CHAPTER 4

GENOME SEQUENCE ANALYSIS AND COMPARATIVE PHYLOGENOMICS OF *PSEUDOMONAS PUTIDA* N4T

4.1 Abstract

Pseudomonas putida is a Gram-negative, rod-shaped bacterium that is metabolically versatile and can be found in different ecological niches. *P. putida* N4T an endophyte, isolated from seven-days old healthy tomato seedlings, was found to possess antagonistic activity against a bacterial phytopathogen *Ralstonia solanacearum*, which causes lethal bacterial wilt disease in more than 400 plant species including tomato. *P. putida* N4T was found to be efficient in protecting tomato seedlings against *R. solanacearum* F1C1 wilt disease. In this study, we performed a Next Generation Sequencing, to analyze the genome of this bacterium and a comparative phylogenomics study of this bacterium was performed with five other *Pseudomonas* strains. Also different *in silico* tools were used to reveal the presence of secondary metabolite gene clusters, possible Plant growth promotion traits, and plant interaction factors. The NGS and the short read de novo assembly generated the 5,529,188bp genome and an average GC content of 63.3%. *In silico* analysis of its genome sequence revealed the presence of secondary metabolite biosynthetic gene clusters that are involved in the biosynthesis of different compounds/metabolites such as pyoverdine, putisolvin, WLIP, lankacidin C and fragin. Possible 3714 genes associated with plant growth promoting traits (PGPTs) were found to be presentin the genome and presence of 38 different factors encoded by a range of genes present in the genome responsible for plant interaction and virulence were identified.

4.2 Introduction

Pseudomonas species are ubiqutious, and can be isolated from sources ranging from plants to contaminated soils and water to human clinical samples [1-3]. The genus *Pseudomonas* is one of the most diverse bacterial genera, containing over 220 validly described species which is one of the highest among Gram negative bacteria [4]. While the genus contains few species that are pathogenic to plants and animals, a large majority of *Pseudomonas* species are commensals, and also the role of various species in bioremediation, biostimulation, and in control of plant diseases has been elucidated [5-

10]. In addition, *Pseudomonas* are known to produce several bioactive secondary metabolites that often support them to dwell in a different lifestyle and play different roles which include iron scavenging, swarming motility, biofilm formation, pathogenicity, cooperation, antagonism or competition against other microorganisms [11, 12]. The genus *Pseudomonas* is also phylogenetically rather heterogenous, containing several subgroups, also genome sequencing has revealed that the size is wide ranging.

Pseudomonas putida, an important member of the genus *Pseudomonas*, is known to exhibit a diverse spectrum of metabolic activities which includes the ability to degrade organic solvents and heavy metals etc. This enables them to reside in highly polluted environments. Also, *P. putida* strains are known to present in the rhizosphere and the role in plant growth promotion has also been elucidated [13-20]. *P. putida* is also been demonstrated to be a potential bio-control agent against different plant pathogens including *Fusarium oxysporumf.sp. radicis-lycopersici*, *Rhizoctonia solani Pectobacterium atrosepticum*, *Sclerotinia sclerotiorum*, *Ralstonia solanacearum*, *Xanthomonas oryzae* pv. *oryzae*, *X. o.* pv. *oryzicola*, *X. citri* subsp. *Citri* and *Uromyces appndiculatus* [21-24]. Although the role of P. *putida* in bio-control of different diseases has been established, the elucidation of the mechanisms involved in bio-control has not been extensively studied yet. For a better understanding of the *P. putida* strains and to identify the factors responsible for traits including plant growth promotion, secondary metabolite production, and antagonism, technologies like genome sequencing can be very useful. Already many strains of *P. putida* have been sequenced and have been analyzed. The analysis and comparison of different available genome sequences of *P. putida* strains and comparison of *P. putida* species with other *Pseudomonas* species will help to find out the evolutionary relationships among the species and also will help in better understanding of the species.

P. putida N4T was isolated from healthy tomato seedlings grown under gnotobiotic condition. The strain was found to exhibit antagonistic activity against a bacterial phytopathogen *Ralstonia solanacearum* F1C1, a causal agent of bacterial wilt. *R. solancearum* F1C1 was isolated from wilted chilli plants and characterized to cause disease in seedlings stages of tomato and eggplant in lab condition [25-30]. It was established that *P. putida* N4T reduced the bacterial wilt disease in tomato seedlings and grown-up plants and the appearance of wilt symptoms were delayed. In this study, we performed whole genome sequencing of the *P. putida* N4T genome using Illumina

platform, and after annotation we performed in silico study to find out possible secondary metabolite producing gene clusters, and also in a separate analysis we identified factors responsible for plant interaction and plant growth promoting traits. Moreover, we have performed a comparative phylogenomics study of *P. putida* N4T with five other *Pseudomonas* strains. The study includes three other *P. putida* strains *P. putida* KT2440, *P. putida* F1, *P. putida* BIRD-1, also two other pathogenic strains *P. aeruginosa* PAO1, an opportunistic human pathogen and *P. syringaepv. syringae* B728a is a plant pathogen of beans. Strain KT2440 is a rhizospheric bacterium isolated from garden soil in Japan. KT2440 is an extensively studied strain and it was the first *P. putida* strain to be the whole genome sequenced [15]. *P. putida* F1 was obtained from a polluted creek in USA and can be used in bioremediation due to its ability to degrade aromatic hydrocarbon compounds [31]. Strain BIRD-1 is a rhizospheric Plant growth promoting bacterium that is highly tolerant to drought and is able to synthesize siderophores, and phytohormones such as IAA solubilize inorganic phosphate [32]. *P. aeruginosa* PAO1 was the first sequenced *Pseudomonas* genome, which is an opportunistic human pathogen and was originally isolated from a wound [33]. *P. syiangae pv. syringae* Strain B728a was isolated from a snap bean leaflet in Wisconsin [34]. We found that *P. putida* N4T acquires an extensive repertoire of traits responsible for plant association, growth promotion and also obtain different secondary metabolite gene clusters.

4.3 Materials and Methods:

4.3.1 Bacterial Growth and DNA Extraction

P. putida N4T was isolated from six to seven days old tomato seedlings grown under gnotobiotic condition as described in chapter 2**.** The strain was found to be exhibiting antagonistic activity against a plant pathogenic bacterium *R. solanacreaum* F1C1 strain which was isolated from wilted chili plants in Tezpur, Assam [27]. The bacterium was further found to be effective in bio-control of disease caused by F1C1 in tomato and brinjal seedlings. A single colony of *P. putida* N4T grown on BG (Bacto Glucose) agar was inoculated into 5ml of BG broth and grown overnight in shaking conditions at 28°C. Bacterial cells were collected in pellets through centrifugation and genomic DNA was isolated with a QiaAmp DNA mini kit (Qiagen, USA) according to the manufacturer's instructions. The quality of gDNA was checked on 1% Agarose gel (loaded 5µl) for the

single intact band. The gel was run at 110 V for 30 mins. 1 µl of each sample was loaded in Nanodrop 8000 for determining the A260/280 ratio and 1 µl of each sample was used for determining concentration using Qubit® 2.0 Fluorometer.

4.3.2 Preparation of 2 x 150 Paired-End (PE) Libraries

The paired-end sequencing library was prepared using Illumina TruSeq Nano DNA HT library preparation kit. 200ng gDNA was fragmented by Covaris to generate a mean fragment distribution of 550bp. Covaris shearing generates dsDNA fragments with 3' or 5' overhangs. The fragments were then subjected to end-repair. This process converts the overhangs resulting from fragmentation into blunt ends using End Repair Mix. The 3' to 5' exonuclease activity of this mixremoves the 3' overhangs and the 5' to 3' polymerase activity fill in the 5' overhangs. A single ‘A’ nucleotide is added to the 3’ ends of the blunt fragments to prevent them from ligating to oneanother during the adapter ligation reaction. A corresponding single ‘T’ nucleotide on the 3’ endof the adapter provides a complementary overhang for ligating the adapter to the fragment. This strategy ensures a low rate of chimera (concatenated template) formation. Indexing adapters were ligated to the ends of the DNA fragments, preparing them for hybridization onto a flow cell. The ligated products were purified using SP beads supplied in the kit. The size-selected product was PCR amplified as described in the kit protocol. The amplified library was analyzed in Bioanalyzer 2100 (Agilent Technologies) using HighSensitivity (HS) DNA chip as per the manufacturer's instructions.

4.3.3 Cluster generation and sequencing

After obtaining the Qubit concentration for the library and the mean peak size from the Bioanalyser profile, library was loaded for cluster generation and sequencing. Paired-End sequencing allows the template fragments to be sequenced in both the forward and reverse directions. The kit reagents will be used in the binding of samples to complementary adapter oligos on the paired-end flowcell. The adapters are designed to allow selective cleavage of the forward strands after re synthesis of the reverse strand during sequencing. The copied reverse strand is then used to sequence from the opposite end of the fragment.

4.3.4 De novo Assembly of N4T genome

De novo assembly of high-quality PE reads of sample N4T was accomplished using CLC Genomics workbench 6[35]. The statistical elements of the assembly were calculated by in house Perlscripts.

4.3.5 N4T Genome Annotation and Analysis

The *P. putida* N4T genome sequence was first annotated by Prokaryotic Dynamic Programming Genefinding Algorithm (Prodigal) v 2.6 which is a microbial gene-finding program. Prodigal is a highly accurate gene finder. Prodigal's false positive rate compares favorably with other gene identification programs and usually falls under 5%. Sensitivity is the number of correct genes divided by the number of genes in the Genbank file and Specificity, is a measure of the false positive rate, and is the number of 'Matching Genes'divided by the number of genes predicted by Prodigal [36]. "Accuracy" is defined as (SN+SP)/2. Each predicted CDS was functionally annotated evaluating the homology by BLASTX search against the NR database. Based on similarity searches with known proteins, we annotated 4849 CDS (about98.79% of the Total CDS) with cut-off E-values of 10^{-5}. Putative coding sequences (CDS) were identified by the Glimmer tool, which is embedded by the OmicsBox pipeline [37, 38]. For the identification of transfer RNA (tRNA) genes, tRNAscan-SE v.1.31 was used. It allows detection and accuratesecondary structure prediction of unusual tRNA species including prokaryotic and eukaryotic selenocysteine tRNA genes, as well as tRNA-derived repetitive elements and pseudogenes [39]. Ribosomal RNA (rRNA) genes were predicted by using RNAmmer 1.2 server. It predicts 5s/8s, 16s/18s, and 23s/28s ribosomal RNA in full genome sequences [40]. Repeat sequences were examined by using MISA software [41]. Genome annotation of N4T was also done by using Rapid Annotation using Subsystem Technology (RAST) service. RAST is a fully-automated service for annotating complete or nearly complete bacterial and archaeal genomes. It provides high-quality genome annotations for these genomes across the whole phylogenetic tree [42]. The N4T genome was also annotated by using the Prokaryotic Genome Annotation Pipeline (PGAP) developed by the National Center for Biotechnology Information (NCBI) [43]. PGAP determines structural annotation by comparing open reading frames (ORFs) to libraries of protein hidden Markov models (HMMs), representative RefSeq proteins and proteins from well-characterized reference genomes. GeneMarkS+ then makes ab initio coding

region predictions for genomic regions that lack HMM or protein evidence and selects start sites for ORFs whose evidence comes from HMMs. After annotation the genome was automatically submitted to NCBI through the PGAP pipeline and ad subsequently to GenBank, which is a genetic sequence database where annotated collections of all publicly available DNA sequences are incorporated. GO annotation was obtained for annotated CDS using blast2GO Pro. GO sequence distributionshelps in specifying all the annotated nodes comprising of GO functional groups. CDS associated with similar functions are assigned to the same GO functional group. The GO sequence distributionwas analyzed for all the three GO domains i.e. biological processes, molecular function and cellular components. Ortholog assignment and mapping of the CDS to the biological pathways were performed using KEGG automatic annotation server (KAAS) [44]. All the genes were compared against the KEGG database using BLASTX with a threshold bit-score value of 60 (default). The identification of secondary metabolite biosynthesis gene clusters in the N4T genome was performed using antiSMASH 6.o. antiSMASH allows the rapid genome-wide identification, annotation, and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genomes. It integrates and cross-links with a large number of in silico secondary metabolite analysis tools that have been published earlier [45]. The genome sequence of N4T has been deposited in GenBank under the accession number WBOI00000000 and the version is WBOI00000000.1.

4.3.6 Comparative Gene Content Analysis of N4T using EDGAR

EDGAR is a tool that provides comparative gene content analysis, Phylogenetic analysis between genomes and also includes features that allow visualization of the results in the form of plots or Venn diagrams. The EDGAR platform for comparative genomics permits the differential gene content of bacteria to be easily analyzed and their common features and differences to be identified. Furthermore, EDGAR provides AAI and ANI matrices as well as the core-genome-based Phylogenetic trees. The EDGAR tool contains numerous numbers of published genomes that are accessible to users. Also using EDGAR analysis of unpublished genomes and comparison of them with already publicly available genomes can be done [46].

Comparative genomics of *P. putida* N4T with three other *P. putida* strains and two of the pathogenic *Pseudomonas* species were performed using the EDGAR tool. The

Pseudomonas strains taken in the study were *P. putida* KT2440, *P. putida* F1, *P. putida* Bird1, *P. aeruginosa* PAO1 and *P. syringaepv. syringae* B728a. All the three P. putida strains are non pathogenicwhere as *P. aeruginosa* PAO1 is an opportunistic human pathogen and *P. syringaepv. Syringae* B728a is a plant pathogen of bean. The genomes of all these strains are already sequenced and the sequence data is present in public databases. N4T genome was included in the EDGAR software platform along with the other five mentioned genomes and a custom project namely Pseudomoas_Tezpur was created for the subsequent comparative genomics study.

The core genome and pan-genome of all the six organisms were calculated using the EDGAR tool. Also, different set size statistics were calculated and based on the statistics and data different plots like core development plot, singleton development plot, pan development plot and pan vs. core development plots were generated by the EDGAR tool. Also using EDGAR some other plots and diagrams like Venn diagram, synteny plot, synteny matrix, circular plot, etc. were generated for visualizing common features between all the genomes, also using these plots the difference between the genomes can also be interpreted.

The phylogenetic tree was constructed using the phylogeny pipeline of the EDGAR tool to understand the phylogenetic relationship between N4T and the other five genomes that were included in the project. The phylogeny pipeline calculated the core genome of all strains included in the project, and then multiple alignments for all core gene sets were calculated. As this pipeline analyses the phylogenetic relationships between genomes based on the thousands of orthologous genes contained in the complete core genome, the phylogenetic trees usually show high stability.

The phylogenetic relationship between all the six genomes was also evaluated by calculating ANI and AAI values. For the AAI method, the average AAIs of all conserved genes in the core genome as computed by the BLAST algorithm is collected. The results can be easily extracted from the EDGAR database. ANI values are computed as described in and as implemented in the popular JSpecies package. For both methods, the resultingphylogenetic distance values are arranged in an AAI/ANI matrix, clustered according to their distance patterns andvisualized as heat maps. The heat map images, as well as theraw AAI/ANI values, can be exported from the web server.

Table 4.1: Strains used for comparative genomics study

Strain	Size (Mb)	GC%	Isolated From	Function
P. putida N4T	5.52	63.3	Tomato seedlings (This study)	Anti-bacterial activity
P. putida KT2440	6.18	61.6	Garden soil	Biotechnological applications
P. putida F1	5.95	61	Polluted creek	Potential for bio-remediation
P. putida BIRD-1	5.7	61.7	Soil	Plant growth promotion
P. aeruginosa PAO1	6.3	66.6	wound	Opportunistic human pathogen
P. syringaepv. syringae B728a	6.09	59.23	Bean leaflet	Plant pathogen of bean

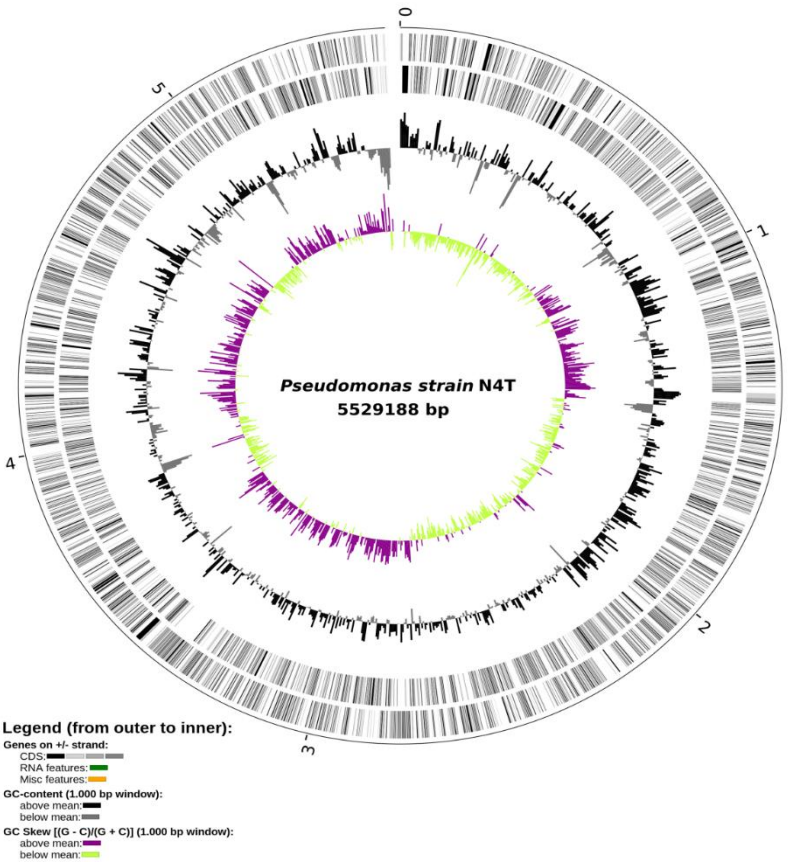


Fig 4.1: Circular representation of *P. putida* N4T genome. Circle (from outside to inside): First, scale bar in megabases; second and third, predicted coding sequences of *P. putida* N4T genome on the leading and the lagging strand, respectively; fourth, G+C content; fifth, G+C skew.

4.4 Results

4.4.1 General features of N4T Genome

The genome of *P. putida* N4T has a single circular chromosome of 5,529,188bp and an average GC content of 63.3%. A total of 4908 CDS were predicted by the Prodigal algorithm, out of those 4849 (98.79% of total CDS) CDS were annotated using the BLASTX tool. A total of 4920 putative coding genes were predicted by the Glimmer tool. Using tRNAscan-SE v.1.31 and RNAmmer 1.2 server total of 50 tRNA and 4 rRNA were identified respectively. A total of 8 different repeat sequences were found in the N4T genome. Using RAST 5089 geneswere annotated out of which 4983 were found to be protein-coding genes, among the coding genes 882 were annotated as a hypothetical protein. Also, a total of 50 tRNA and 5rRNA were identified by the RAST tool. Using PGAPtotal of 5035 genes was annotated in which 4973 CDS were identified. Out of the 4973 CDS, 4875 were identified as protein-coding genes. A total of 62 RNA genes were identified, out of which 51 were tRNA genes, 7were rRNA genes and 4ncRNA genes.

Table 4.2: Characteristic features of *P. putida* N4T genome

Features	
Size (bp)	5,529,188
GC content	63.3%
Predicted no. of CDS	4908 (Prodigal), 4920 (Glimmer),
rRNA genes	4 (RNAmmer 1.2), 5 (RAST), 7 including partial sequence (PGAP)
tRNA genes	50 (tRNAscan-SE v.1.31 and RAST), 51 (PGAP)
Annotated genes	4849 (BLASTX), 5089 (RAST), 5035 (PGAP)
Hypothetical protein	882 (RAST)
Repeat sequence	8 (MISA)

4.4.2 Secondary metabolite biosynthesis gene clusters

The genome of *P. putida* N4T was analyzed for the presence of secondary metabolite biosynthetic gene clusters using the antiSMASH6 platform. A total of 11 putative secondary metabolite biosynthesis clusters were predicted for N4T. Seven of these

werefound to be NRPS (Non-Ribosomal Peptide Synthetase) or NRPS like clusters. Apart from NRPS clusters there were lassopeptide, betalactone, hserlactone, bacteriocin, butyrolactone, redox-cofactor and NAGGN type clusters were predicted in N4T genome. These clusters were found to have genes that are involved in the biosynthesis of different compounds/metabolites such as pyoverdine, putisolvin, WLIP, lankacidin C and fragin.

One NRPS cluster belonging to region 3.3 showed 100% similarity with WLIP biosynthesis genes present in a reference organism *P. putida* strain RW10S2. This suggests that the same WLIP is also present in *P. putida* N4T, as all the relevant genes required for the biosynthesis of WLIP are shared between the two strains. On the other hand, three gene clusters present in region 1.3, region 12.1 and region 14.1 were found to have no similarity with genes that are known to have a role in synthesizing secondary metabolites/compounds.There is a possibility that these clusters produce a completely unknown molecule or some molecule that has not been reported yet. Also, there are four regions 1.1, 3.1, 9.1 and 17.1 belongs to NRPS and betalactone type BGCs which showed very less percentage similarity (less than 10% in each) with pyoverdin of *P. protegens*Pf-5. This suggests that because minorities of the genes are shared so the pyoverdin will be different from Pf-5 and other known pyoverdine. In region 3.2, an NRPS, bacteriaocin type BGC was present, which showed 50% similarity with a biosurfactant, putisolvin from *P. putida* PCL1445. Region 21.1, an NRPS like BGC was found to exhibit 37% gene similarity with fragin producing BGC of *Burkholderia cenocepacia* H111.

Table 4.3: Prediction of secondary metabolite biosynthesis gene clusters

Region/Cluster	**Type**	**Metabolite/compound**	**similarity**	**Organism**
Region 1.1	NRPS	pyoverdin	9%	*Pseudomonas protegens* Pf-5
Region 1.3	lassopeptide	-	-	-
Region 3.1	Betalactone, NRPS	pyoverdin	2%	*P. protegens* Pf-5
Region 3.2	NRPS, bacteriocin	putisolvin	50	*P. putida* PCL1445
Region 3.3	NRPS	WLIP	100%	*P. putida* RW10S2
Region 6.1	Redox-cofactor	Lankacidin C	13%	*Streptomyces rochei*
Region 9.1	NRPS	pyoverdin	5%	*P. protegens* Pf-5
Region 12.1	hserlactone,	-	-	-

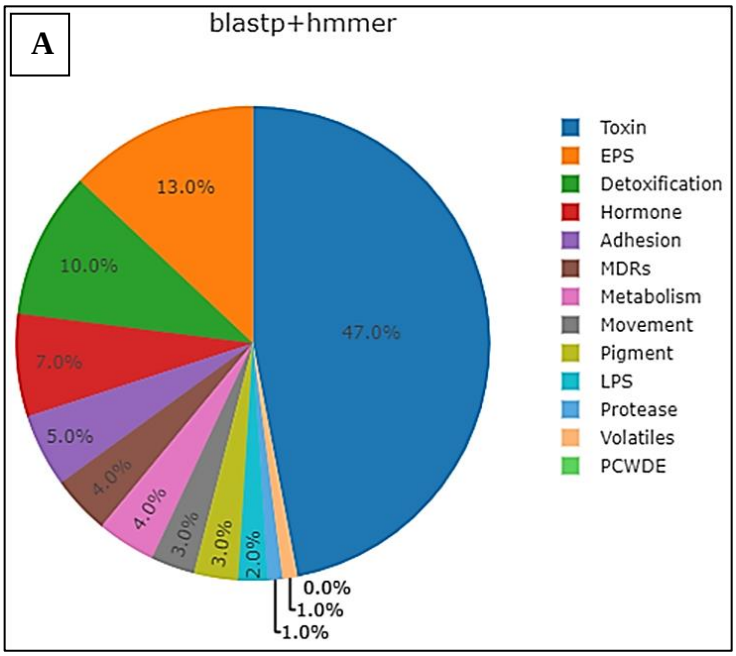
	butyrolactone			
Region 14.1	NAGGN	-	-	-
Region 17.1	NRPS	pyoverdin	2%	*P. protegns* Pf-5
Region 21.1	NRPS-like	fragin	37%	*Burkholderiacenocepacia* H111

4.4.3 Association of *P. putida* N4T with plants

P. putida N4T isolated from tomato seedlings was found to be localized inside vascular tissues of tomato seedlings as well in grown up plants in laboratory conditions. PLaBase web resource was used to predict the traits for plant growth promotion and plant association of *P. putida* N4T.PLaBAse encompasses various features including (i) a database for screening 5,565 plant associated bacteria (PLaBA-db), (ii) a tool for predicting plant growth-promoting traits (PGPTs) of bacteria using their genome data (PGPT-Pred), and (iii) a tool for the prediction of bacterial plant-association by marker gene identification (PIFAR-Pred) [47].

PGPT-Pred analysis of the *P. putida* N4T genome revealed the presence of possible 3714 genes associated to PGPTs in the genome. The genes were found to be linked with different functions including colonizing plant system, competitive exclusion, stress control, bio-fertilization, phytohormone/plant signal production, bio-remediation, plant immune response stimulation and putative functions.

PIFAR-Pred analysis of the N4T genome predicted the presence of possible 38 factors encoded by different genes present in the genome those play role in plant interaction and virulence. These factors were responsible for important functions such as toxins, EPS, detoxification, hormone, adhesion, MDRs, metabolism, movement, pigment, LPS, protease, volatiles and PCWDE.



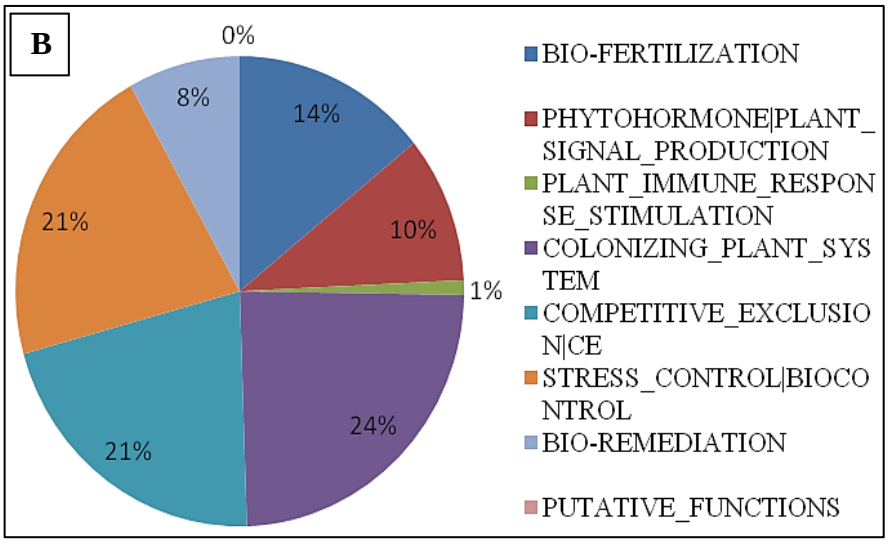


Fig 4.2: A. Plot generated by PGPT –Pred tool represents the percentage of genes predicted involved in functions that contributes to plant growth promotion, B. Plot generated by PIFAR –Pred tool represents percentage of genes possible involved in traits responsible for plant association and virulence.

Table 4.4: Possible factors responsible for plant interaction and colonization (PIFAR-Pred)

	Possible factors responsible for plant interaction and colonization (PIFAR- Pred)
1	Haemagg_ac
2	Pilin
3	Usher
4	YadA
5	attC
6	cbb
7	copA
8	dps
9	isothiocyanate_resistance
10	katB
11	katE
12	katG
13	pip
14	galU
15	gpsX
16	gumH
17	cytokinin_ptz
18	salycilic_hydroxylase
19	galU
20	rfb303
21	wzt
22	ACR_tran
23	Multi_Drug_Res
24	OEP
25	acnB
26	aroC
27	aroK
28	aroQ
29	mqo
30	purine_biosynthesis_purC
31	purine_biosynthesis_purD
32	trpCG
33	motN
34	pilA
35	lipA
36	htrA
37	tagetitoxin
38	budB

4.4.4 Phylogenetic analysis of *Pseudomonas* strains

The Phylogenetic tree constructed by the neighbor-joining method taking the orthologous genes of all the five genomes showed that *P. putida* Bird1 and *P. putida* KT2440 were closely related. The next related strain to these two was found to be *P. putida* F1. *P. putida* N4T was distantly related to the three strains. Whereas the two pathogenic strains *P. aeruginosa* PAO1 and *P. syringae pv. syringae* B728a were not very much similar to N4T and the other three non-pathogenic strains as appeared in the Phylogenetic tree.

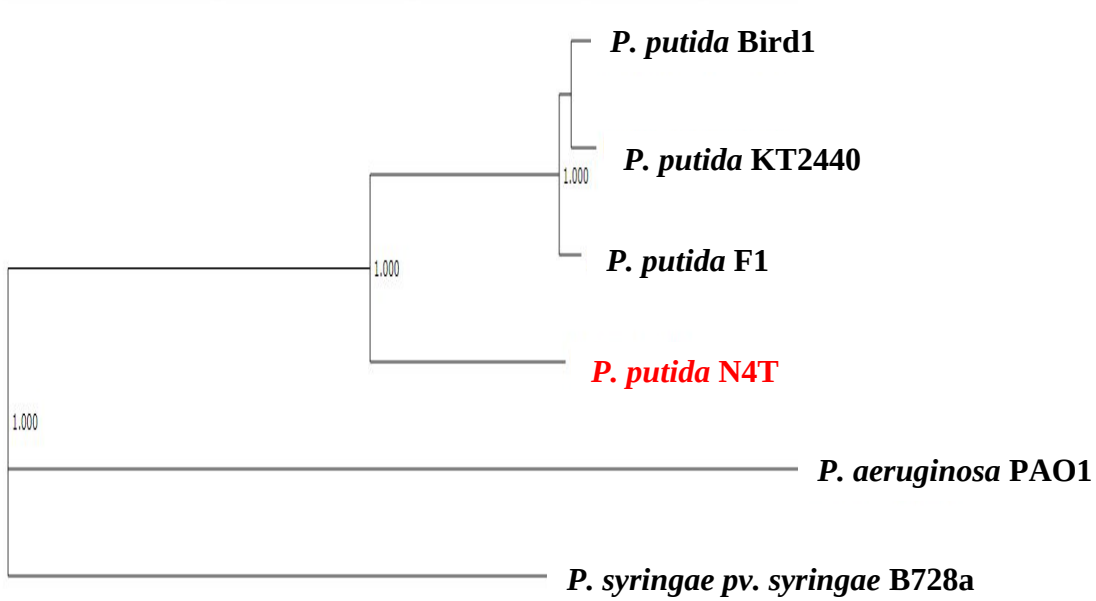


Fig 4.3: Phylogenetic tree of the six strains based on the multiple alignment of the core genome

4.4.5 Phylogenetic analysis by ANI and AAI

The heatmaps generated by calculating both the ANI and AAI values to establish the Phylogenetic relationship between all the six genomes including N4T showed that the strains F1, BIRD-1 and KT2440 were close to each other. The ANI and AAI percentage of N4T was not very similar to the otherthree *P. putida* strains mentioned above. It was observed that N4T was not a part of the same cluster that all the three above mentioned *P. putida* strains formed. Also, the heat maps showed that the opportunistic human pathogen PAO1 and plant pathogen B728a were not so similar with all the four non-pathogenic strains; they both shared less than 80% ANI and AAI with all the four strains. PAO1 and B728a shared less than 80% ANI and AAI between them too.

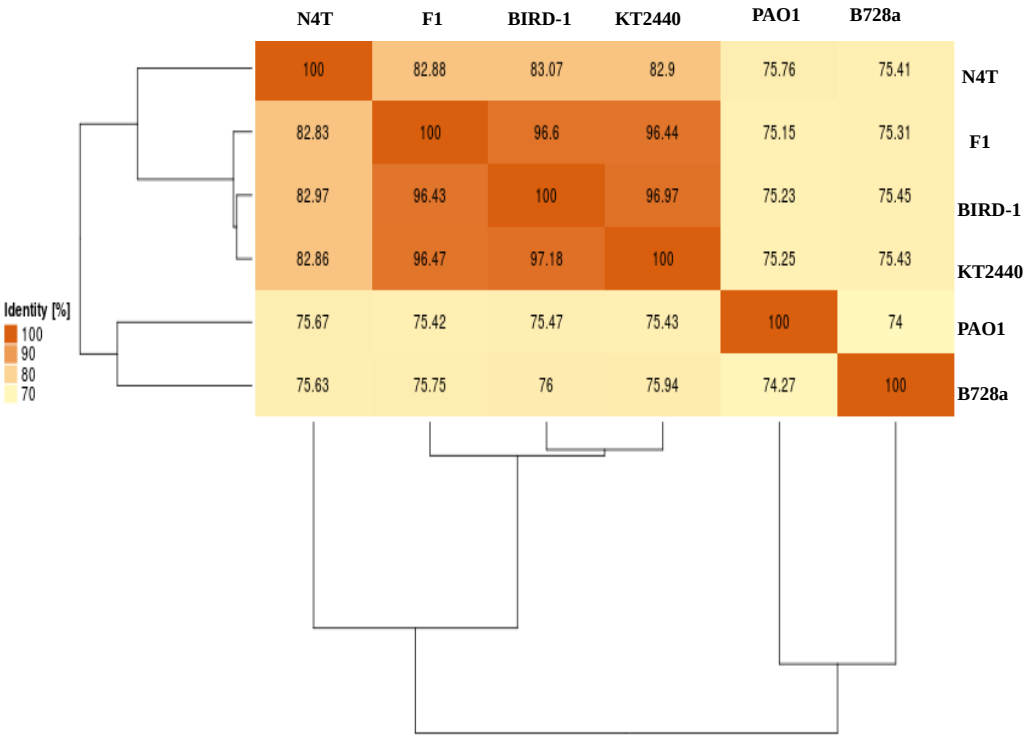


Fig 4. 4: Heat map of ANI mean percentage between different *Pseudomonas* strains, including lab isolate N4T.

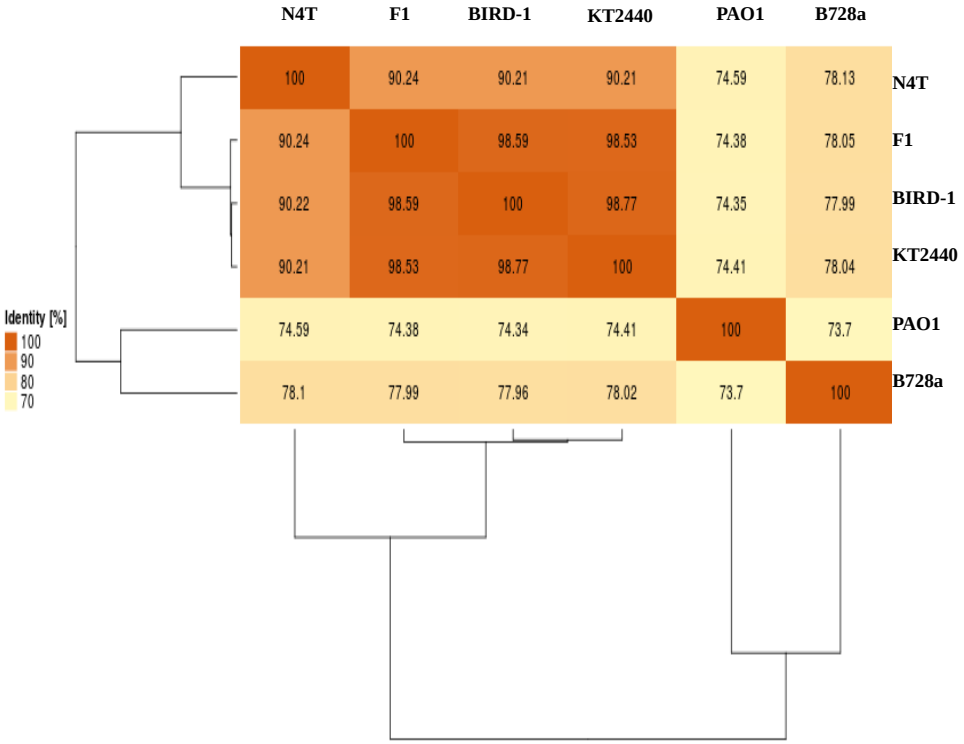


Fig 4.5: Heat map of AAI mean percentage between different *Pseudomonas* strains, including lab isolate N4T.

4.4.6 Core genome V/S Pan genome analysis

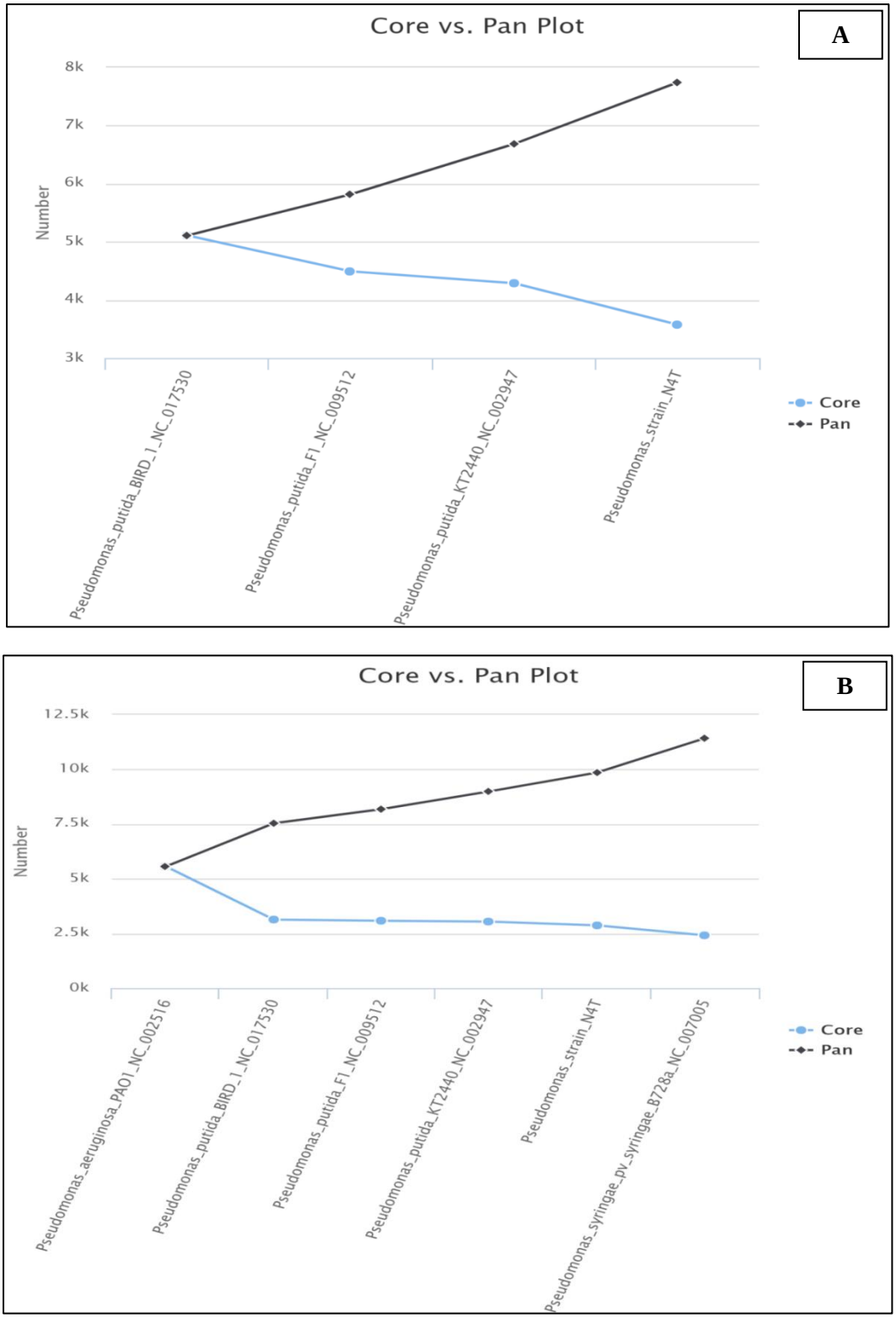


Fig 4.6: Core versus pan genome development plot of the four *P. Putida* strains including the lab isolate N4T (A) and all the six strains (B)

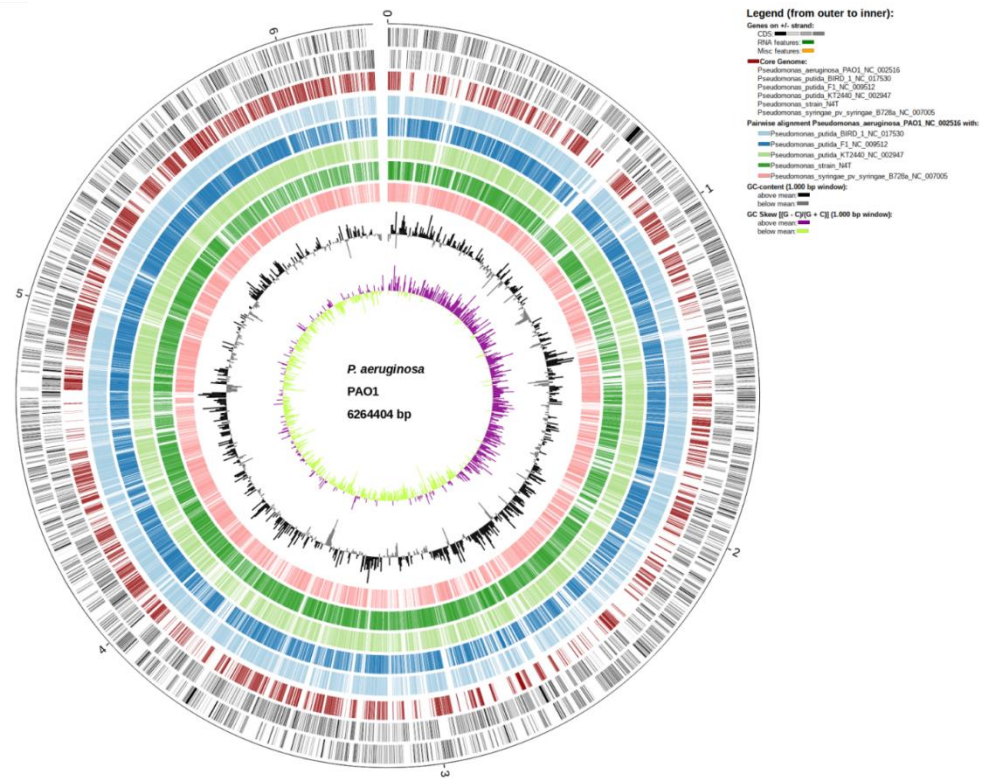


Fig 4.7: Circular plot representation gene clusters present in the all six genomes. The other five genomes were pairwise aligned with the reference *P. aeruginosa* PAO1 genome; the scale represents genome size in KB. Circle ranges from 1 (outer circle) to 10 (inner circle). Circle 1 and 2 represents CDSs present in + and – strand of PAO1. Circle 3 represents the core genome of the six strains used in the study. Then circle 4 to 8 represnts genomes of BIRD1, F1, KT2440, N4T and B728a respectively in different colours. GC content and GC skew are represented in the circle 9 and 10 resepectively.

Using EDGAR tool core genome and pan genome of the six strains were calculated. The pan genome was found to be consists of 11426 genes. And the total number of core genes was found to be 2414 among the six strains. Also, the pan genome and core genome for the four *P. putida* strains were calculated. A total number of 7743 genes were found in the pan genome and 3577 genes were present in the core genome.

4.5 Disscusion

In the present work, the genome of *P. putida* N4T was investigated after sequencing and annotation. Firstly, the general genome features were explored and then different other properties like prediction of secondary metabolite biosynthesis gene clusters, presence of potential plant growth promoting traits and potential genes involved in plant association

were elucidated using different tools. After delving in to these different properties of *P. putida* N4T, a comparative analysis was carried out recruiting genome information of five other *Pseudomonas* strains to understand genomic similarity and variability among each other. This comparative analysis involved both non-pathogenic and pathogenic strains of *Pseudomonas* spp. The presence of 11 secondary metabolite biosynthesis gene clusters were predicted. These gene clusters included different genes that were involved in producing metabolites like pyoverdine, putisolvin, WLIP, lankacidin C and fragin. Another study revealed the presence of 38 different factors that were associated with different functions are responsible plant association of the bacterium. The functions included for plant association included toxins, EPS, detoxification, hormone, adhesion, MDRs, metabolism, movement, pigment, LPS, protease, volatiles and PCWDE. Also, total 3714 genes were linked with plant growth promoting traits of the bacterium. The genes were related to having role in colonizing the plant system, competitive exclusion, stress control, bio-fertilization, phytohormone/plant signal production, bio-remediation, plant immune response stimulation and putative functions. The N4T genome contains diverse important traits encoded by a variety of genes that are responsible for nitrogen fixation, chemotaxis, siderophore production etc. which are well known characteristic of plant associated bacteria. Apart from these the genome also possesss traits that are related to the Type six secretion system, which are involved in different roles including inter-bacterial competition. The Type six secretion system acts as a molecular weapon that delivers toxic effectors to prey cells [48]. The anti-microbial property of Type six secretion systems has been elucidated in many organisms. N4T genome is encomprises with arsenal of genes that are associated with different functions that enables the bacterium to be a successful plant colonizer also these traits bestow different roles on the bacterium. A more detailed analysis of the genome data might unravel new insight into the N4T strain.

The phylogenetic analysis with the five genomes recruited in this study validated that N4T was not closely related to the three non pathogenic strains F1, BIRD-1 and KT2440 as well with the two pathogenic strains PAO1 and B728a. On the contrary F1, BIRD-1 and KT2440 shared lots of similarity. Inclusion of more genome in comparative genomics study will help to better understand the diversity and stratification patterns of *Pseudomonas* species. The rapid advancements in next generation sequencing and availibilty of new genome sequences will help to excavate new findings in this field.

Bibliography

[1] Jensen, L. J., Skovgaard, M., Sicheritz-Pontén, T., Hansen, N. T., Johansson, H., Jørgensen, M. K., Ussery, D. Comparative genomics of four *Pseudomonas* species *Pseudomonas*, *Springer*, pp. 139-164, 2004.

[2] Sun, D., Zhuo, T., Hu, X., Fan, X., & Zou, H. Identification of a *Pseudomonas putida* as biocontrol agent for tomato bacterial wilt disease. *Biological Control, 114*, 45-50, 2017.

[3] Palleroni, N. J. The road to the taxonomy of *Pseudomonas*. *Pseudomonas: genomics and molecular biology*, 1-18, 2008.

[4] Lalucat, J., Mulet, M., Gomila, M., & García-Valdés, E. Genomics in bacterial taxonomy: impact on the genus *Pseudomonas*. *Genes, 11*(2), 139, 2020.

[5] Höfte, M., & Vos, P. D. Plant pathogenic *Pseudomonas* species *Plant-associated bacteria Springer*, pp. 507-533, 2007.

[6] Wiklund, T. *Pseudomonas anguilliseptica* infection as a threat to wild and farmed fish in the Baltic Sea. *Microbiology Australia, 37*(3), 135-136, 2016.

[7] Beaton, A., Lood, C., Cunningham-Oakes, E., MacFadyen, A., Mullins, A. J., Bestawy, W. E., Dalzell, C. Community-led comparative genomic and phenotypic analysis of the aquaculture pathogen *Pseudomonas* baetica a390T sequenced by Ion semiconductor and Nanopore technologies. *FEMS microbiology letters, 365*(9), 2018.

[8] Wasi, S., Tabrez, S., & Ahmad, M. Use of *Pseudomonas* spp. for the bioremediation of environmental pollutants: a review. *Environmental monitoring and assessment, 185*(10), 8147-8155, 2013.

[9] Weller, D. M. *Pseudomonas* biocontrol agents of soilborne pathogens: looking back over 30 years. *Phytopathology, 97*(2), 250-256, 2007.

[10] Girard, L., Lood, C., Höfte, M., Vandamme, P., Rokni-Zadeh, H., van Noort, V., . . . De Mot, R. The ever-expanding *Pseudomonas* genus: description of 43 new

species and partition of the *Pseudomonas putida* group. *Microorganisms*, *9*(8), 1766, 2021.

[11] Gross, H., & Loper, J. E. Genomics of secondary metabolite production by *Pseudomonas* spp. *Natural product reports, 26*(11), 1408-1446, 2009.

[12] Götze, S., & Stallforth, P. Structure elucidation of bacterial nonribosomal lipopeptides. *Organic & Biomolecular Chemistry*, *18*(9), 1710-1727, 2020.

[13] Dejonghe, W., Boon, N., Seghers, D., Top, E. M., & Verstraete, W. Bioaugmentation of soils by increasing microbial richness: missing links. *Environmental Microbiology, 3*(10), 649-657, 2001.

[14] Ye, L., Hildebrand, F., Dingemans, J., Ballet, S., Laus, G., Matthijs, S., . . . Cornelis, P. Draft genome sequence analysis of a *Pseudomonas putida* W15Oct28 strain with antagonistic activity to Gram-positive and *Pseudomonas* sp. pathogens. *PLoS One, 9*(11), e110038, 2014.

[15] Nelson, K., Weinel, C., Paulsen, I., Dodson, R., Hilbert, H., Martins dos Santos, V., . . . Holmes, M. Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environmental Microbiology, 4*(12), 799-808, 2020.

[16] Rojas, A., Duque, E., Mosqueda, G., Golden, G., Hurtado, A., Ramos, J. L., & Segura, A. Three efflux pumps are required to provide efficient tolerance to toluene in *Pseudomonas putida* DOT-T1E. *Journal of bacteriology, 183*(13), 3967-3973, 2001.

[17] Ramos, J. L., Duque, E., Godoy, P., & Segura, A. Efflux pumps involved in toluene tolerance in *Pseudomonas putida* DOT-T1E. *Journal of bacteriology, 180*(13), 3323-3329, 1998.

[18] Roca, A., Pizarro-Tobías, P., Udaondo, Z., Fernández, M., Matilla, M. A., Molina-Henares, M. A., . . . Ramos, J. L. Analysis of the plant growth-promoting properties encoded by the genome of the rhizobacterium *Pseudomonas putida* BIRD-1. *Environmental Microbiology, 15*(3), 780-794, 2013.

[19] Matilla, M. A., Ramos, J. L., Bakker, P. A., Doornbos, R., Badri, D. V., Vivanco, J. M., & Ramos-González, M. I. *Pseudomonas putida* KT2440 causes induced systemic resistance and changes in Arabidopsis root exudation. *Environmental Microbiology Reports*, *2*(3), 381-388, 2010.

[20] Espinosa-Urgel, M., Kolter, R., & Ramos, J.-L. Root colonization by *Pseudomonas putida*: love at first sight. *Microbiology*, *148*(2), 341-343, 2002.

[21] Espinosa-Urgel, M., & Ramos, J.-L. Cell density-dependent gene contributes to efficient seed colonization by *Pseudomonas putida* KT2440. *Applied and environmental microbiology*, *70*(9), 5190-5198, 2004.

[22] Oliver, C., Hernández, I., Caminal, M., Lara, J. M., & Fernàndez, C. *Pseudomonas putida* strain B2017 produced as technical grade active ingredient controls fungal and bacterial crop diseases. *Biocontrol Science and Technology, 29*(11), 1053-1068, 2019.

[23] Ohno, M., Kataoka, S., Numata, S., Yamamoto-Tamura, K., Fujii, T., Nakajima, M., Hasebe, A. Biological control of *Rhizoctonia* damping-off of cucumber by a transformed *Pseudomonas putida* strain expressing a chitinase from a marine bacterium. *Japan Agricultural Research Quarterly: JARQ*, *45*(1), 91-98, 2011.

[24] Andreote, F. D., de Araújo, W. L., de Azevedo, J. L., van Elsas, J. D., da Rocha, U. N., & van Overbeek, L. S. Endophytic colonization of potato (Solanum tuberosum L.) by a novel competent bacterial endophyte, *Pseudomonas putida* strain P9, and its effect on associated bacterial communities. *Applied and environmental microbiology*, *75*(11), 3396-3406, 2009.

[25] Abo-Elyousr, K. A., Abdel-Rahim, I. R., Almasoudi, N. M., & Alghamdi, S. A. Native endophytic *Pseudomonas putida* as a biocontrol agent against common bean rust caused by *Uromyces appendiculatus*. *Journal of Fungi*, *7*(9), 745, 2021.

[26] Kumar, R., Barman, A., Phukan, T., Kabyashree, K., Singh, N., Jha, G., . . . Ray, S. K. *Ralstonia solanacearum* virulence in tomato seedlings inoculated by leaf clipping. *Plant Pathology*, *66*(5), 835-841, 2017.

[27] Singh, N., Phukan, T., Sharma, P., Kabyashree, K., Barman, A., Kumar, R., . . . Ray, S. An innovative root inoculation method to study *Ralstonia solanacearum* pathogenicity in tomato seedlings. *Phytopathology, 108*(4), 436-442, 2018.

[28] Kumar, R., Barman, A., Jha, G., & Ray, S. K. Identification and establishment of genomic identity of *Ralstonia solanacearum* isolated from a wilted chilli plant at Tezpur, North East India. *Current Science*, 1571-1578, 2013.

[29] Phukan, T., Kabyashree, K., Singh, R., Sharma, P. L., Singh, N., Barman, A., . . . Ray, S. K. *Ralstonia solanacearum* virulence in eggplant seedlings by the leaf-clip inoculation. *Phytopathology Research, 1*(1), 1-11, 2019.

[30] Singh, N., Kumar, R., & Ray, S. K. An innovative approach to study *Ralstonia solanacearum* pathogenicity in 6 to 7 days old tomato seedlings by root dip inoculation. *Bio-protocol, 8*(21), 2018.

[31] Singh, N. *Ralstonia solanacearum*: an emerging threat as a mysterious plant pathogen. *Current Science, 120*(1), 20, 2021.

[32] Wu, X., Monchy, S., Taghavi, S., Zhu, W., Ramos, J., & van der Lelie, D. Comparative genomics and functional analysis of niche-specific adaptation in *Pseudomonas putida*. *FEMS microbiology reviews, 35*(2), 299-323, 2011.

[33] Matilla, M. A., Pizarro-Tobias, P., Roca, A., Fernández, M., Duque, E., Molina, L., Segura, A. Complete genome of the plant growth-promoting rhizobacterium *Pseudomonas putida* BIRD-1. *Journal of bacteriology, 193*(5), 1290-1290, 2011.

[34] Stover, C. K., Pham, X. Q., Erwin, A., Mizoguchi, S., Warrener, P., Hickey, M., Lagrou, M. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature, 406*(6799), 959-964, 2000.

[35] Feil, H., Feil, W. S., Chain, P., Larimer, F., DiBartolo, G., Copeland, A., Goltsman, E. Comparison of the complete genome sequences of *Pseudomonas syringae pv. syringae* B728a and pv. tomato DC3000. *Proceedings of the National Academy of Sciences, 102*(31), 11064-11069, 2005.

[36] Matvienko , M. CLC Genomics Workbench. *Plant and Animal Genome. Sr. Field Application Scientist, CLC Bio, Aarhus, DE*, 2015.

[37] Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics, 11*(1), 1-11, 2010.

[38] Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Conesa, A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research, 36*(10), 3420-3435, 2008.

[39] Delcher, A. L., Harmon, D., Kasif, S., White, O., & Salzberg, S. L. Improved microbial gene identification with GLIMMER. *Nucleic Acids Research, 27*(23), 4636-4641, 1999.

[40] Lowe, T. M., & Chan, P. P. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Research, 44*(W1), W54-W57, 2016.

[41] Lagesen, K., Hallin, P., Rødland, E., Stærfeldt, H., Rognes, T., & Ussery, D. RNammer: consistent annotation of rRNA genes in genomic sequences. *Nucleic Acids Res, 35*(9), 3100-3108, 2007.

[42] Beier, S., Thiel, T., Münch, T., Scholz, U., & Mascher, M. MISA-web: a web server for microsatellite prediction. *Bioinformatics, 33*(16), 2583-2585, 2017.

[43] Aziz RK , Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics 9:75, 2008.

[44] Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., Ostell, J. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research, 44*(14), 6614-6624, 2016.

[45] Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., & Kanehisa, M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research, 35*(suppl_2), W182-W185, 2007.

[46] lin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., Van Wezel, G. P., Medema, M. H., & Weber, T. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Research, 49*(W1), W29-W35, 2021.

[47] Dieckmann , MA, Beyvers, S, Nkouamedjo-Fankep RC, Hanel PHG, Jelonek L, Blom J , Goesmann, A. EDGAR3.0: comparative genomics and phylogenomics on a scalable infrastructure. Nucleic Acids Research 49: 185-192, 2021.

[48] Patz , S., Gautam, A., Matthias, B., Ruppel, S., Palenzuela, P. R., & Huson, D. H. (2021). PLaBAse: A comprehensive web resource for analyzing the plant growth-promoting potential of plant-associated bacteria. *bioRxiv*, 2021.

[49] Bernal, P., Allsopp, L. P., Filloux, A., & Llamas, M. A. The *Pseudomonas putida* T6SS is a plant warden against phytopathogens. *The ISME journal, 11*(4), 972-987, 2017.