. Some NGS technologies have been generally compared to each other in Table 1.2.

Table 1.2. Comparison of some different NGS technologies (modified from (Ambardar et al., 2016)).

Sequencing platforms	Error rate (%)	Read length (nts)	No. of million reads	Yield Gb/run
Pyrosequencing (454 Roche)	1	500	1	0.5
Sequencing by reverse terminator (Illumina Hiseq 2500)	0.26	2 × 100	8000 PE	720–800
Sequencing by reverse terminator (Illumina NextSeq)	0.8	2 × 150	800 PE	100–120
Sequencing by reverse terminator (Illumina Miseq)	0.8	2 × 300	44–50 PE	13.2–15
Sequencing by detection of hydrogen ion (Ion Torrent)	1.78	200	80	10
Sequencing by ligation SOLiD	0.01	35	1400	155
Single molecule real time sequencing (Pacific Biosciences)	13	40,000	0.1	0.1
Oxford nanopore technologies (minION)	38.2	2000	0.03	1
Heliscope™ single molecule sequencing	0.5	25	12–20	35

2. AIM AND SCOPE OF THE STUDY

Whole genome sequencing studies are accelerated and cost-effective since NGS technologies has already been dynamically developed on the last years. Besides, whole genome sequencing studies gives comprehensive information about the organism.

The long-term goal is the identification of eco-friendly microorganisms that can be used for agricultural purposes. To reach this goal, there are two focuses: first one is the isolation of plant growth-promoting bacteria that can be used to induce nitrogen fixation, phosphate solubilization, production of indole-3-acetic acid, etc.; second one is identification and characterization the genomics and biochemical content of those bacteria. Thus, we can better understand the nature of those bacteria in details.

In this thesis project, our general hypothesis is that the bacteria isolated from the roots of einkorn wheat (Local name: Iza) (*Triticum monococcum spp. monococcum)* in Seben/Bolu, Turkey can be used as a source for phosphate solubilization in agricultural areas. To test this hypothesis, we used computational approaches based on our working hypothesis which is whole genome sequencing should reveal at least one conserved gene for phosphate solubilization.

Thus, main objective in this study is to reveal the whole genome annotation of the phosphate solubilizing bacteria that were isolated from the root of einkorn wheat (*Triticum monococcum spp. monococcum)* in Seben, Bolu. In terms of this information, it is to reveal that

- General features of the genome
- Functional and structural annotation
- 16s rRNA and Multi-locus Phylogenetic Analyses
- To predict the genes involved in phosphate solubilization mechanism

Thus, the genetic information obtained from the phosphate solubilizing bacteria will contribute to understand the basic genetic of mechanism of phosphate

solubilization mechanism. This information will also contribute for the next genomic and/or metagenomic studies.

3. MATERIALS AND METHODS

3.1 Bacteria and DNA Isolation

Bacteria strain *Pseudomonas* spp. *ESU1531* was isolated from the roots of einkorn wheat. According to ongoing PhD study of Ömer Can Ünüvar (Academic advisor: Ercan Selçuk ÜNLÜ, Ph.D.), *Pseudomonas* spp. *ESU1531* has the highest phosphate activity among the other species isolated by sonication was selected. Since 16S, ITS and Maldi-TOF comparisons did not provide consistent data to known *Pseudomonas* species *ESU1531* which is the stock code of the strain in the laboratory of Dr. Ercan Selçuk Ünlü is assigned to the strain to track the studies till the species name is fully elucidated.

In order to isolate genomic DNA from the bacterial cells, Zymo Quick DNA Fungal/Bacterial Miniprep Kit (Cat. No: D6005) was used.

100 mg of bacterial cells were resuspended in 200 uL sterile dH_2O and 750 uL BashingBead Buffer was added. The solution was beated at maximum speed for 5 minutes. After that, it was centrifuged at 10,000 x g for 1 minute. 400 uL of supernatant was transferred to the Zymo-Spin III-F Filter in a collection tube and it was centrifuged at 8,000 x g for 1 minute before the lysis step. Genomic Lysis Buffer (1,200 uL) was added to the filtrate in the collection tube for the lysis. After the discarding the flow through from the collection tube 200 uL DNA Pre-Wash Buffer was added to the Zymo-Spin IICR Column in the new collection tube and it was centrifuged at 10,000 x g for 1 minute. After the pre-washing step 500 uL g-DNA Wash Buffer Was added to the Zymo-Spin IICR Column and it was centrifuged again 10,000 x g for 1 minute. The Zymo-Spin IICR column was transferred to a 1.5 mL microcentrifuge tube and 100 uL DNA Elution Buffer directly was added to the column. The centrifugation was applied at 10,000 g for 30 seconds to elute the DNA.

3.2 Whole Genome Sequencing and Analysis

In this study, whole genome sequencing was performed in the core facilities of Gen Era Diagnostic Incorporated. Whole genome sequencing of *Pseudomonas* spp. *ESU1531* was performed shotgun sequencing on Illumina NextSeq 550 platform using NextSeq 550 System High-Output Kit, generating 150-bp paired-end reads which of were totally 655,555,100 bp with 76-fold coverage.

It was constructed a pipeline to analyze genomic data that has been shown in Figure 3.1. Firstly, preprocessing steps, which includes trimming/removing adapter, and quality filtering, were performed on NGS data. After normalization, *de novo* assembly, annotation and phylogenetic steps were applied on the data, respectively.

Illumina pair ends reads qualities were checked using FastQC version 0.11.3. During genome sequencing, used Illumina index and adapters were removed using Trimmomatic v.0.39 (Bolger et al., 2014) and the first 10 and the last 10 bases were trimmed using FASTX-toolkit version 0.0.14.

.

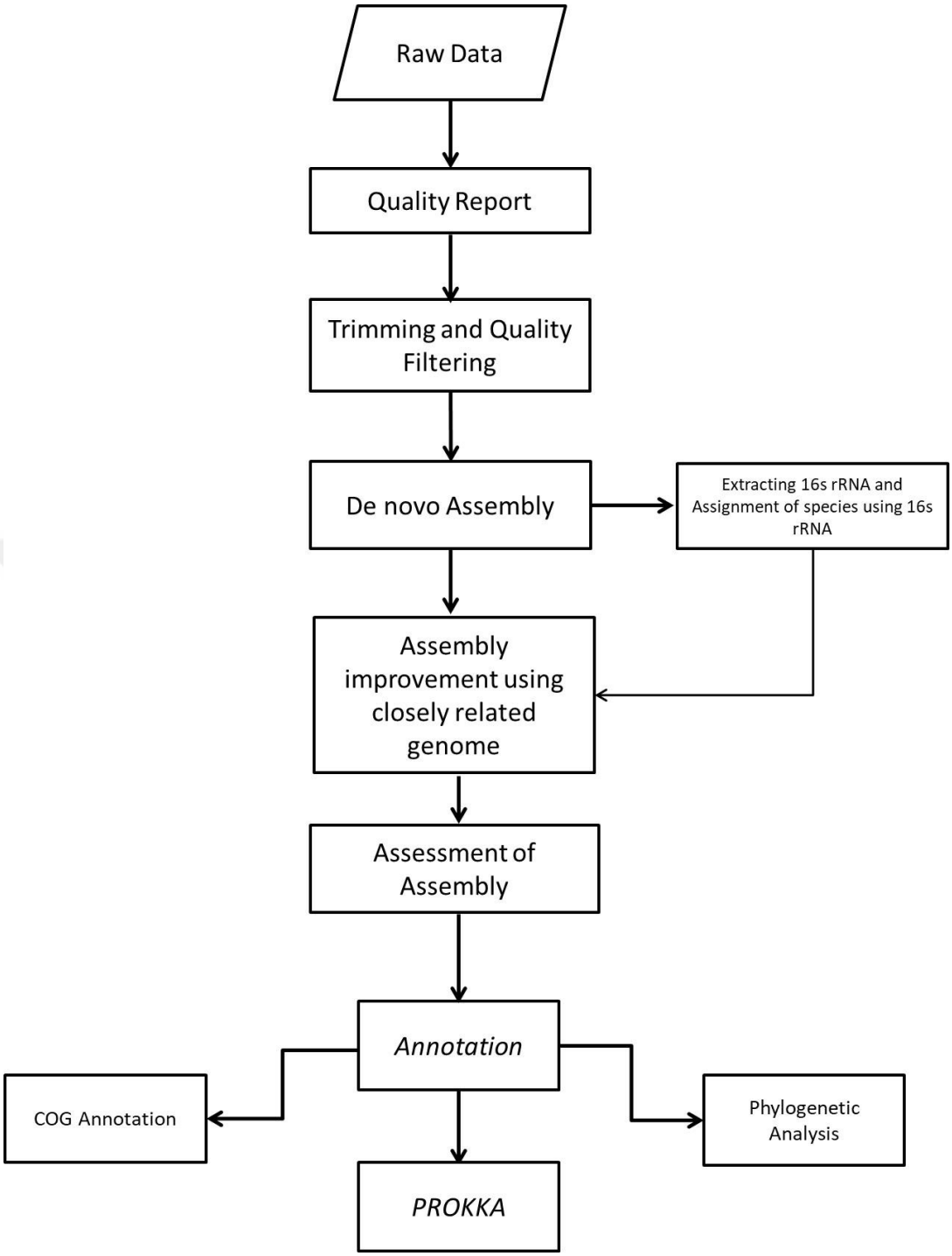

Figure 3.1. Pipeline of whole genome analysis of *Pseudomonas* spp. *ESU1531*

3.2.1 *De novo* Assembly

A total of 2,188,517 bp of paired-end sequencing was assembled using SPAdes v.3.13.0 (Bankevich et al., 2012) with 61, 81, 101, and 121 k-mers. 16s rRNA was extracted from assembly using barrnap version 0.9. Next, extracted 16s rRNA region was made BLAST on NCBI. 110 *Pseudomonas* species and the assembly were aligned using MUSCLE v.3.8.1551 (Edgar, 2004). Poorly aligned regions were removed using trimAl 1.2rev59 (Capella-Gutiérrez et al., 2009) and a maximum likelihood phylogenetic tree of 16s rRNA was constructed using IQ-TREE v.1.6.12 (Nguyen et al., 2015). Two closely related species was selected based on the tree. Then, reference genome of these species was used for re-assembly of the data for the input of assembler. For this, IDBA-Hybrid v.1.1.3 (Peng et al., 2012) was used *Pseudomonas thivervalensis* and *Pseudomonas brassicacearum* were entered as reference genome input, respectively. The best assembly result, that was to use *Pseudomonas thivervalensis,* was chosen. After this, SPAdes 3.13.0 was re-used to assemble. This time, SPAdes was executed using output of IDBA-Hybrid assembly. For this, trusted-contigs parameter of SPAdes was used. All these procedures were also repeated using *Pseudomonas brassicacearum* genome as reference and without any references. All assemblies output were evaluated using QUAST 4.6.3 (Gurevich et al., 2013).

Selected assembly was also evaluated using BUSCO (Simão et al., 2015) on the Bacteria, Proteobacteria, and G-Proteobacteria datasets that were downloaded from BUSCO website.

3.2.2 Genome Annotation and Analysis

Pseudomonas spp. ESU1531 genome was first predicted its protein using Prodigal v2.6.3. Next, a local Uniprot database was built using BLAST+ v.2.7.1 tool (Camacho et al., 2009) and all predicted protein were searched against the local database. A similarity cut off was performed on the data based on the result. Predicted proteins were divided into three groups as between 50% and 70%, 70% and 90%, and over 90% based on the Uniprot result. Cluster of orthologous Group (COG) was predicted using eggNOG-mapper v.5.0 (Huerta-Cepas et al., 2017) on

the proteins divided into three groups, respectively. After the COG annotation, the results were visualized on R studio using R script that has been manually written. Then, all the predicted proteins by Prodigal were functionally annotated using eggNOG-mapper. This result also was visualized on R platform.

For functional and structural annotation, PROKKA v1.14.6 (Rapid Prokaryotic Genome Annotation) (Seemann, 2014) was used on *Pseudomonas* spp. *ESU1531* genome. When PROKKA was run, *Pseudomonas thivervalensis* protein annotation file that was downloaded from NCBI was entered as hint input.

3.2.3 Phylogenetic Analyses of *Pseudomonas* spp. *ESU1531*

107 *Pseudomonas* species and *Pseudomonas* spp. *ESU1531* were used for phylogenetic analyses of the 16s rRNA gene and four housekeeping genes (16s rRNA, *gyrB*, *rpoB*, and *rpoD*) downloaded from NCBI. These genes ID have been used in the study of (Duan et al., 2013) before. The accession numbers of the genes used for the analysis have been shown in Appendix A.

First of all, 16s rRNA gene analysis was done. All species was aligned using MUSCLE v.3.8.1551 and visualize in SeaView v4.4.2 (Guindon et al., 2010). Then, poorly aligned region automatically fixed using trimAl 1.2rev59. A maximum likelihood tree was constructed using IQ-TREE v.1.6.12 with 1000 ultrafast bootstrap. *Pseudomonas aeuriginosa* was chosen as outgroup. Outgroup was selected according to study of Duan et al. (2013). Model was detected by ModelFinder that is found in IQ-TREE. TVMe+I+G4 was selected as the best fit model out of 88 models.

Secondly, four concatenated housekeeping genes analysis was done. After all the genes were downloaded from NCBI, they were added end-to-end, respectively. The order of the genes was *gyrB*, *rpoB*, *rpoD*, and 16s rRNA. All the species were aligned using MUSCLE v.3.8.1551 and visualize in SeaView v4.4.2. Then, poorly aligned region automatically fixed using trimAl 1.2rev59. A maximum likelihood tree was constructed using IQ-TREE v.1.6.12 with 1000 ultrafast bootstrap. *Pseudomonas aeuriginosa* was chosen as outgroup. Model was

detected by ModelFinder. GTR+F+I+G4 was chosen as the best fit model out of 88 models.

All the trees were visualized by MEGA-X (Kumar et al., 2018).

3.2.4 Average Nucleotide Identity Analysis

14 *Pseudomonas* genomes that have been showed in Table 3.1 were downloaded from NCBI. Comparison of average nucleotide identity of genomes was done using fastANI v1.3 (Jain et al., 2018). All the genomes were compared to each other, respectively and 2-dimensional matrix was constructed. 2-D matrix was visualized and colored in MS Excel.

Table 3.1 Pseudomonas species

Species
Pseudomonas brassicacearum subsp. *Brassicacearum NFM421*
Pseudomonas frederiksbergensis strain ERDD5
Pseudomonas orientalis strain F9 chromosome
Pseudomonas brassicacearum strain LBUM30
Pseudomonas chlororaphis strain NCTC7357
Pseudomonas fluorescens strain NCTC10038
Pseudomonas corrugate strain RM1-1-4
Pseudomonas trivialis strain IHBB745
Pseudomonas putida NBRC 14164
Pseudomonas thivervalensis
Pseudomonas aeuriginosa
Pseudomonas stutzeri
Pseudomonas lini
Pseudomonas kilonensis strain ZKA7

3.2.5 Chromosomal Genome and Circular Genome Visualization

All contigs of assembly was re-ordered and chromosomal genome was predicted using ABACAS v.1.3.1 (Assefa et al., 2009) with reference genome input of *Pseudomonas thivervalensis.*

Circular genome of *Pseudomonas* spp. *ESU1531* was visualized using Artemis v16.0.0 (Carver et al., 2012) and DNAplotter that is found in Artemis was used. The labels were manually added after obtaining image.

4. RESULTS AND DISCUSSIONS

4.1 Genome Assembly and Assessment

Seven assemblies have been constructed using SPAdes and IDBA assembler tools. All assemblies have been compared with each other using QUAST and selected the best assembly to go downstream analysis.

IDBA-UD is a *de novo* assembler for single cell and metagenomic sequencing data. It presents an approach which takes a closely related organism input as a reference to make *de novo* assembly (Peng et al., 2012). The other assembler tool, SPAdes, is a *de novo* assembler that works on NGS data. SPAdes assembler tool provides the trusted-contigs parameter that takes a contig fasta file that comes from the other assembly as an input. Therefore, SPAdes and IDBA-UD tools have allowed this combined approach to be implemented.

Combined approach with IDBA and SPAdes in *de novo* assembly using reference has given better results than the other approaches we performed. While using *Pseudomonas thivervalensis* as a reference in IDBA-assembler has given 560 contigs, using *Pseudomonas brassicacearum* as a reference has given 564 contigs. When these results have combined with SPAdes, the contigs have still remains less contigs in the output of *Pseudomonas thivervalensis.* In the result of combining output of IDBA with SPAdes, number of contigs is 252 and 257 in IDBA+SPAdes* (Reference genome: *Pseudomonas thivervalensis*) and IDBA+SPAdes** (Reference genome: *Pseudomonas brassicacearum*), respectively that have been shown in Table 4.1. On the other hand, both N50 statistics of them is the same. Nevertheless, since bigger than 1,000, 5,000, 10,000 bp contigs in IDBA+SPAdes* in Table 4.1, is more than IDBA+SPAdes**, IDBA+SPAdes* has been selected for downstream analysis of *Pseudomonas* spp. *ESU1531.*

In the other assembly results, using only SPAdes and only IDBA without reference have given 350 and 1374 contigs, respectively.

Table 4.1. Comparison of the assemblies and their assessment using QUAST

Assembly	IDBA+SPAdes*	IDBA+SPAdes**	IDBA+SPAdes***
# contigs (>= 0 bp)	252	257	419
# contigs (>= 1000 bp)	205	209	362
# contigs (>= 5000 bp)	166	167	264
# contigs (>= 10000 bp)	143	144	199
# contigs (>= 25000 bp)	93	93	83
# contigs (>= 50000 bp)	43	44	28
Total length (>= 0 bp)	6520191	6519782	6521150
Total length (>= 1000 bp)	6502398	6501297	6495692
Total length (>= 5000 bp)	6389149	6378263	6211238
Total length (>= 10000 bp)	6218913	6210122	5756787
Total length (>= 25000 bp)	5393586	5356827	3857038
Total length (>= 50000 bp)	3564586	3595465	1952742
# contigs	214	219	380
Largest contig	234170	234170	160162
Total length	6509537	6509128	6509928
GC (%)	61.03	61.03	61.03
N50	56484	56484	31218
N75	30989	30348	16938
L50	38	38	62
L75	75	76	134
# N's per 100 kbp	0	0	0

*Reference genome *Pseudomonas thivervalensis, **Pseudomonas brassicacearum, ***without reference genome

Table 4.1. Comparison of the assemblies and their assessment using QUAST (continued)

Assembly	Only SPAdes	Only IDBA*	Only IDBA**	Only IDBA***
# contigs (>= 0 bp)	350	560	564	1374
# contigs (>= 1000 bp)	305	459	463	1184
# contigs (>= 5000 bp)	236	333	333	413
# contigs (>= 10000 bp)	188	223	221	144
# contigs (>= 25000 bp)	86	79	78	14
# contigs (>= 50000 bp)	28	8	9	1
Total length (>= 0 bp)	6529940	6520959	6520625	6519801
Total length (>= 1000 bp)	6509518	6482160	6481438	6412224
Total length (>= 5000 bp)	6300296	6107189	6095320	4310683
Total length (>= 10000 bp)	5951439	5300014	5271238	2434557
Total length (>= 25000 bp)	4241769	2947714	2926339	539760
Total length (>= 50000 bp)	2112445	530094	595951	72192
# contigs	317	477	482	1286
Largest contig	160170	88120	88120	72192
Total length	6519731	6494974	6494934	6490177
GC (%)	61.03	61.03	61.03	61.02
N50	36688	22858	22647	7316
N75	20096	12855	12656	4053
L50	54	92	92	239
L75	115	186	187	537
# N's per 100 kbp	0	0	0	0

*Reference genome *Pseudomonas thivervalensis,* **Pseudomonas brassicacearum,* ***without reference genome

A total of 2,188,517 pair-end reads was assembled into 252 contigs with a total genome of 6,520,191. The largest contig is 234,170 kb and N50 is 56,484 kb. The results of BUSCOs in different groups were 99.3%, 99.1%, and 99.5%, respectively (Table 4.2). For bacteria database, there were 147 complete genes and 1 missing gene in total of 148 BUSCOs. For Proteobacteria, there were 218 complete genes, 1 complete and duplicated gene, and 2 missing genes in total of 221 BUSCOs. For Gamma-Proteobacteria, there were 448 complete genes, 2 complete and duplicated genes, and 2 missing genes in total of 454 BUSCOs. BUSCOs result has shown that our assembly is enough good to go downstream analysis.

Table 4.2. BUSCOs Assessment Results

Database	BUSCOs (%)
Bacteria	99.3
Proteobacteria	99.1
G-Proteobacteria	99.5

4.2 General Features of *Pseudomonas* spp. *ESU1531* and Annotation

Pseudomonas spp. *ESU1531* with single circular chromosome has 6,520,191 with whole contigs and an average G+C content of 61.03%. Total of coding sequence in the genome was manually predicted as 5735 by Prodigal to use for COGs annotation (Fig. 4.1).

As manual genome annotation is time-consuming process, rapid prokaryotic genome annotation (PROKKA) is a preferable and common annotation method for prokaryotes method in the prokaryote. Thus, we evaluated the genome annotation of *Pseudomonas* spp. *ESU1531* based on PROKKA results (Table 4.3). After PROKKA, genome size of *Pseudomonas* spp. *ESU1531* has been reduced since PROKKA did not include the contigs less than 200 bp in the analysis. According to PROKKA, total of 5652 coding sequences (CDSs) were predicted. The total number of RNA coding genes such as 16s rRNA and tRNAs, in the genome were found as 67. 4547 protein coding genes were assigned to function whereas 1105 genes were assigned to hypothetical protein or unknown function. In addition to the annotation, there were no CRISPR genes as well as plasmids in the genome. General features of *Pseudomonas* spp. *ESU1531* has been shown in Table 4.3.

Table 4.3. General Features of *Pseudomonas* spp. *ESU1531*

Features	Genome
Size (bp)	6,520,191
G+C content (%)	61.03
Number of Coding Sequences	5652
tRNAs	66
rRNA	1
Genes with assigned function	4547
Genes with assigned to COGs	5279 out of 5735
CRISPR repeats	0

According to annotation of Cluster of Orthologous Groups (COGs), 5279 out of 5735 protein were classified into 21 categories. As it has been shown in Fig. 4.1, the most abundant category was “Function unknown (S)” (27.09%) category and followed by “Transcription (K)” (8.66%) and “Amino acid transport and metabolism (E)” (8.58%). The detailed COGs numbers has been shown in Table 4.4.

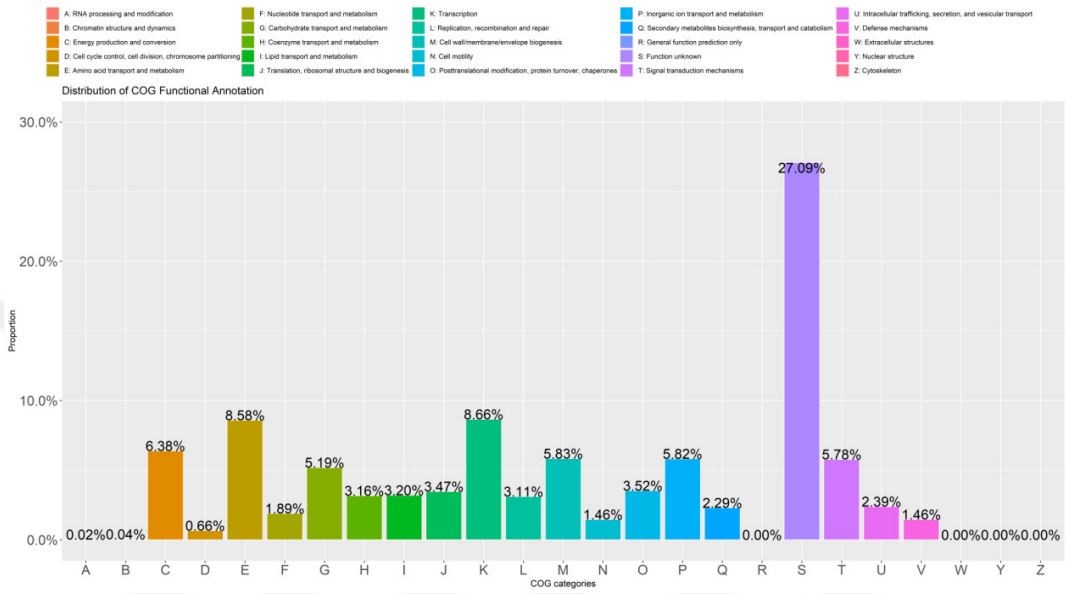

Figure 4.1. Distribution of COGs Functional Annotation

There were no proteins which assigned into W, Y, and Z which are called as "Extracellular structure", "Nuclear structure" and "Cytoskeleton", respectively. 456 of 5735 protein have not assigned to any COGs categories. As shown in the Table 4.4., there was no protein that has been interestingly assigned to General function prediction only (R category). In addition, function of 1430 proteins was unknown.

Table 4.4. COGs Functional Categories of *Pseudomonas* spp. *ESU1531*

COG Category		Number of COG
A	RNA processing and modification	1
B	Chromatin structure and dynamic	2
C	Energy production and conversion	337
D	Cell cycle control, cell division, chromosome partitionin	35
E	Amino acid transport and metabolism	453
F	Nucleotide transport and metabolism	100
G	Carbohydrate transport and metabolism	274
H	Coenzyme transport and metabolism	167
I	Lipid transport and metabolism	169
J	Translation, ribosomal structure and biogenesis	183
K	Transcription	457
L	Replication, recombination and repair	164
M	Cell wall/membrane/envelope biogenesis	308
N	Cell motility	77
O	Posttranslational modification, protein turnover, chaperones	186
P	Inorganic ion transport and metabolism	307
Q	Secondary metabolites biosynthesis, transport and catabolism	121
R	General function prediction only	0
S	Function unknown	1430
T	Signal transduction mechanisms	305
U	Intracellular trafficking, secretion, and vesicular transport	126
V	Defense mechanisms	77
W	Extracellular structures	0
Y	Nuclear structure	0
Z	Cytoskeleton	0

The genes assigned on Uniprot database were separated into three groups based on similarities threshold, between 50% and 70%, 70% and 90% and over 90%, respectively, to make COG functional annotation to assess the data in terms of speciation of *Pseudomonas* spp. *ESU1531.*

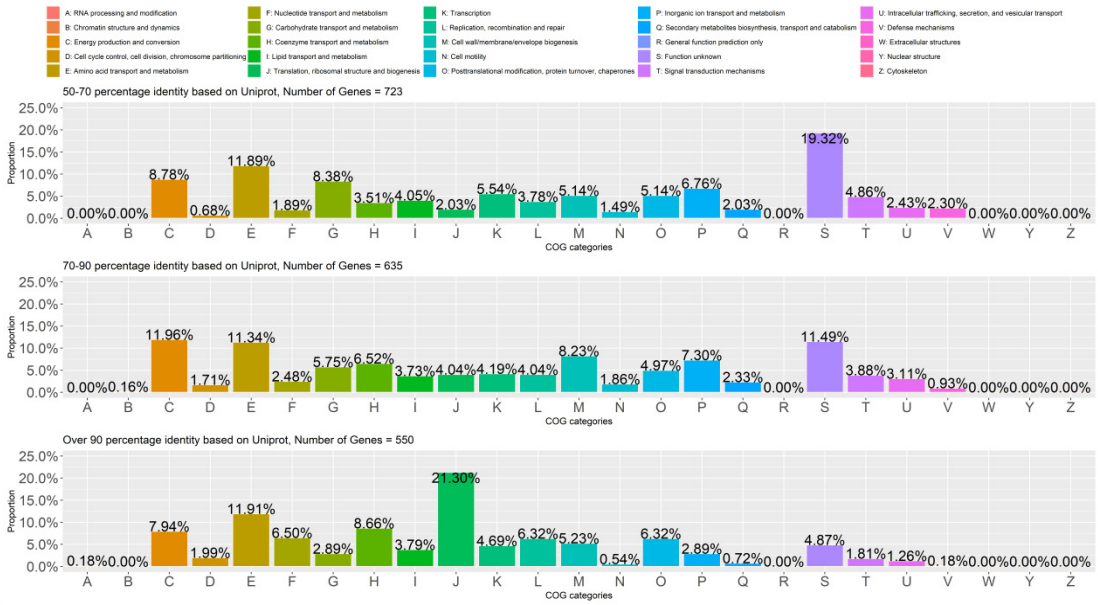

Figure 4.2. COGs Functional Annotation among the genes divided into three groups. Descriptions of categories are top of the Figure. Groups of 50-70, 70-90, and over 90 are in order from top to bottom, respectively.

In Figure 4.2, according to COGs results based on three groups, G, P, T, U and V categories have increased as it has moved from 90 percent similarities to 50 percent similarities. On the contrary, F and J categories have decreased. As G, P, T, U and V categories relate with external environment, those categories are expected to evolve faster. On the other hand, F and J are more structural categories in the genome. In addition, A category has been grouped into group of 90 percent. Thus, "RNA processing and modification" category also expected to be conserved region in the genome. This result may give an insight into speciation. Because *Pseudomonas* spp. ESU1531 has been isolated from the root of einkorn wheat grown in geographically isolated regions from the main farming areas, these orthologous groups in the categories are expected to being more conserved, while orthologous groups relating with external environment are expected to evolve faster.

Reference-based circular genome of *Pseudomonas* spp. ESU1531 has been constructed and represented in Figure 4.3. As the genome size of *Pseudomonas* thivervalensis is bigger than *Pseudomonas* spp. ESU1531, some gaps have been occurred in the genome. This heuristic approach has given us information that the location of the genes in the genome how they are. Moreover, to support this

approach, it has shown us that the long read sequences are necessary to get the actual structure of the genome of *Pseudomonas* spp. ESU1531.

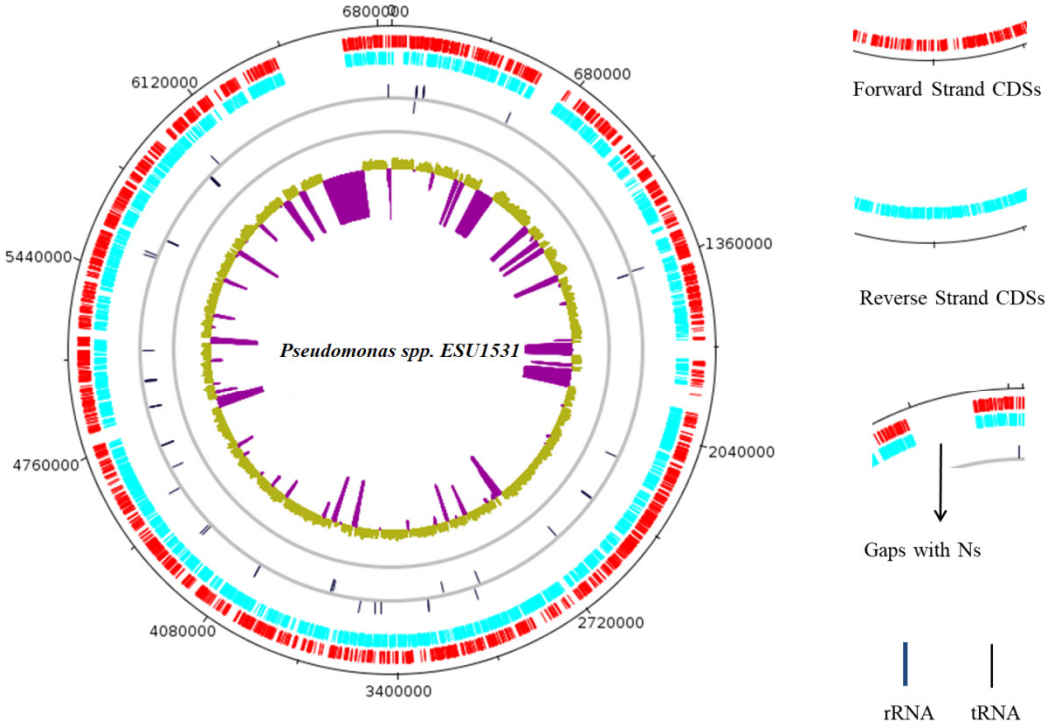

Figure 4.3. Single Circular Genome of *Pseudomonas* spp. *ESU1531.* It was heuristically constructed using reference genome of *Pseudomonas thivervalensis* that is closely related species according to 16s rRNA and multi-locus gene analysis results.

The data indicated two possible outcomes: over the years, the strain may have evolved by genome reduction event, or short reads may have been insufficient to get the actual structure of the genome.

4.3 Phylogenetic analysis of *Pseudomonas spp. ESU1531*

16s rRNA phylogeny analysis is very common method to identify and classify among bacterial species since it is highly conserved among bacteria species.

ML-based phylogenetic tree was constructed using 16s rRNA among 108 *Pseudomonas* species. According to 16s rRNA analysis in Fig. 4.4, *Pseudomonas* spp. *ESU1531* have grouped into *Pseudomonas brassicacearum, Pseudomonas*

thivervalensis, and *Pseudomonas kilonensis.* While this group's bootstrap value is 88, bootstrap value of branch of *Pseudomonas spp. ESU1531* is 74. These bootstrap values give good confidence to classify *Pseudomonas spp. ESU1531* using 16s rRNA region.

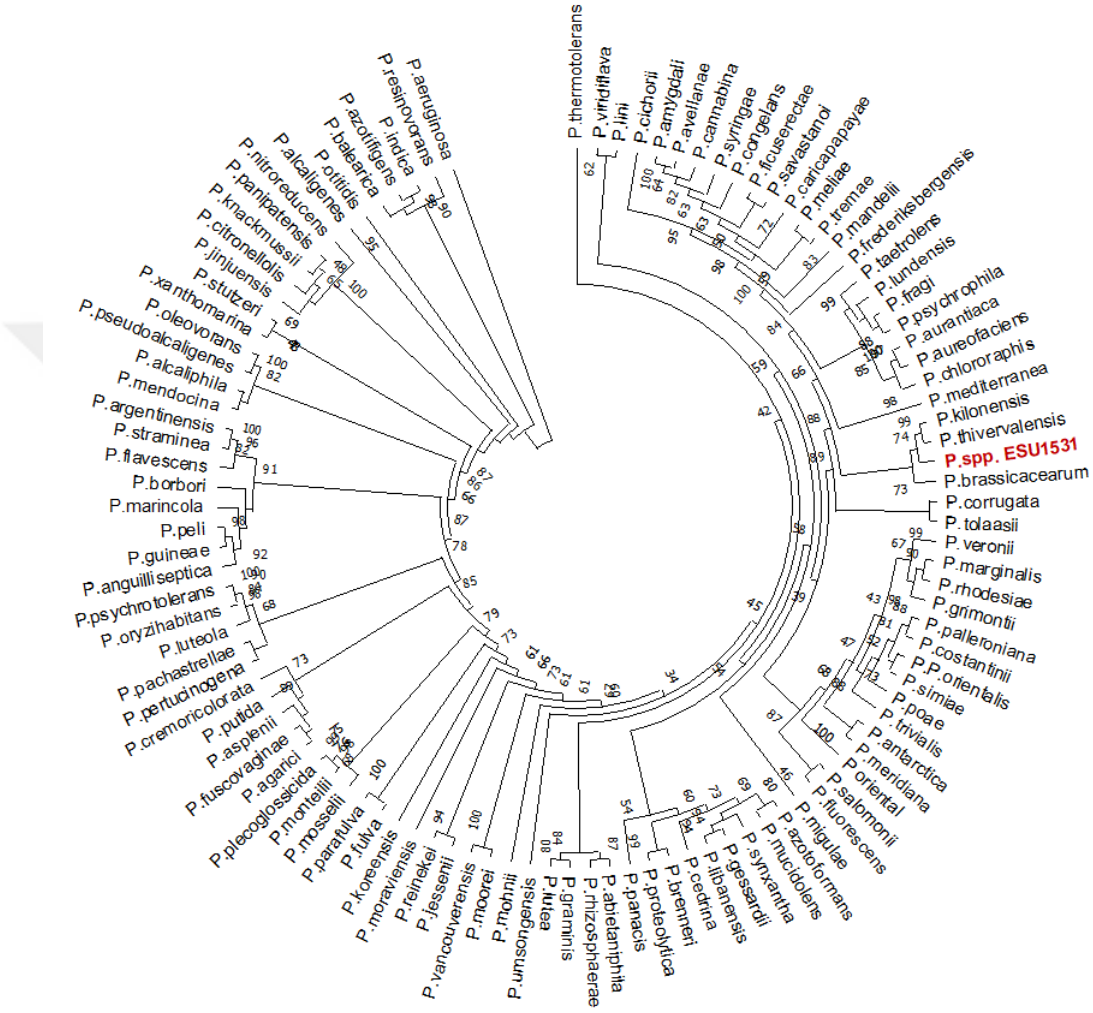

Figure 4.4. Phylogenetic Tree of 16s rRNA of *Pseudomonas* species

However, since 16s rRNA phylogenetic analyses generally give low confidence to classify *Pseudomonas* species, multi-locus sequence analysis (MLSA) has been additionally done. In the MLSA, concatenated genes (16s rRNA, *gyrB*, *rpoD*, and *rpoB*) were used and analyzed among 108 *Pseudomonas* species and constructed a ML-based tree that has been shown in Fig. 4.5.

According to MLSA result, *Pseudomonas spp. ESU1531* has branched with *Pseudomonas kilonensis.* On the other hand, the phylogenetic tree has shown that

Pseudomonas brassicacearum and *Pseudomonas thivervalensis* are very close to *Pseudomonas* spp. *ESU1531.* While housekeeping genes of *Pseudomonas* spp. *ESU1531* are grouped with those *Pseudomonas* species, node of bootstrap values is 79 that support the good confidence level to classify them.

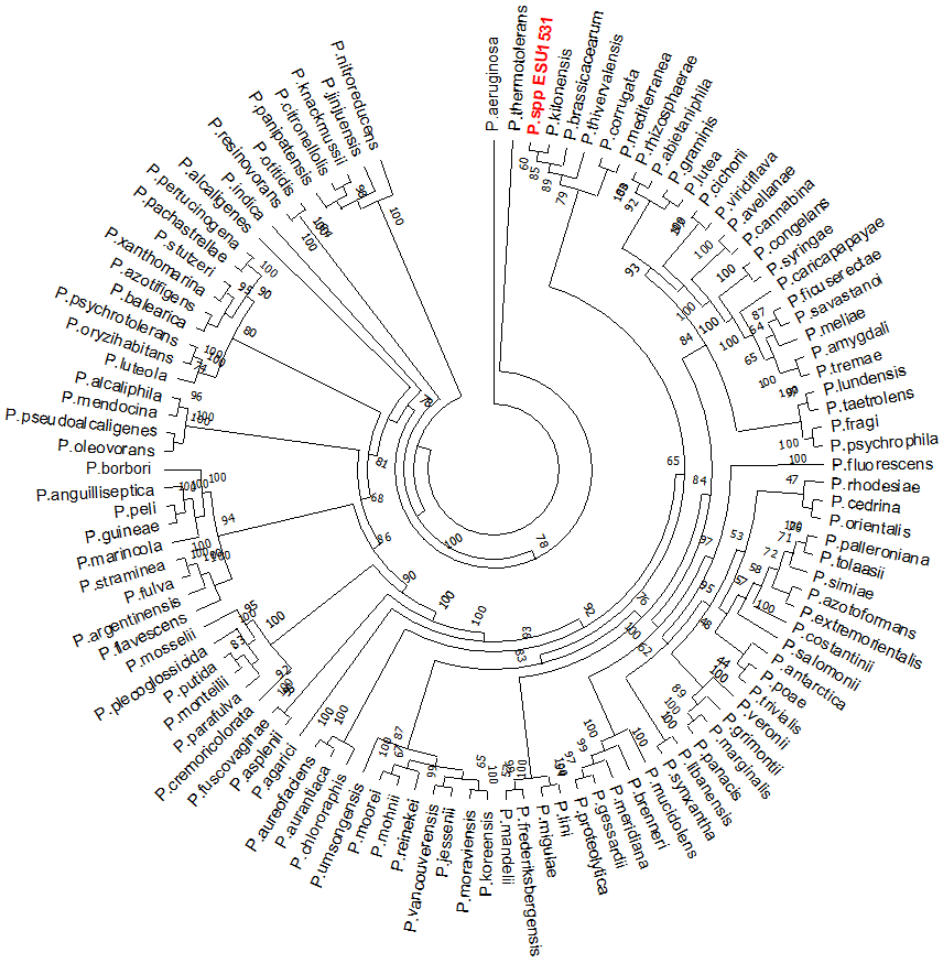

Figure 4.5. Phylogenetic Tree of Concatenated four housekeeping genes of Pseudomonas species

Both of phylogenetic trees have positively correlated to each other. In order to classify *Pseudomonas* spp. *ESU1531*, species of *Pseudomonas brassicacearum, Pseudomonas thivervalensis,* and *Pseudomonas kilonensis* should be in consideration because of the genomic similarities among the species.

The taxonomy of *Pseudomonas* has always been confusing since it was first genus discovered. *Pseudomonas* species have been revised and reclassified many times over the years. 16s rRNA phylogenetic analysis is insufficient to show species differentiation in prokaryote genera. In other words, it causes controversy when very

closely related species are analyzed. To increase the resolution power of the comparative analysis, multi-locus sequence analysis (MLSA) has been selected as a featured approach since it provides more convenient comparison results. This robust approach is based on housekeeping or core genes among bacteria species, like *Pseudomonas*. In *Pseudomonas*, *gyrB*, *rpoB*, and *rpoD* are housekeeping genes that are often used for phylogenetic studies.

4.3.1 Analysis of Average Nucleotide Identity

Average nucleotide identity was calculated against *Pseudomonas* spp. *ESU1531* (Table 4.5) using 13 species and 2-dimensional (2D) matrix was constructed (Fig. 4.6). In 2D matrix, as the colors changes from yellow to red, the similarity increases between species. On the contrary, it decreases as it is from red to yellow.

According to the result in Table 4.6, over 90 percent of species, similar to *Pseudomonas* spp. *ESU1531*, were *Pseudomonas brassicacearum* subsp. *brassicacearum NFM421* (91.8%), *Pseudomonas brassicacearum strain LBUM30* (91.8%), and *Pseudomonas thivervalensis* (93.5%), and *Pseudomonas kilonensis strain ZKA7* (91.6).

Table 4.5. Comparison of average nucleotide identity (ANI)

Species	**AANI (%)**
Pseudomonas brassicacearum subsp. *brassicacearum NFM421*	91.8434
Pseudomonas frederiksbergensis strain ERDD5	83.7902
Pseudomonas orientalis strain F9 chromosome	83.1972
Pseudomonas brassicacearum strain LBUM30	91.8196
Pseudomonas chlororaphis strain NCTC7357	84.7389
Pseudomonas fluorescens strain NCTC10038	83.1932
Pseudomonas corrugate strain RM1-1-4	87.1981
Pseudomonas trivialis strain IHBB745	83.1532
Pseudomonas putida NBRC 14164	81.1136
Pseudomonas thivervalensis	93.5378
Pseudomonas aeuriginosa	79.9727
Pseudomonas stutzeri	79.1656
Pseudomonas lini	84.3483
Pseudomonas kilonensis strain ZKA7	91.5566

On the other hand, in 2D matrix, *Pseudomonas spp. ESU1531, Pseudomonas thivervalensis,* and *Pseudomonas kilonensis strain ZKA7* showed 90 percent similarity with *Pseudomonas brassicacearum subsp. Brassicacearum NFM421* and *Pseudomonas brassicacearum strain LBUM30. Pseudomonas thivervalensis* showed 92.2%, 92.3% and (91.8% similarity with *Pseudomonas brassicacearum subsp. Brassicacearum NFM421* (*Pseudomonas brassicacearum strain LBUM30* (), and *Pseudomonas kilonensis strain ZKA7*), respectively. Following those values, *Pseudomonas corrugate strain RM1-1-4* was the highest similarity value with *Pseudomonas spp. ESU1531* (87.2%) and *Pseudomonas thivervalensis* (87.2%). This result has shown that *Pseudomonas spp. ESU1531* is the most similar to *Pseudomonas thivervalensis.* Moreover, it is also positively correlated with phylogenetic analysis.

On the other hand, ANI under 95% implies a different species among the bacterial taxonomy (Lalucat et al., 2020). Therefore, this result supports the other results as well we discussed above.

	Pseudomonas spp. ESU1531	*Pseudomonas aeuriginosa*	*Pseudomonas brassicacearum strain LBUM30*	*Pseudomonas brassicacearum subsp brassicacearum NFM421*	*Pseudomonas chlororaphis strain NCTC7357*	*Pseudomonas corrugata strain RM1-1-4*	*Pseudomonas fluorescens strain NCTC10038*	*Pseudomonas frederiksbergensis strain ERDD5*	*Pseudomonas lini*	*Pseudomonas orientalis strain F9 chromosome*	*Pseudomonas putida NBRC 14164*	*Pseudomonas stutzeri*	*Pseudomonas thivervalensis*	*Pseudomonas trivialis strain IHBB745*	*Pseudomonas kilonensis strain ZKA7*
Pseudomonas spp. ESU1531	0	80	91.8	91.8	84.7	87.2	83.2	83.8	84.3	83.2	81.1	79.2	93.5	83.2	91.5
Pseudomonas aeuriginosa	80	0	80	80	81	79.7	79.1	79.3	79.3	79.6	80.2	80.8	80	79.4	79.9
Pseudomonas brassicacearum strain LBUM30	91.8	80	0	99.4	84.8	87.5	83.3	84	84.5	83.5	81.3	79.5	92.3	83.3	94.5
Pseudomonas brassicacearum subsp brassicacearum NFM421	91.8	80	99.4	0	84.9	87.6	83.2	84	84.7	83.5	81.3	79.4	92.2	83.3	94.5
Pseudomonas chlororaphis strain NCTC7357	84.7	80.9	84.8	84.8	0	83.8	83.7	84.7	84.7	83.9	81.9	80	84.8	83.8	84.7
Pseudomonas corrugata strain RM1-1-4	87.2	79.5	87.5	87.6	84	0	82.4	83.2	83.6	82.6	80.9	79.2	87.2	82.6	87.3
Pseudomonas fluorescens strain NCTC10038	83.2	79	83.2	83.1	83.7	82.3	0	83.4	83.4	87.1	80.9	78.8	83.3	87.9	83.1
Pseudomonas frederiksbergensis strain ERDD5	84	79.2	84.1	84.1	84.8	83.2	83.5	0	85.9	83.2	80.9	79.1	84	83.4	83.9
Pseudomonas lini	84.4	79.1	84.4	84.7	84.7	83.5	83.4	85.8	0	83.3	80.8	78.7	84.4	83.4	84.6
Pseudomonas orientalis strain F9 chromosome	83.2	79.6	83.5	83.4	84	82.7	87.3	83.2	83.3	0	81.1	79.2	83.4	87.6	83.3
Pseudomonas putida NBRC 14164	81.2	80.2	81.3	81.3	81.9	81.1	80.9	80.9	80.7	81.1	0	79.9	81.4	81	81.2
Pseudomonas stutzeri	79.3	80.9	79.4	79.4	79.8	79.2	78.9	78.8	78.9	79.2	79.8	0	79.6	78.7	79.2
Pseudomonas thivervalensis	93.6	79.9	92.3	92.2	84.6	87.2	83.1	84	84.4	83.4	81.2	79.6	0	83.2	91.8
Pseudomonas trivialis strain IHBB745	83.1	79.4	83.3	83.3	83.8	82.6	87.9	83.4	83.4	87.4	80.9	78.9	83.2	0	83.1
Pseudomonas kilonensis strain ZKA7	91.5	79.9	94.5	94.5	84.7	87.3	83.1	83.9	84.6	83.3	81.2	79.2	91.8	83.1	0

Figure 4.6. 2-Dimensional Similarity matrix of average nucleotide identity. It was constructed using FastANI tool. FastANI takes a query input and reference input. Since these two entries are exchanged when creating a 2-dimensional matrix, some percentages among the same genomes have decimal level deviation in the matrix.

4.4 Genes Associated with Phosphate Solubilization Mechanism

In *Pseudomonas* spp. *ESU1531* genome, genes associated with organic and inorganic phosphate mechanism have been predicted which has shown in Table 4.6.

Table 4.6. Predicted genes involved in solubilization of inorganic and organic phosphate

Gene	Product	*Pseudomonas* spp. *ESU1531* ORF ID (PROKKA)
pqqB *pqqC* *pqqD* *pqqE* *pqqF*	Pyrroquinoline quinone biosynthesis protein	03071, 05420, 03072, 05419 05715 04156, 04656 05421
gcd	Glucose/quinate/shikimate family membrane-bound PQQ-dependent dehydrogenase	03067
phoD	Alkaline phosphatase	05702
IdhA	Lactate dehydrogenase	04258
icd	Bifunctional isocitrate dehydrogenase kinase/phosphatase	00755

All the genes involved in phosphate solubilization were annotated based on PROKKA result. The results showed that pyrroquinoline quinone (*pqq BCDEF*) gene cluster, which produces PQQ co-enzyme that is required for glucose dehydrogenase (GDH), and *gcd* gene, which produces gdh enzyme that secretes gluconic acid, are located in *Pseudomonas* spp. *ESU1531* genome. The gene, *gcd,* is the most-well known and predominantly found gene among these genes in phosphate solubilizing bacteria (Liang et al.,, 2020). Although *Pseudomonas* spp. *ESU1531* lacked *pqqA* gene, it is not required for biosynthesis of PQQ in *Pseudomonas* species (Shen et al., 2013).

We also annotated that is alkaline phosphatase encoding *phoD* gene which is other important enzyme that contributes to phosphate solubilization. Moreover, *IdhA*, and *icd* genes that potentially contributes to phosphate solubilization mechanism are located in *Pseudomonas* spp. *ESU1531* genome.

According to PROKKA annotation, five open reading frames (ORFs), which of 01102, 03902,04236,04847,05702 have been existed in the genome. These ORFs

have been identified as alkaline phosphatase family protein in the genome. When these ORF IDs aligned with *phoD* gene, in Uniprot database, which is found in *Pseudomonas fluorescens* genome, 05702 ORF ID aligned with it concordantly (Fig. 4.7).

Figure 4.7. Multiple sequence alignment of *phoD* gene between *Pseudomonas* spp. *ESU1531* (BOBJLDPC_05702) and *Pseudomonas fluorescens* (tr|A05E6YH30)

Other IDs may be related with the other alkaline phosphatase genes. These genes can be manually checked one by one in Uniprot database. We focused on only *phoD* gene in the genome to make sure which ORF ID belongs to *phoD.*

We revealed that the most known genes likely associated with the phosphate mechanism in the literature are existed in *Pseudomonas* spp. *ESU1531*.

5. CONCLUSIONS AND RECOMMENDATIONS

In this study, *Pseudomonas* spp. *ESU1531* genome was sequenced using next generation sequencing method. *De novo* assembly was carried out and genome annotation was performed by PROKKA. In addition, general features of the genome and phylogenetic analyses were carried out as well as prediction of the genes involved in phosphate solubilizing mechanism in *Pseudomonas* spp. *ESU1531.*

We revealed that *Pseudomonas* spp. *ESU1531* has several phosphate solubilizing genes in its genome. Thus, it can be used as a source for phosphate solubilization in agricultural areas. In the terms of agricultural and biotechnological studies, *Pseudomonas* spp. *ESU1531* will provide basis of information in the use of as a biofertilizer and/or biocontrol agent since we present draft genome information.

Besides, it is very close to *Pseudomonas thivervalensis* under *Pseudomonas* genus. Moreover, it is likely that *Pseudomonas* spp. *ESU1531* is a new species that was isolated from the einkorn wheat (Local name: Iza) in Seben/Bolu.

This study will contribute to future studies and help to understand the phosphate solubilizing mechanism in phosphate solubilizing bacteria and plant growth-promoting bacteria. Further, it can be used as a reference in the future genomic, metagenomic and variant studies in the same region, Seben/Bolu, and thus can provide information in the terms of speciation.

6. REFERENCES

Adhya T, Gattupali NK, Reddy G, Podile AR, Bee H, and Samantaray B (2015) "Microbial mobilization of Soil Phosphorus and Sustainable P Management in Agricultural Soils", Current Science, 108: 1280-1287.

Aliyu S (2014) "Bacterial Whole Genome Sequencing: The Future of Clinical Bacteriology", Annals of Nigerian Medicine, 8(2): 51-57.

Alori ET, Glick BR, and Babalola OO (2017) "Microbial Phosphorus Solubilization and Its Potential for Use in Sustainable Agriculture", Frontiers in Microbiology, 8: 971.

Ambardar S, Gupta R, Trakroo D, Lal R, and Vakhlu J (2016) "High Throughput Sequencing: An Overview of Sequencing Chemistry", Indian Journal of Microbiology, 56(4): 394-404.

Arif MS, Shahzad SM, Yasmeen T, Riaz M, Ashraf M, Ashraf MA, and Kausar R (2017) "Improving Plant Phosphorus (P) Acquisition by Phosphate-Solubilizing Bacteria", In M. Naeem, A. A. Ansari, & S. S. Gill (Eds.), Essential Plant Nutrients (pp. 513–556), Springer International.

Assefa S, Keane TM., Otto TD, Newbold C, and Berriman M (2009) "ABACAS: Algorithm-Based Automatic Contiguation of Assembled Sequences", Bioinformatics, 25(15): 1968-1969.

Bankevich A, Nurk S, Antipov, D, Gurevich AA, Dvorkin M, Kulikov AS, and Pevzner PA (2012) "SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing", Journal of Computational Biology, 19(5): 455-477.

Biswas K, Kumar R, Prakash V, Tarafdar A, and Kumar S (2016) Soil Microbes and Their Interaction with Plants, Sharma Publications and Distributors, India.

Bolger AM, Lohse M, and Usadel B (2014) "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data", Bioinformatics, 30(15): 2114-2120.

Brady NC, and Weil RR. (2017) The Nature and Properties of Soils, 15th Edition, Pearson, Columbus, OH, USA.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, and Madden TL (2009) "BLAST+: Architecture and Applications", BMC Bioinformatics, 10 :421.

Capella-Gutiérrez S, Silla-Martínez JM, and Gabaldón T (2009) "trimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses", Bioinformatics, 25(15): 1972-1973.

Carver T, Harris SR, Berriman M, Parkhill J, and McQuillan JA. (2012) "Artemis: An Integrated Platform for Visualization and Analysis of High-Throughput Sequence-Based Experimental Data", Bioinformatics, 28(4): 464-469.

Duan J, Jiang W, Cheng Z, Heikkila JJ, and Glick BR (2013) "The Complete Genome Sequence of The Plant Growth-Promoting Bacterium *Pseudomonas* sp. UW4", PLoS ONE, 8(3).

Edgar RC (2004) “MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput”, Nucleic Acids Research, 32(5): 1792-1797.

Gomila M, Peña A, Mulet M, Lalucat J, and García-Valdés E (2015) “Phylogenomics and Systematics in *Pseudomonas*”, Frontiers in Microbiology, 6: 214.

Gouda S, Kerry RG, Das G, Paramithiotis S, Shin HS, and Patra JK (2018) “Revitalization of Plant Growth Promoting Rhizobacteria for Sustainable Development in Agriculture”, Microbiological Research, 206: 131-140.

Gouy M, Guindon S, and Gascuel O (2010) “Sea View Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building”, Molecular Biology and Evolution, 27(2): 221-224.

Gurevich A, Saveliev V, Vyahhi N, and Tesler G (2013) “QUAST: Quality Assessment Tool for Genome Assemblies”, Bioinformatics, 29(8): 1072-1075.

Heather JM, Chain B (2016) “The Sequence of Sequencers: The History of Sequencing DNA”, Genomics, 107(1): 1-8.

Hesse C, Schulz F, Bull CT, Shaffer BT, Yan Q, Shapiro N, and Loper JE (2018) “Genome-Based Evolutionary History of *Pseudomonas* spp”, Environmental Microbiology, 20(6): 2142-2159.

Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, Von Mering C, and Bork P (2017) “Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-mapper”, Molecular Biology and Evolution, 34(8): 2115-2122.

Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, and Aluru S (2018) “High Throughput ANI Analysis of 90K Prokaryotic Genomes Reveals Clear Species Boundaries”, Nature Communications, 9(1): 5114.

Kchouk M, Gibrat JF, and Elloumi M (2017) “Generations of Sequencing Technologies: From First to Next Generation”, Biology and Medicine, 9(3): 395.

Khan MS, and Zaidi A (2009) Phosphate Solubilizing Microbes for Crop Improvement, Nova Science Publishers Inc., UK.

Kircher M, Heyn P, and Kelso J (2011) “Addressing Challenges in The Production and Analysis of Illumina Sequencing Data”, BMC Genomics, 12: 382.

Kumar S, Stecher G, Li M, Knyaz C, and Tamura K (2018) “MEGA X: Molecular evolutionary genetics analysis across computing platforms”, Molecular Biology and Evolution, 35(6): 1547-1549.

Lalucat J, Mulet M, Gomila M, and García-Valdés E (2020) “Genomics in Bacterial Taxonomy: Impact on the Genus *Pseudomonas*”, Genes, 11(2): 139.

Liang JL, Liu J, Jia P, Yang T tao, Zeng Q wei, Zhang S chang, and Li J tian (2020) “Novel Phosphate-Solubilizing Bacteria Enhance Soil Phosphorus Cycling Following Ecological Restoration of Land Degraded by Mining”, ISME Journal, 14(6): 1600-1613.

Mardis ER (2008) “Next-Generation DNA Sequencing Methods”, Annual Review of Genomics and Human Genetics, 9: 387-402.

Masoudi-Nejad A, Narimani Z, and Hosseinkhan N (2013) Next Generation Sequencing and Sequence, First Edition, Springer-Verlag, New York.

Meyer JB, Frapolli M, Keel C, and Maurhofer M (2011) "Pyrroloquinoline Quinone Biosynthesis Gene pqqC, A Novel Molecular Marker for Studying the Phylogeny and Diversity of Phosphate-Solubilizing Pseudomonads", Applied and Environmental Microbiology, 77(20): 7345-7354.

Nguyen LT, Schmidt HA, Von Haeseler A, and Minh BQ (2015) "IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-likelihood Phylogenies", Molecular Biology and Evolution, 32(1): 268-274.

Peng Y, Leung HCM, Yiu SM, and Chin FYL (2012) "IDBA-UD: A *De novo* Assembler for Single-Cell and Metagenomic Sequencing Data with Highly Uneven Depth", Bioinformatics, 28(11): 1420-1428.

Ramakrishna W, Yadav R, and Li K (2019) "Plant Growth Promoting Bacteria in Agriculture: Two Sides of a Coin", Applied Soil Ecology, 138: 10-18.

Rodríguez H, and Fraga R (1999) "Phosphate Solubilizing Bacteria and Their Role in Plant Growth Promotion", Biotechnology Advances, 17(4-5): 319-339.

Seemann T (2014) "Prokka: Rapid Prokaryotic Genome Annotation", Bioinformatics, 30(14): 2068-2069.

Sharma SB, Sayyed RZ, Trivedi MH, and Gobi TA (2013) "Phosphate Solubilizing Microbes: Sustainable Approach for Managing Phosphorus Deficiency in Agricultural Soils", SpringerPlus, 2: 587.

Shen X, Hu H, Peng H, Wang W, and Zhang X (2013) "Comparative Genomic Analysis of Four Representative Plant Growth-Promoting Rhizobacteria in *Pseudomonas*", BMC Genomics, 14: 271.

Shrivastava M, Srivastava PC, and D'Souza SF (2018) Role of Rhizospheric Microbes in Soil: Volume 2: Nutrient Management and Crop Improvement, Springer Singapure.

Silby MW, Winstanley C, Godfrey SAC, Levy SB, and Jackson RW (2011) "*Pseudomonas* Genomes: Diverse and Adaptable", FEMS Microbiology Reviews, 35(4): 652-680.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, and Zdobnov EM (2015) "BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs", Bioinformatics, 31(19): 3210-3212.

Siqueira JF, Fouad AF, and Rôças IN (2012) "Pyrosequencing As A Tool for Better Understanding of Human Microbiomes", Journal of Oral Microbiology, 4: 10.

Watson JD, and Crick FHC (1953) "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid", Nature, 171(4356): 737-738.

Winsor GL, Griffiths EJ, Lo R, Dhillon BK, Shay JA, and Brinkman, FSL (2016) "Enhanced Annotations and Features for Comparing Thousands of *Pseudomonas* Genomes in the *Pseudomonas* Genome Database", Nucleic Acids Research, 44(1): 646-653.

APPENDICES

7. APPENDICES

Appendix A

Genbank Accession Numbers of Genes Used in the Phylogenetic Analysis of *Pseudomonas spp. ESU1531*

Species	**16S rRNA**	***gyrB***	***rpoD***	***rpoB***
P. abietaniphila	AJ011504	FN554166	FN554447	AJ717416
P. aeruginosa	AF094713	AJ633104	AJ633568	AJ717442
P. agarici	AJ308298	AB039456	AB039560	AJ717477
P. alcaligenes	AB680567	AB039388	AB039606	AJ748191
P. alcaliphila	AB030583	FN554167	FN554448	AJ717463
P. amygdali	NR_036999	AB039474	AB039510	AJ717462
P. anguilliseptica	AB021376	FN554168	FN554449	FN554726
P. antarctica	HE586386	FN554169	FN554450	FN554727
P. argentinensis	AY691188	FN554170	FN554451	FN554728
P. asplenii	AB021397	AB039473	AB039594	AJ717432
P. avellanae	AJ889838	FN554173	FN554454	AJ717469
P. azotiFigens	AB189452	FN554174	FN554455	FN554729
P. azotoformans	PSEIAM08	AB039411	AB039547	AJ717458
P. balearica	AF054936	AB039394	AJ633565	AJ717480
P. borbori	AM114527	FN554175	FN554456	FN554730
P. brassicacearum	AJ292381	AM084675	AM084334	AJ717436
P. brenneri	AF268968	FN554176	FN554457	AJ717482
P. cannabina	AJ492827	FN554177	FN554458	AJ717453
P. caricapapayae	PSEATCC09	AB039454	AB039507	AJ717437
P. cedrina	AF064461	FN554178	FN554459	AJ717424
P. chlororaphis	FJ652610	FJ652718	D86036	AJ717478
P. aurantiaca	AB021412	FN554171	FN554452	AJ717426
P. aureofaciens	AJ308300	FN554172	FN554453	AJ717421

P. cichorii	AB021398	AB039436	AB039524	AJ717418
P. citronellolis	AB021396	AB039452	AB039604	FN568270
P. congelans	AJ492828	FN554179	FN554460	FN554731
P. corrugata	AF348508	AB039466	AB039574	AJ748175
P. costantinii	AF374472	FN554180	FN554461	FN554732
P. cremoricolorata	AB060137	FN554181	FN554462	AJ717476
P. extremorientalis	AF405328	FN554182	FN554464	FN554733
P. ficuserectae	AB021378	AB039418	AB039501	AJ717457
P. flavescens	AJ308320	FN554183	FN554465	AJ717468
P. fluorescens	AF094725	AB039492	AB039624	HE586409
P. fragi	AF094733	FN554184	FN554466	AJ717444
P. frederiksbergensis	FR750403	AM084676	AM084335	AJ786298
P. fulva	AB046998	AB039440	AB039599	AJ717419
P. fuscovaginae	AB021381	FN554185	FN554467	AJ717433
P. gessardii	AF074384	FN554186	FN554468	AJ717438
P. graminis	NR_026395	FN554187	FN554469	AJ717429
P. grimontii	AF268029	FN554188	FN554470	AJ717439
P. guineae	AM491810	FN554189	FN554471	FN554734
P. indica	AB681925	FN554190	FN554472	AJ717481
P. jessenii	NR_024918	FN554191	FN554473	AJ717447
P. jinjuensis	AB681927	FN554192	FN554474	FN554735
P. kilonensis	AJ292426	AM084677	AM084336	AJ717472
P. knackmussii	NR_041702	FN554193	FN554475	FN554736
P. koreensis	AB495129	FN554194	FN554476	FN554737
P. libanensis	NR_024901	FN554195	FN554477	AJ717454
P. lini	AB649011	FN554196	FN554478	AJ717466
P. lundensis	AB021395	FN554197	FN554479	AJ717428
P. lutea	NR_029103	FN554198	FN554480	FN554738
P. luteola	AJ871471	FN554199	FN554481	AJ717452
P. mandelii	AF058286	FN554200	FN554482	AJ717435
P. marginalis	AB021401	AB039472	AB039578	AJ717425
P. marincola	AB301071	FN554201	FN554483	FN554739
P. mediterranea	JQ904748	AM084678	AM084337	AJ717449

P. meliae	AB021382	FN554202	FN554484	AJ717486
P. mendocina	AJ308310	AJ633103	AJ633567	AJ748169
P. meridiana	AJ537602	FN554203	FN554485	FN554740
P. migulae	NR_024927	FN554204	FN554486	AJ717446
P. mohnii	AM293567	AM293561	FN554487	FN554741
P. monteilii	AF064458	FN554205	FN554488	AJ717455
P. moorei	AM293566	AM293560	FN554489	FN554742
P. moraviensis	AY970952	FN554206	FN554490	FN554743
P. mosselii	AF072688	FN554207	FN554491	FN554744
P. mucidolens	PSEIAM16	AB039409	AB039546	AJ717427
P. nitroreducens	AM088473	FN554208	FN554492	AJ717448
P. oleovorans	AB680450	AB039396	AB039601	AJ717461
P. orientalis	AF064457	FN554209	FN554493	AJ717434
P. oryzihabitans	GQ250598	FN554210	FN554494	AJ717470
P. otitidis	NR_043289	FN554211	FN554495	FN554745
P. pachastrellae	AB125366	FN554212	FN554496	FN554746
P. palleroniana	AY091527	FN554213	FN554497	FN554747
P. panacis	AY787208	FN554214	FN554498	FN554748
P. panipatensis	EF424401	FN554215	FN554499	FN554749
P. parafulva	AB060132	FN554216	FN554500	AJ717471
P. peli	AM114534	FN554217	FN554501	FN554750
P. pertucinogena	AB021380	DQ350613	FN554502	AJ717441
P. plecoglossicida	AB009457	FN554218	FN554503	AJ717456
P. poae	AJ492829	FN554219	FN554504	FN554751
P. proteolytica	AJ537603	FN554220	FN554505	FN554752
P. pseudoalcaligenes	AB021379	AB039439	AB039598	AJ717430
P. psychrophila	NR_028619	FN554221	FN554506	AJ717464
P. psychrotolerans	NR_042191	FN554222	FN554507	FN554753
P. putida	AF094736	AB039451	AB039581	AJ864841
P. reinekei	AM293565	AM293559	FN554508	FN554754
P. resinovorans	AB021373	FN554223	FN554509	AJ717479
P. rhizosphaerae	AY152673	FN554224	FN554510	FN554755
P. rhodesiae	NR_024911	FN554225	FN554511	AJ717431

P. salomonii	AY091528	FN554226	FN554512	FN554756
P. savastanoi	DQ318862	AB039468	AB039513	AJ717422
P. simiae	AJ936933	FN554227	FN554513	FN554757
P. straminea	NR_036908	AB039410	AB039600	FN554758
P. stutzeri	AF094748	AB039392	AB039613	AJ748189
P. synxantha	PSEIAM24	AB039415	AB039550	AJ717420
P. syringae	AJ308316	AB039428	AB039516	FN554759
P. taetrolens	PSEIAM26	AB039412	AB039523	AJ717423
P. thermotolerans	AJ311980	FN554228	FN554514	FN554760
P. thivervalensis	AF100323	AM084679	AM084338	AM084680
P. tolaasii	AF348507	AB039423	AB039561	AJ748160
P. tremae	AJ492826	FN554229	FN554463	FN554761
P. trivialis	AJ492831	FN554230	FN554515	FN554762
P. umsongensis	NR_025227	FN554231	FN554516	FN554763
P. vancouverensis	AJ011507	FN554232	FN554517	AJ717473
P. veronii	AB021411	FN554233	FN554518	AJ717445
P. viridiflava	AF094751	AB039427	AB039520	FN554764
P. xanthomarina	AB176954	AM905836	AM905872	FN554765

8. CURRICULUM VITAE

Name SURNAME : Uğur ÇABUK

Place and Date of Birth : Bakırköy - 10.06.1990

Universities

Bachelor's Degree : Bolu Abant İzzet Baysal University

e-mail : ugurcabuk23@gmail.com

Address : Turgut Reis Mah. 506. Sk. No:25/4 Esenler/ İstanbul, TURKEY