

**Sequence Analysis of the Genome of the Plant Growth-  
Promoting Bacterium *Pseudomonas putida* UW4**

by

Jin Duan

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Biology

Waterloo, Ontario, Canada, 2012

© Jin Duan 2012

## **Author's declaration**

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

## Abstract

The plant growth-promoting bacterium (PGPB) *Pseudomonas putida* UW4, previously isolated from the rhizosphere of common reeds growing on the campus of University of Waterloo, promotes plant growth in the presence of different environmental stresses, such as flooding, high concentration of salt, cold, heavy metals, drought and phytopathogens. The known mechanisms used by *P. putida* UW4 to promote plant growth include 1-aminocyclopropane-1-carboxylate (ACC) deaminase, indole-3-acetic acid (IAA) synthesis and siderophore production. In this work, the genome sequence of UW4 was obtained by pyrosequencing and the gaps between the contigs were closed by directed PCR. The *P. putida* UW4 genome contains a single circular chromosome that is 6,183,388 bp with a 60.05% G+C content. The bacterial genome contains 5,431 predicted protein-coding sequences that occupy 87.4% of the genome. Nineteen genomic islands were predicted and thirty one complete putative insertion sequences were identified. Genome analyses were conducted in order to better characterize the general features of the UW4 genome. Genes potentially involved in plant growth promotion such as IAA biosynthesis, trehalose production, siderophore production, and acetoin synthesis were identified, which will facilitate a better understanding of the mechanisms of plant-microbe interactions. Moreover, genes that contribute to the environmental fitness of UW4 were also determined including genes responsible for heavy metal resistance such as nickel, copper, cadmium, zinc, molybdate, cobalt, arsenate, and chromate. Central metabolic pathways helped elucidate the physiological roles of diverse metabolites of UW4. Unexpectedly, whole-genome comparison with other completely sequenced *Pseudomonas* sp. revealed that UW4 is more similar to the *fluorescens* group rather than to the *putida* group. More surprisingly, a putative type III secretion system (T3SS) was found in the UW4 genome, and T3SS was thought to be essential for

bacterial pathogenesis. Although putative T3SS was observed in other non-pathogenic *Pseudomonas* spp. previously, this is the first report indicating that a T3SS in a *Pseudomonas* sp. is highly similar to the one from *Salmonella* spp.

## Acknowledgments

Foremost, I would like to thank my supervisors, Dr. Bernard Glick and Dr. John Heikkila, who gave me this opportunity to work on this exciting and challenging project, for their guidance, motivation, patience, and continuous support of my study and research. I could not have imagined having better advisors for my Ph.D study.

I want to thank my committee members, Dr. Trevor Charles, Dr. Brendan McConkey, and Dr. Zoya Leonenko for providing a wealth of knowledge, experience and insightful comments throughout my study. Their valuable feedback helped me to improve the quality of this dissertation in many ways.

Many thanks to my thesis examiners, Dr. Barry Warner and Dr. Joyce Loper for thoroughly reading my thesis and providing constructive comments. Their thoughts and time are very much appreciated.

My sincere thanks also go to Dr. Xiangjun Liu for his valuable and practical advice on genome analyses. Thanks to Owen Woody for his assistance on bioinformatics softwares.

I am grateful to my past and current fellow lab members in Dr. Glick and Dr. Heikkila's lab and my friends: Youai Hao, Zhenyu Cheng, Susanne Vesely, Jiming Song, Wei Jiang, Shimaila Rashid, Jianmin Yan, and Saad Khan, who encouraged me and gave me help whenever I asked for it.

My success in graduate school would not have been possible without continuous support from my family. Thanks to my parents, Mingxing Duan and Caifeng Tian, who inspired, encouraged and always believed in me. Thanks to my husband, Baoan Lin, for always being there for me. Without him I would not have finished my Ph.D. I dedicate this thesis to my baby daughter, Carrie, who was born during the writing of this thesis and brought endless happiness into our family.

# Table of Contents

Authors Declaration .....	ii
Abstract .....	iii
Acknowledgments .....	v
Table of Contents .....	vii
List of Figures .....	x
List of Tables .....	xii
List of Abbreviations .....	xiv
1 Introduction .....	1
1.1 Plant Growth-Promoting Bacteria .....	1
1.1.1 Phosphate Solubilization .....	3
1.1.2 Nitrogen Fixation .....	6
1.1.3 Siderophore Production .....	11
1.1.4 Biocontrol of Phytopathogens .....	14
1.1.5 IAA Production .....	19
1.1.6 Cytokinins and Gibberellins Production .....	21
1.1.7 ACC Deaminase .....	24
1.2 <i>Pseudomonas putida</i> UW4 .....	28
1.3 Bacterial Genome Sequencing .....	31
1.3.1 Sanger DNA Sequencing .....	32
1.3.2 Roche 454 GS FLX Genome Sequencer .....	33
1.3.3 Illumina/Solexa Genome Analyzer .....	35

1.3.4 ABI SOLiD System.....	38
1.3.5 Helicos HeliScope Sequencer .....	41
1.3.6 Pacific BioSciences PACBIO RS .....	44
1.3.7 Polonator G.007 .....	46
1.3.8 Ion Torrent Personal Genome Machine .....	47
1.3.9 Summary .....	49
2 Objectives.....	50
3 Materials and Methods.....	51
3.1 Bacterial Growth and DNA Extraction .....	51
3.2 Whole Genome Pyrosequencing and Sequence Assembly .....	51
3.3 Genome Annotation and Analysis.....	52
3.4 Phylogeny of 16S rRNA genes of <i>Pseudomonas</i> genomes.....	54
3.5 Analysis of Heat Shock Protein Genes and $\sigma^{32}$ promoters.....	54
4 Results .....	55
4.1 General Genome Features .....	55
4.2 Genes Involved in Plant Growth Promotion and <i>P. putida</i> UW4 Lifestyle .....	82
4.2.1 ACC Deaminase .....	82
4.2.2 Siderophores.....	86
4.2.3 IAA Production .....	86
4.2.4 Trehalose .....	94
4.2.5 Acetoin and 2,3-butanediol .....	94
4.2.6 Antimicrobial Compounds and Antibiotics Resistance.....	99
4.2.7 Polyhydroxyalkanoates Biosynthesis.....	104

4.2.8 Degradation of Aromatic Compounds .....	107
4.2.9 Heavy Metal Resistance .....	114
4.3 <i>P. putida</i> UW4 Central Metabolic Pathways and Protein Secretion Systems .....	144
4.4 <i>Pseudomonas</i> Genome Comparisons and Phylogeny .....	164
4.5 16S rRNA Genes Phylogenetic Analysis .....	185
4.6 Heat Shock Protein Genes Phylogenetic and Promoter Analyses.....	193
5 Discussion .....	205
5.1 Taxonomy of <i>P. putida</i> UW4: <i>putida</i> or <i>fluorescens</i> .....	205
5.2 Comparative Genomics of UW4 with Other <i>Pseudomonas</i> .....	207
5.2.1 ACC Deaminase .....	212
5.2.2 Siderophores .....	212
5.2.3 IAA Production .....	213
5.2.4 Trehalose .....	214
5.2.5 Acetoin .....	215
5.2.6 Antimicrobial Compounds .....	215
5.2.7 PHAs Biosynthesis .....	216
5.2.8 Degradation of Aromatic Compounds .....	217
5.3 Type III Secretion System in Non-pathogenic <i>Pseudomonas</i> .....	218
5.4 Heat Shock Proteins and $\sigma^{32}$ Promoter Analysis.....	219
6 Conclusions .....	221
References .....	223
Appendix A. Examples of Primers Used for Gaps Closure .....	294
Appendix B. Examples of Agarose Gel Pictures for Filling the Gaps Between the Contigs.....	299

## List of Figures

Figure 4-1 Circular genome map of <i>P. putida</i> UW4 .....	56
Figure 4-2 Nucleotide sequence of UW4 <i>dnaA</i> upstream region analysis.....	63
Figure 4-3 Genomic islands of <i>P. putida</i> UW4 .....	65
Figure 4-4 <i>P. putida</i> UW4 predicted protein localizations .....	84
Figure 4-5 Pyoverdine synthesis genes in <i>P. putida</i> UW4 .....	87
Figure 4-6 <i>P. putida</i> UW4 pyoverdine II biosynthesis pathway .....	90
Figure 4-7 Amino acids sequence alignment of <i>pvdYII</i> in UW4 and <i>P. aeruginosa</i> .....	92
Figure 4-8 <i>P. putida</i> UW4 IAA biosynthesis pathways .....	95
Figure 4-9 <i>P. putida</i> UW4 trehalose biosynthesis pathways .....	97
Figure 4-10 <i>P. putida</i> UW4 acetoin synthesis pathway.....	100
Figure 4-11 <i>P. putida</i> UW4 4-hydroxybenzoate synthesis pathway .....	102
Figure 4-12 <i>P. putida</i> UW4 PHA biosynthesis and degradation pathways.....	108
Figure 4-13 Benzoate degradation pathway in <i>P. putida</i> UW4.....	110
Figure 4-14 4-hydroxybenzoate and L-quininate degradation pathways in <i>P. putida</i> UW4.....	112
Figure 4-15 Genetic organization of nickel transporters located on the genome of UW4.....	119
Figure 4-16 Multiple sequence alignment of NikA of <i>P. putida</i> UW4 .....	121
Figure 4-17 CueAR copper resistance system in <i>P. putida</i> UW4 .....	124
Figure 4-18 CopABCD copper resistance system in <i>P. putida</i> UW4.....	127

Figure 4-19 CinQARS copper resistance system in <i>P. putida</i> UW4.....	132
Figure 4-20 Multiple sequence alignments of CopRS, PputUW4_2050-2051, and PputUW4_4498-4499.....	137
Figure 4-21 Amino acid sequence alignment of PputUW4_2454 and CopG of KT2440 .....	139
Figure 4-22 Amino acid sequence alignments of CadAR from UW4 and <i>P. putida</i> 06909 .....	141
Figure 4-23 Amino acid sequence alignment of metallothionein in KT2440, Pf0-1, SBW25, and UW4. ....	145
Figure 4-24 Zinc transporter system in UW4 and amino acids sequence alignment of ZnuA between UW4 and <i>E. coli</i> K12.....	147
Figure 4-25 Sequence comparison of ChrA between (A) UW4 and <i>P. aeruginosa</i> (B) UW4 and <i>P.</i> <i>putida</i> KT2440.....	149
Figure 4-26 Genetic organization of <i>oscA</i> in the genome of <i>P. putida</i> UW4 .....	151
Figure 4-27 Schematic overview of metabolic pathways and transport systems in UW4.....	154
Figure 4-28 Phylogenetic tree of 21 different <i>Pseudomonas</i> species.....	166
Figure 4-29 Comparative synteny line plots of the complete six-frame translations of the whole genome sequences of UW4 with other <i>P. fluorescens</i> and <i>P. putida</i> genomes .....	183
Figure 4-30 Maximum likelihood phylogenetic tree of 16S rRNA sequences from completely sequenced <i>Pseudomonas</i> genomes .....	186
Figure 4-31 16S rRNA gene sequences alignment of UW4, KT2440, SBW25, and the type strain of <i>P. putida</i> IAM1236r (D84020) and <i>P. fluorescens</i> IAM12022 (D84013) .....	188
Figure 4-32 HSPs maximum likelihood trees .....	194
Figure 4-33 $\sigma^{32}$ promoter sequence alignment of HSPs.....	203

## List of Tables

Table 1-1 Top Twenty Sequenced Bacterial Genomes as of May 25 <sup>th</sup> , 2012 .....	2
Table 4-1 General Features of <i>P. putida</i> UW4 Genome.....	58
Table 4-2 COG Functional Categories of <i>P. putida</i> UW4.....	59
Table 4-3 <i>P. putida</i> UW4 Pseudogenes .....	60
Table 4-4 <i>P. putida</i> UW4 tRNAs.....	61
Table 4-5 Genomic Islands of <i>P. putida</i> UW4 Predicted by IslandViewer.....	67
Table 4-6 <i>P. putida</i> UW4 IS Elements .....	72
Table 4-7 <i>P. putida</i> UW4 Phage Related CDSs .....	73
Table 4-8 Tandem Repeats Identified in <i>P. putida</i> UW4 .....	74
Table 4-9 <i>P. putida</i> UW4 Protein Localizations.....	83
Table 4-10 Genes Associated With Pyoverdine Synthesis in <i>P. putida</i> UW4 .....	89
Table 4-11 <i>P. putida</i> UW4 Multidrug Efflux Systems.....	105
Table 4-12 Genes Involved In 3-hydroxyphenylpropionate Degradation.....	115
Table 4-13 Genes Potentially Involved In Metal Resistance of <i>P. putida</i> UW4.....	117
Table 4-14 Protein Secretion Systems in <i>P. putida</i> UW4.....	159
Table 4-15 Type IV Pilus Genes in <i>P. putida</i> UW4 .....	165
Table 4-16 Putative Orthologous Relations Between UW4 And Completely Sequenced <i>Pseudomonas</i> genomes.....	168

Table 4-17 Predicted CDSs That Share Similarities With Other <i>Pseudomonas</i> sp.....	169
Table 4-18 Putative Unique CDSs In <i>P. putida</i> UW4 .....	171
Table 4-19 Predicted UW4 CDSs That Share Sequence Similarities To Those In Other Genera Only.....	178
Table 4-20 16S rRNA Genes Sequences Pairwise Identities .....	192
Table 5-1 General Features of the <i>Pseudomonas</i> Genomes .....	208

## List of Abbreviations

2ODDs	2-oxoglutarate dependent dioxygenases
3-HPP	3-hydroxyphenylpropionate
4-Cl-IAA	4-chloroindole-3-acetic acid
ABC	ATP-binding-cassette
ACC	1-aminocyclopropane-1-carboxylate
AcdB	glycerophosphoryl diester phosphodiesterase
<i>acdR</i>	ACC deaminase regulatory gene
<i>acdS</i>	ACC deaminase structural gene
ACT	artemis comparison tool
AHLs	acyl homoserine lactones
AIs	autoinducers
aLRT	approximate likelihood ratio test
APS	adenosine-5'-phosphosulfate
ATs	autotrasporters
BASys	bacterial annotation system
CCD	charge-coupled device
CDS	coding sequence
COG	cluster of orthologous group
CRP	cAMP receptor protein
Cy	cyanine
dATP $\alpha$ S	deoxyadenosine $\alpha$ -thiotriphosphate
ddNTP	dideoxynucleotide

dNTP	deoxyribonucleotide
ECF	extracytoplasmic function
<i>ent</i> -K	<i>ent</i> -kaurene
EPS	exopolysaccharides
FeMoCo	iron molybdenum cofactor
FNR	fumarate and nitrate reduction protein
Fur	ferric uptake regulator
GAs	Gibberellins
GDH	glucose dehydrogenase
GGPP	geranyl geranyl diphosphate
GI	genomic island
GTR	general time-reversible model
HCN	hydrogen cyanide
Hcp	haemolysin coregulated protein
HMA	heavy-metal-associated
HR	hypersensitive response
HSP	heat shock protein
IAA	indole-3-acetic acid
IAAld	indole-3-acetaldehyde
IaaH	IAM hydrolase
IaaM	tryptophan-2-monooxygenase
IAM	indole-3-acetamide
IAN	indole-3-acetonitrile
IBA	indole-3-butyric acid

IpdC	indole-3-pyruvate decarboxylase
IPyA	indole-3-pyruvate
IS	insertion sequence
ISGA	integrative services for genomic analysis
ISR	induced systemic resistance
JA	jasmonic acid
LB	Luria-Bertani
LPS	lipopolysaccharides
Lrp	leucine-responsive regulatory protein
MEME	multiple em for motif elicitation
ML	maximum likelihood
N <sub>2</sub>	dinitrogen
NCBI	National Center for Biotechnology Information
NGS	next generation sequencing
<i>nif</i>	nitrogen fixation
NR	non-redundant
NRPS	non-ribosomal peptide synthetases
<i>oriC</i>	origin of replication
P	phosphorous
PAF	<i>Pseudomonas</i> Agar F
PAP2	phosphatidic acid phosphatase
PCR	polymerase chain reaction
PGM	personal genome machine
PGPB	plant growth-promoting bacteria

PGPR	plant growth-promoting rhizobacteria
PPP	pentose phosphate pathway
PQQ	pyrroloquinoline quinone
PHAs	Polyhydroxyalkanoates
PHB	poly-hydroxybutyrate
Pre-Q <sub>0</sub>	7-cyano-7-deazaguanine
Prodigal	prokaryotic dynamic programming genefinding algorithm
PSB	phosphorous solubilizing bacteria
QS	Quorum sensing
RNAP	RNA polymerase
rrn	ribosomal RNA
SA	salicylic acid
SAM	S-adenosylmethionine
SAR	systemic acquired resistance
SBL	sequencing by ligation
SBS	sequencing by synthesis
Sec	general secretory pathway
SMRT	single molecule real time
SOLiD	support oligonucleotide ligation detection
T1SS	type I secretion system
T2SS	type II secretion system
T3SS	type III secretion system
T5SS	type V secretion system
T6SS	type VI secretion system

TAM	tryptamine
Tat	twin arginine translocation
TCA	tricarboxylic acid
TPS	two-partner secretion system
TSB	tryptic soy broth
TSO	tryptophan side-chain oxidase
VgrG	valine-glycine repeat protein
ZMW	zero-mode waveguide

# 1 INTRODUCTION

The number of sequenced bacterial genomes has grown exponentially in the past decade mainly because of the new and improved Next Generation Sequencing (NGS) technologies. The underlying reasons for sequencing the genome of various bacteria are either because they are highly virulent to humans, animals or plants, or they can be applied to bioremediation or bioenergy production. In fact, as of May 2012, among the 2120 complete microbial genomes available in the National Center for Biotechnology Information (NCBI) genome database (May 25<sup>th</sup>, 2012), the top 20 sequenced bacterial genomes are all pathogenic organisms (Table 1-1), whereas the proportion of plant growth-promoting bacteria (PGPB) genomes is very small. In this work, the genome of a very well studied PGPB, *Pseudomonas putida* UW4, was sequenced using one of the NGS sequencing techniques, 454 pyrosequencing and subsequently finished by directed PCR and primer walking. Then, the genome sequences were annotated and analyzed using various computational tools. Genomic analyses of *P. putida* UW4 provided valuable information towards a comprehensive understanding of the physiology of this microorganism, as well as insights into the molecular mechanisms used by this PGPB to promote plant growth.

## 1.1 Plant Growth-Promoting Bacteria

Plant-microbe interactions can occur at all parts of the plant including spermosphere, rhizosphere, phyllosphere, vascular tissue and endophytic regions (Beattie 2006). Because as much as 40% of plant photosynthates are exuded from plant roots, this makes the rhizosphere a most attractive place for microorganisms (Lynch and Whipps 1990). There are three groups of bacteria that associate with plants. First, most bacteria have no detectable effect on plant growth and development and they are mostly found on plant surfaces (Beattie 2006). Second,

Table 1-1. Top Twenty Sequenced Bacterial Genomes as of May 25<sup>th</sup>, 2012.

Organisms	Number
<i>Escherichia coli</i>	60
<i>Chlamydia trachomatis</i>	59
<i>Bacillus cereus</i>	40
<i>Helicobacter pylori</i>	39
<i>Staphylococcus aureus</i>	33
<i>Salmonella enterica</i>	31
<i>Streptococcus pneumoniae</i>	24
<i>Mycobacterium tuberculosis</i>	21
<i>Bacillus thuringiensis</i>	19
<i>Clostridium difficile</i>	18
<i>Listeria monocytogenes</i>	16
<i>Neisseria meningitidis</i>	16
<i>Streptococcus pyogenes</i>	16
<i>Clostridium beijerinckii</i>	15
<i>Buchnera aphidicola</i>	14
<i>Corynebacterium pseudotuberculosis</i>	14
<i>Rickettsia prowazekii</i>	14
<i>Sulfolobus islandicus</i>	14
<i>Acinetobacter baumannii</i>	13
<i>Corynebacterium diphtheriae</i>	13
<i>Streptococcus suis</i>	13

phytopathogenic bacteria may induce diverse deleterious symptoms and can be found on any part of a plant. Third, mutualistic bacteria may enhance plant growth by providing nutrients and/or inhibiting proliferation of pathogenic bacteria. These beneficial bacteria are generally referred to plant growth-promoting bacteria (PGPB) (Bashan and Holguin 1998). Among the PGPB, bacteria that colonize plant roots are termed plant growth-promoting rhizobacteria (PGPR) (Kloepper and Schroth 1978); they are reputed to exert beneficial effects on plant growth.

PGPB can promote plant growth and development directly and/or indirectly (Glick 1995; Glick et al. 1999). Direct mechanisms involve providing the plants with nutrients when they are insufficient, such as nitrogen, phosphate and iron. Indirect mechanisms include suppression of soil-borne diseases by synthesizing antibiotics and lytic enzymes, competing for nutrients and colonization sites, and inducing systemic resistance. Furthermore, some PGPB are capable of synthesizing phytohormones and/or 1-aminocyclopropane-1-carboxylate (ACC) deaminase, which can modulate phytohormone levels within plants and stimulate plant growth directly or indirectly. Most PGPB are not limited to one mechanism to enhance plant growth, rather they employ multiple strategies to maximize the positive effects with their mutualistic partner.

### ***1.1.1 Phosphate Solubilization***

Phosphorous (P) is an essential macronutrient necessary for plant growth. In soil, P is present at  $\sim 400\text{-}1200 \text{ mg}\cdot\text{kg}^{-1}$  (Rodríguez and Fraga 1999). However, soluble P that is available for plants under most soil conditions is only  $\sim 1 \text{ mg}\cdot\text{kg}^{-1}$ , therefore it is considered as a major constraint to plant growth (Goldstein 1994; Hinsinger 2001; Feng et al. 2004). Insoluble P present in soils is mainly in two groups: mineral and organic forms. Mineral forms are represented by apatite, hydroxyapatite, and oxyapatite, while organic forms consist of inositol phosphate (soil phytate), phosphomonoesters, phosphodiester, and phosphotriesters (Rodríguez

and Fraga 1999; Khan et al. 2007). Because of limited bioavailability of P, large amount of expensive chemical P fertilizers are applied to the soil through high-energy-intensive processes every year (Goldstein et al. 1993). However, at least 75% of the P fertilizers applied to soil are reimmobilized into insoluble forms by metal-cation complexes, thus excess application of P fertilizer is required (Goldstein 1986; Stevenson 1986; Omar 1998; Zaidi et al. 2009).

Many bacteria are known to solubilize P through different mechanisms, thereby improving plant growth. Those genera include *Bacillus* (De Freitas et al. 1997; Toro et al. 1997; Rojas et al. 2001; Sahin et al. 2004), *Burkholderia* (Tao et al. 2008; Jiang et al. 2008; Park et al. 2010), *Enterobacter* (Toro et al. 1997; Sharma et al. 2005), *Erwinia* (Liu et al. 1992), *Methylobacterium* (Jayashree et al. 2011), *Pantoea* (Castagno et al. 2011), *Pseudomonas* (Di Simone et al. 1998; Gulati et al. 2008; Park et al. 2009; Malboobi et al. 2009; Meyer et al. 2011), and *Rhizobium* (Halder et al. 1991; Abd-Alla 1994; Chabot et al. 1996).

The main mechanism used by P solubilizing bacteria (PSB) to solubilize inorganic P is the production and excretion of low molecular weight organic acids, i.e. acetate, citrate, gluconate, ketogluconate, lactate, succinate, etc. (Goldstein 1995; Nahas 1996; Kim et al. 1998; Jones 1998). These organic acids chelate the P-bound cations via hydroxyl and carboxyl groups, resulting in the release of plant-available P into the soil (Bnayahu 1991; Kpombrekou and Tabatabai 1994; Sagoe et al. 1998; Gyaneshwar et al. 2002; Vyas and Gulati 2009). The major organic acid used by PSB to solubilize inorganic P is gluconic acid, which is derived from oxidation of glucose. The reaction is catalyzed by glucose dehydrogenase (GDH) and its cofactor pyrroloquinoline quinone (PQQ) (Buurman et al. 1994; Buch et al. 2008). Several studies with *Pseudomonas* and *Enterobacter* spp. showed that bacterial mutants deficient in GDH activity and/or PQQ production failed in P solubilization processes (Gyaneshwar et al. 1999; Han et al. 2008; De Werra et al. 2009).

Besides the organic acid theory of solubilization of mineral P, it has been proposed that other mechanisms may play important roles as well. Illmer and Schinner (1992) observed the lack of correlation between pH and the amount of solubilized P, and therefore suggested that H<sup>+</sup> excreted from the cytoplasm to the outer surface may help solubilize P directly at the cell surface. In addition, Yi et al. (2008) found that exopolysaccharides (EPS) and organic acids act synergistically on the solubilization of tricalcium phosphate. However, the amount of solubilized P depends on the origin and concentration of EPS in the medium. Further studies are necessary to confirm the mechanism of EPS holding phosphorus.

Organic P in soil comes from plant and animal remains. PSB can mineralize organic P through synthesis of phosphatases, which can hydrolyze phosphoric esters or phosphoanhydride bonds (Rodríguez and Fraga 1999). Phosphatases can be grouped into acid or alkaline, specific or nonspecific (Rossolini et al. 1998; Rodríguez and Fraga 1999). One of the specific phosphatases is phytase, which can sequentially remove P from phytate (*myo*-inositol hexakisphosphate), the most abundant (up to 80%) organic P in soil (Turner et al. 2002). It has been reported that a phytase-negative mutant of PSB did not promote plant growth when its phytase synthesis gene *phyA* was disrupted (Idriss et al. 2002). Furthermore, a phytase gene (*phyC*) from *Bacillus amyloliquefaciens* FZB45 is controlled by the phosphate starvation-induced PhoPR two-component system. Makarewicz et al. (2006) proposed a model to explain how PhoP~P and RNA polymerase (RNAP) activate the transcription of *phyC* during phosphate limitation. The promoter of the *phyC* gene in *B. amyloliquefaciens* FZB45 is E $\sigma^A$  dependent, which usually has an optimal spacing of 17 to 19 bp between the -35 and 10 regions. However, in the case of FZB45, a 21 bp window was observed between the -35 and -10 regions of the *phyC* promoter, indicating improper spacing. When the phosphate concentration is high, unphosphorylated PhoP is not able

to bind to the two PhoP boxes around the -35 region. Therefore,  $E\sigma^A$  RNAP will only bind to either the -35 or -10 region because of the improper spacing, leading to a silent *phyC*. Under phosphate starvation, dimeric PhoP~P bind to the two PhoP boxes at -32 and -49, resulting in binding of  $E\sigma^A$  RNAP at the -10 region due to the formation of a PhoP~P-RNAP protein complex, causing subsequent transcription activation of *phyC* (Makarewicz et al. 2006). More recently, Makarewicz et al. (2008) demonstrated that the expression of *phyC* in FZB45 has a second level of control by AbrB, which is a transition state regulator. The transcription of *phyC* is directly repressed by AbrB during exponential growth, due to the binding of this protein to the two binding sites within the *phyC* promoter region. During the transition from the exponential to the stationary phase, phosphorylated Spo0A, the regulator responsible for the initiation of sporulation, binds to the P2 promoter of *abrB* and lowers its transcription, thus relieving the repression of *phyC* transcription (Makarewicz et al. 2008).

The efficiency of PSB depends on many factors, such as carbon and nitrogen sources, specificity for the host plant, soil pH and type (Kim et al. 1998; Rodríguez and Fraga 1999; Gyaneshwar et al. 2002; Gamalero and Glick 2011). In addition, higher efficiency in stimulating plant growth was observed when PSB were co-inoculated with PGPR having other physiological capabilities (Ray et al. 1981; Azcón-Aguilar et al. 1986; Toro et al. 1997; Rojas et al. 2001; Babana and Antoun 2006; Valverde et al. 2006a; Matias et al. 2009; Bianco and Defez, 2010; Leungvutiviroj et al. 2010; Castagno et al. 2011).

### ***1.1.2 Nitrogen Fixation***

Nitrogen is an essential element in living cells because it is a primary component of all amino acids and nucleic acids. Although 78% of the earth's atmosphere is composed of nitrogen,

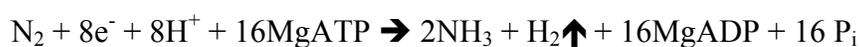
it cannot be used directly in plant growth because it exists as dinitrogen (N<sub>2</sub>). Plants become stunted and chlorotic when the N source is insufficient. As a result, chemical fertilizers are widely used in agriculture at the cost of the environment and, likely, human health (Peplow 2005; Townsend and Howarth 2010). In addition, large amount of natural gas, high pressure and high temperature are required to produce ammonia (Vance 2001). However, at least half of the fertilizer applied in the field is lost due to leaching, run-off into streams, or metabolism by soil microorganisms (Bhattacharjee et al. 2008).

Certain eubacteria (Fischer 1994), cyanobacteria (Sohm et al. 2011) and actinomycetes (Benson, 1988) are able to convert N<sub>2</sub> to ammonia or nitrate, a process called biological nitrogen fixation. Those prokaryotic organisms are defined as diazotrophs, which can be classified into three subgroups including symbiotic, such as *Rhizobium*, *Bradyrhizobium*, *Mesorhizobium*, *Sinorhizobium* (Masson-Boivin et al. 2009); free-living, such as *Klebsiella*, *Azotobacter*, *Clostridium*, *Rhodospirillum*, *Azospirillum*, *Acetobacter*, *Herbaspirillum*, *Azoarcus* (Klipp et al. 2004); and endophytic, such as *Gluconoacetobacter*, *Azospirillum*, *Burkholderia*, *Herbaspirillum*, *Pantoea*, *Pseudomonas*, *Rhanella*, *Klebsiella*, *Enterobacter* (Hill et al. 1983; Paula et al. 1991; Sevilla et al. 2001; Oliveira et al. 2002; 2009; Caballero-Mellado et al. 2004; Iniguez et al. 2004; Reis et al. 2004; Govindarajan et al. 2006; 2008; Montanez et al. 2009; Peng et al. 2009).

The most studied nitrogen fixation system is the symbiosis between rhizobia and their leguminous host plants. The whole nodulation process is highly regulated by complex communications between the plant and the bacteria (Göttfert 1993). Briefly, the plant signals, flavonoids secreted by the roots, bind to the *nodD* gene product from rhizobia, activate the expression of nodulation genes by binding to their promoter regions called a *nod* box, and produce lipochitooligosaccharide signals called Nod factors (Perret et al. 2000). These molecules stimulate the legume root hairs to curl due to the expression of *nodABC* genes. Next, rhizobia

invade the root through the root hair tip where they induce the formation of an infection thread, which is a tube-like structure growing from the root hair tip toward the root cell body, through which the rhizobia enter, travel, and are released into the dividing cortical cells of legume roots, leading to the formation of root nodules. Rhizobial cells released in the plant cells differentiate into their symbiotic form, nitrogen fixing bacteroids.

Inside the root nodule, the nitrogen fixation reaction is catalyzed by the nitrogenase enzyme, which requires high energy (Seefeldt et al. 2009).



The structure of nitrogenase is complex. It has two components and both are oxygen sensitive (Rubio and Ludden 2008; Glick et al. 2010). Component I catalyzes the reduction of  $\text{N}_2$ . It has two identical  $\alpha$ -protein subunits, two identical  $\beta$ -protein subunits, 24 molecules of iron, 2 molecules of molybdenum and an iron molybdenum cofactor (FeMoCo). Component II acts as an electron donor to component I. It consists of two  $\alpha$ -protein subunits, which are different from the ones in component I, and a number of iron molecules. Besides the two components, about 20 additional proteins are required to make nitrogenase fully functional.

The identification of nitrogen fixation (*nif*) genes was through genetic complementation, which was achieved by screening a wild-type library that can restore nitrogen fixation to various mutants of the original strain (Fischer 1994). In the 1970's, the *nif* gene cluster was first isolated and characterized from the free-living diazotroph, *Klebsiella pneumonia* M5a1, which became the model organism for the genetic study of nitrogen fixation (Dixon 2004). The gene cluster contains 20 proteins and they are *nifA*, *nifL*, *nifD*, *nifK*, *nifH*, *nifF*, *nifJ*, *nifQ*, *nifB*, *nifN*, *nifE*, *nifV*, *nifM*, *nifS*, *nifW*, *nifZ*, *nifT*, *nifY*, *nifU*, and *nifX*. NifA is a positive regulatory protein, which turns on the transcription of all of the *nif* operons except its own. NifL is a negative

regulatory protein, which acts as antagonist of the NifA protein in the presence of either oxygen or high levels of fixed nitrogen, therefore it turns off the transcription of all other *nif* genes (Dixon 2004). Rhizobia have a different number of *nif* genes compared to those in *K. pneumonia* M5a1. For example, *Bradyrhizobium* has 15 *nif* genes, *S. meliloti* has 9, and *Rhizobium leguminosarum* bv. *viciae* has 8. The core *nif* genes shared by rhizobia are *nifH* encoding dinitrogenase reductase, *nifDK* encoding the  $\alpha$  and  $\beta$  subunits of dinitrogenase, *nifEN* serving as the molecular scaffold for assembly of the Fe-Mo cofactor, *nifB* involved in FeMoCo synthesis, and *nifA*. It is possible that different organisms have different nitrogenase assembly machinery, on the other hand the missing *nif* products are likely replaced by novel proteins or proteins with novel functions (Masson-Boivin et al. 2009).

The concentration of oxygen is critical in the nitrogen fixation process. It inhibits nitrogenase activity but is required for bacteroid respiration. To solve this problem, *Rhizobium* can be engineered to produce leghemoglobin, which can bind free oxygen tightly. It has been demonstrated that *R. etli* that is transformed with a hemoglobin gene from *Vitreoscilla* sp. has a much higher respiratory rate as well as nitrogenase activity than the wild-type strain (Ramírez et al. 1999; Glick et al. 2010). To further improve the efficiency of nitrogen fixation, it will be preferable to save the energy from producing  $H_2$ . This can be achieved by introducing hydrogenase gene to a diazotroph. Hydrogenase can take up  $H_2$  from the atmosphere and convert it into  $H^+$ . Meanwhile, more ATP will be generated and enter into nitrogen fixation (Báscones et al. 2000).

Nitrogen fixation in free-living bacteria has been demonstrated in *Azotobacter vinelandii*, which has three systems designated as Nif, Vnf and Anf based on three types of nitrogenases. The three nitrogenases are different because of their metal content. For example, the Nif system

needs molybdenum for nitrogenase activity; Vnf system requires vanadium; while the Anf system requires neither (Joerger and Bishop 1988; Joerger et al. 1989 and 1990; Dos Santos and Dean 2011). A total of 82 genes were found in its genome that are known or might be involved in nitrogen fixation and they are organized into at least 16 operons (Setubal et al. 2009). The localization of the nitrogen fixation gene clusters is close to the origin of replication, which implies their high level of expression during active growth (Setubal et al. 2009). Furthermore, expression of specific *nif* genes, such as *nifU*, *nifS*, *nifV* and *nifB* is required for all three systems (Joerger et al. 1986; Kennedy and Dean 1992; Drummond et al. 1996). Recently, it has been observed that ClpX2, a component of an ATP-dependent protease system, regulates protein levels of NifB and NifEN in *A. vinlandii*. *A. vinlandii* carries two copies of *clpX* genes in its genome and *clpX2* is located between the *nifM* and *nifF* genes in the major *nif* cluster. Although it is not essential for the nitrogen fixation process (Jacobson et al. 1989), it was suggested that ClpX2, together with housekeeping protein ClpP, could degrade NifB and NifEM to maintain Fe homeostasis in nitrogen fixing cells (Martínez-Noël et al. 2011).

Compared to free-living diazotrophs, nitrogen-fixing endophytes have the advantage of living inside plants, which facilitates nutrient exchange more readily. After moving from soil to the root area in response to plant signals, endophytes enter at lateral root junctions or at wound sites. They colonize the plant vascular tissues or intercellular spaces and are able to move up to stems and leaves (Doty 2011). *Pseudomonas stutzeri* A1501 is a nitrogen-fixing endophyte isolated from paddy rice and its genome sequence was published in 2008 (Yan et al. 2008). It has a nitrogen fixation island that is 49 Kb in size and consists of 59 genes. This region was probably acquired by horizontal gene transfer because its GC content is higher than the average of the entire genome and the island isn't found in most other *Pseudomonas* species. The organization of *nif* genes in A1501 shares a high degree of similarity with that of *A. vinlandii* (Yan et al. 2008).

Global transcriptional profiling analysis of nitrogen fixation and ammonium repression in A1501 showed that 166 genes were upregulated under nitrogen fixation conditions but downregulated 10 minutes after ammonium shock. Furthermore, a new gene *pnfA*, which was annotated as a hypothetical protein in other diazotrophic microorganisms, was found to have influence on nitrogenase activity (Yan et al. 2010).

Due to the complexity of the nitrogen fixation process, a nondiazotrophic strain is unlikely to fix nitrogen by simply introducing *nif* genes from diazotrophs. However, complete genome sequences of diazotrophs and their whole genome transcriptome analysis will allow us to better understand the gene regulatory network controlling nitrogen fixation, providing a foundation for further improving the efficiency of nitrogen-fixing bacteria.

### ***1.1.3 Siderophore Production***

Iron is essential for nearly all living organisms. It is required in many important biological processes such as electron transport, respiration, photosynthesis, DNA biosynthesis, and nitrogen fixation. Although it is the fourth most abundant element in the earth's crust, it is hardly available because the prevalent form in nature is ferric iron, which has very low solubility ( $K_{sp}$  of  $\text{Fe}(\text{OH})_3=10^{-39}$ ) in aerobic environments (Sandy and Butler 2009). The concentration of soluble iron in soil is around  $10^{-18}$  M at pH 7.4, whereas iron requirements for bacterial growth are in micromolar levels (Sandy and Butler 2009; Glick et al. 2010). Plants are usually not harmed by iron deficiency because they need much lower iron concentration. However, it becomes a problem in calcareous soil, which accounts for 30% of the world's farmland (Hider and Kong 2010). To acquire enough iron for growth, plants and microorganisms have evolved multiple approaches to extract iron from the environments. One of the most common strategies that bacteria and plants use to solubilize and transport iron to the cell is siderophore production.

Siderophores are low molecular weight compounds (~400- to 1,000-dalton) that can chelate and bind  $\text{Fe}^{3+}$  with very high affinity ( $K_d = 10^{-20}$  to  $10^{-50}$  M) (Drechsel and Jung 1998; Glick et al. 2010). Once it binds with  $\text{Fe}^{3+}$ , the complex can be recognized by outer membrane receptor proteins and then transported across cell membranes through periplasmic binding and inner membrane transport proteins, or ABC-type transport proteins (Templeton 2002; Crosa et al. 2004; Brown and Holden 2002). The reduction of  $\text{Fe}^{3+}$  to  $\text{Fe}^{2+}$  can occur either outside the cell by extracellular or membrane bound ferric-chelate reductases, or in the cytoplasm by intracellular ferric-siderophore reductases or hydrolases (Miethke and Marahiel 2007).

Bacterial siderophores are structurally diverse molecules, which can be categorized into three types: catecholates (enterobactin, vibriobactin, yersiniabactin and pyochelin), hydroxamates (alcaligin and desferrioxamine B) and carboxylates (staphyloferrin A and achromobactin). One strain can synthesize more than one type or even mixed-type siderophores (heterobactin B, mycobactin T, petrobactin and aerobactin) (Miethke and Marahiel 2007).

Siderophore biosynthesis and transport have been extensively studied in fluorescent pseudomonads. Under iron-limited conditions, fluorescent pseudomonads can produce green-yellow fluorescent pigments, pyoverdines, which are mixed catecholate-hydroxamate siderophores. Pyoverdines have three components including a fluorescent dihydroxyquinoline chromophore, an acyl side chain bound to the amino group of the chromophore, and a strain-specific peptide chain linked by an amide group bound to the C1 or C3 carboxyl group of the chromophore (Visca et al. 2007a). Most studies on pyoverdines have been carried out with *Pseudomonas aeruginosa* PAO1 and this strain has served as a model organism for understanding pyoverdine synthesis and regulation.

Based on mutations and bioinformatics analyses, most of the pyoverdine synthesis genes (*pvd* genes) in PAO1 are clustered at a single locus on the chromosome, with a few genes located

on a second minor region (Ankenbauer et al. 1986; Hohnadel et al. 1986; Visca et al. 1992). A total number of 32 genes that span about 90 Kb on the chromosome were identified to be associated with the process (Tsuda et al. 1995; Visca et al. 2007b). The pyoverdine biosynthetic enzymes consist of four non-ribosomal peptide synthetases (NRPS) encoded by *pvdD*, *pvdI*, *pvdJ* and *pvdL* (Merriman et al. 1995; Lehoux et al. 2000; Mossialos et al. 2002). NRPS is a large enzyme family that catalyzes the formation of peptide bonds between amino acids that cannot be joined ribosomally (Finking and Marahiel 2004). PvdA is an L-ornithine-N<sup>5</sup>-oxygenase that is responsible for converting L-ornithine to L-N<sup>5</sup>-hydroxyornithine (Ge and Seah 2006), which can be subsequently converted to L-N<sup>5</sup>-formyl-N<sup>5</sup>-hydroxyornithine by the product of *pvdF* (McMorran et al. 2001). Then PvdI and PvdJ can incorporate the substrate L-N<sup>5</sup>-formyl-N<sup>5</sup>-hydroxyornithine into pyoverdine. PvdH is an aminotransferase that catalyzes the formation of L-Dab, which is a substrate of pyoverdine being incorporated by PvdL (Vandenende et al. 2004). The roles of other *pvd* genes are not clear since biochemical studies have not been carried out on them. However, their functions can be predicted based on a bioinformatic analysis. For example, PvdE could be an ATP-binding-cassette (ABC) membrane transporter protein, based on sequence similarity, and PvdG is predicted to be a thioesterase (Visca et al. 2007b).

The major iron responsive regulator in bacteria is the ferric uptake regulator (Fur), which is a repressor protein maintaining homeostasis in cells. Fur is a 17 kDa metalloprotein functioning as a dimer. The N-terminal domain is responsible for DNA binding, while the C-terminal domain helps in dimerisation. When the iron concentration is sufficient, Fur protein bound with Fe<sup>2+</sup> will recognize specific DNA sequences, Fur boxes, of relevant genes, repressing the transcription of those genes, thereby preventing iron uptake (Hantke K, 1981; Andrews et al. 2003). In addition, two small RNAs, PrrF1 (116 nt) and PrrF2 (114 nt), are involved in positive iron regulation of gene expression in *P. aeruginosa* PAO1 (Wilderman et al. 2004). They are

tandemly arranged on the chromosome and are only transcribed under iron-limited conditions because their transcription is repressed by Fur-Fe<sup>2+</sup> when iron is sufficient. PrrF RNAs can base pair at the 5' end of the transcripts and function at the post-transcriptional level by increasing the rate of decay of mRNA of target genes (Wilderman et al. 2004). More recently, it was observed that the two PrrF RNAs could be transcribed as a single transcript, defined as PrrH (325 nt) (Oglesby-Sherrouse and Vasil 2010). PrrH covers the unique intergenic region (95 nt) of PrrF1 and PrrF2, indicating PrrH has more potential mRNA targets (Oglesby-Sherrouse and Vasil 2010).

Although Fur is a key player for iron responsive in the cell, the regulation of pyoverdine biosynthesis and uptake directly related to PvdS, an extracytoplasmic function (ECF) family sigma factor (Cunliffe et al. 1995; Leoni et al. 2000). All pyoverdine biosynthesis and uptake genes have an iron starvation box in their promoter region, which is recognized by PvdS (Rombel et al. 1995; Wilson and Lamont 2000). However, the *pvdS* gene itself is under the control of Fur and it is expressed only in iron-starved bacteria (Ochsner et al. 1995). This regulatory cascade ensures the tight iron-regulated expression of pyoverdine genes.

A number of pieces of evidence have shown that siderophore-producing bacteria may enhance plant iron nutrition. Mutants of siderophore-producing strains lost their ability to promote plant growth and restored the activity by complementation experiments (Kloepper et al. 1991; Sharma et al. 2003; Katiyar and Goel 2004; Vansuyt et al. 2007). Furthermore, PGPR with high siderophore activity have the advantage of suppressing fungal pathogens by competing for the iron source with them, preventing the proliferation of those disease-causing organisms (Gupta et al. 2002; Chaiharn et al. 2009).

#### ***1.1.4 Biocontrol of Phytopathogens***

Some PGPB have biocontrol power against pathogens and the mechanisms include antibiosis, competition for niches and nutrients, production of hydrolytic enzymes, degradation of virulence factors, and induced systemic resistance.

Common antibiotics produced by *Pseudomonas* sp. are cyclic lipopeptides, hydrogen cyanide (HCN), phenazines, pyrrolnitrin, pyoluteorin, and 2,4-diacetyl phloroglucinol. The modes of action of these antibiotics include insertion into cell membranes, inhibition of metalloenzymes especially cytochrome oxidase, inhibition of electron transport, suppression of fungal respiratory chains (Chin-A-Woeng et al. 1998; 2003; Kirner et al. 1998; Nowak-Thompson et al. 1999; Haas and Keel 2003; Haas and Défago 2005; Mavrodi et al. 2006). *Bacillus cereus* is able to produce zwittermycin A and kanosamine (Silo-Suh et al. 1994; Milner et al. 1996). Zwittermycin A has a broad-spectrum activity against Gram-positive, Gram-negative, and eukaryotic microorganisms (Silo-Suh et al. 1998) and the biosynthetic genes as well as the potential regulatory gene were identified by mutation analysis (Emmert et al. 2004). Kanosamine was shown to have very high inhibitory effect on the growth of pathogenic oomycetes. Its production was regulated by plant factors. For example, kanosamine accumulation was greatly enhanced by adding alfalfa seedling exudate to minimal medium (Milner et al. 1996).

Although it is relatively easy to isolate antibiotic-producing PGPB, their antagonist effects on phytopathogens are not persistent all the time. The production of phenazines is strongly influenced by temperature, salinity, and the levels of ferric, phosphate, sulfate and ammonia ions (Van Rij et al. 2004). In addition, antibiotics can be degraded under certain conditions, resulting in less active derivatives that are not as efficient as the original compound (Bottiglieri and Keel 2006).

The plant rhizosphere is a significant carbon sink. One strategy used by PGPB to outcompete phytopathogens is through competition for the niches and nutrients surrounding plant

roots (Compant et al. 2005). Bacteria reach root surfaces by flagella and are guided by chemical attractants to root exudates such as organic acids, amino acids, and sugars (De Weger et al. 1987; Steenhoudt and Vanderleyden 2000; Turnbull et al. 2001a and 2001b; De Weert et al. 2002; Nelson 2004; Welbaum et al. 2004). Some exudates can also be antimicrobial agents. Therefore, if PGPB can produce enzymes to detoxify them, it gives advantage to those PGPB to survive (Bais et al. 2004). Bacterial lipopolysaccharides (LPS) can also contribute to root colonization, although it might be strain dependent (De Weger et al. 1989; Duijff et al. 1997; Dekkers et al. 1998). In addition, several reports have indicated that high bacterial growth rate, vitamin B1, NADH dehydrogenases, and type IV pili are all important to colonization by bacteria (Strom and Lory 1993; Simons et al. 1996; Hahn 1997; Dörr et al. 1998; Dekkers et al. 1998; Steenhoudt and Vanderleyden 2000). In some PGPB, efficient root colonization is attributed to their capability to produce a site-specific recombinase *sss*, which confers the ability of bacteria to adapt to new environments, subsequently resulting in successful plant-microbe interactions (Dekkers et al. 1998 and 2000; Mavrodi et al. 2006).

Some PGPB produce hydrolytic enzymes such as chitinases (E.C. 3.2.1.14) and glucanases (E.C. 3.2.1.39). Chitin is a homopolymer of  $\beta$ -1,4 linked *N*-acetyl-D-glucosamine and is a major structural component of fungal cell walls. Chitinases can hydrolyze chitin, thereby disrupting fungal cell walls directly, meanwhile releasing oligo-*N*-acetyl glucosamines that can activate plant defense response (Ren and West 1992; Podile and Prakash 1996; Gohel et al. 2006; Neeraja et al. 2010). Glucanase is an enzyme that targets  $\beta$ -1,3-glucans in fungal cell walls. Palumbo et al. (2005) reported that *Lysobacter enzymogenes* strain C3 synthesized three extracellular glucanases that together accounted for the total enzyme activity that was detected. Mutations in all three glucanase genes (*gluA*, *gluB*, *gluC*) resulted in reduced biocontrol activity of the strain (Palumbo et al. 2005). Although the two groups of hydrolytic enzymes showed

promise against phytopathogens, it was observed that the synergistic action of chitinases and glucanases was more effective in inhibition of fungal pathogens than either enzyme alone (Sela-Buurlage et al. 1993; Jongedijk et al. 1995; Lawrence et al. 1996; Anfoka and Buchenaue 1997; Vogeli et al. 1988).

Degradation of virulence factors is another mechanism used by PGPB for controlling disease-causing organisms. Fungal pathogens secrete extracellular hydrolytic enzymes such as pectolytic enzymes, cellulases and cutinase, which facilitate the fungal penetration by disrupting plant cell walls (Kishore et al. 2006). *Bacillus megaterium* B 153-2-2 can produce a calcium-dependent extracellular endoproteinase to suppress the activities of the enzymes produced by the pathogen *Rhizoctonia solani* (Bertagnolli et al. 1996). Another biocontrol agent, *P. aeruginosa* GSE 18, was able to reduce rotting and wilting of groundnut by >60% in an *Aspergillus niger*-infested potting mixture via activating host defence responses and inhibiting fungal cell wall-degrading enzymes (Kishore et al. 2006).

In addition, many bacteria express virulence factors only at a high cell density, when the quorum sensing signal molecules reach a certain level. Quorum sensing (QS) is a phenomenon in which bacteria use small signal molecules, called autoinducers (AIs), to monitor their population density and coordinate their gene regulation (Miller and Bassler 2001). The most common AIs in Gram-negative bacteria are acyl homoserine lactones (AHLs). When bacterial cell densities are low, an AI synthase gene synthesizes only a small amount of signal molecules, which diffuse out of the cells and are then diluted in the surrounding area. As the bacterial population increases, the concentration of AIs accumulates. Then they can bind specifically to a transcriptional regulator protein and the activated regulators interact with target DNA sequences so as to enhance or block the transcription of QS controlled genes, including pathogenicity factors (Lazdunski et al. 2004). It has been observed that QS is involved in the synthesis of cell wall degrading enzymes of the

pathogen *Erwinia carotovora* (Andersson et al. 2000). Also, the pathogenicity of *P. aeruginosa* is dependent on its ability to secrete several virulent compounds and degradative enzymes including toxins, proteases and hemolysin, all of which are not expressed until late logarithmic phase of growth when the cell density is high. Studies have shown that QS is essential for the expression of *P. aeruginosa* virulence factors and biofilm formation (Passador et al. 1993; Latifi et al. 1995). Because of the importance of QS in regulating virulence factors in many pathogens, it is desirable for a biocontrol strain to produce QS inhibitors, such as AHL lactonases and acylases, both of which are able to degrade AHL signal molecules (Uroz et al. 2009). In fact, two *Microbacterium testaceum* strains isolated from the leaf surface of *Solanum tuberosum* (potato) plants produced AHL-lactonase which interrupted infection caused by the plant pathogen *Pectobacterium carotovorum* subsp. *carotovorum* (Morohoshi et al. 2009). Moreover, when a lactonase-encoding gene (*aiiA*) from *Bacillus* sp. A24 was introduced into *Pseudomonas fluorescens* P3, which is not an effective biocontrol agent, the transformant significantly reduced symptoms of plant disease caused by *E. carotovora* and *Agrobacterium tumefaciens* (Molina et al. 2003).

In 1991, two research groups independently discovered that induced systemic resistance (ISR) is a mechanism used by PGPB to suppress plant disease (Van Peer et al. 1991; Wei et al. 1991). ISR is a “state of enhanced defensive capacity developed by a plant reacting to specific biotic or chemical stimuli” (Bakker et al. 2007). It can work effectively against different types of pathogens. For example, *P. fluorescens* WCS374r triggered ISR and protected radish from Fusarium wilt, resulting in a 40% increase in yield (Leeman et al. 1995). ISR is phenotypically similar to systemic acquired resistance (SAR), which develops when plants successfully activate their defense mechanism in response to primary infection by necrotizing pathogen (Sticher et al. 1997). Usually SAR requires endogenous or exogenous salicylic acid (SA) as a stimulation signal, whereas ISR needs ethylene or jasmonic acid (JA) to trigger the pathway (Gaffney et al.

1993; Sticher et al. 1997; Pieterse et al. 1998). Bacterial traits that can trigger ISR have been extensively studied in *Pseudomonas* sp. Several determinants include flagella (Meziane et al. 2005), lipopolysaccharides (Van Peer and Schippers 1992; Leeman et al. 1995; Van Wees et al. 1997; Meziane et al. 2005), siderophores (Maurhofer et al. 1994; Leeman et al. 1996; Audenaert et al. 2002; Meziane et al. 2005; Ran et al. 2005) and 2,4-diacetylphloroglucinol (Iavicoli et al. 2003; Siddiqui and Shoukat 2003). Some PGPB can have several determinants to activate ISR, making it harder to identify the bacterial metabolites using a mutational approach. However, this redundancy ensures that ISR occurs even if one determinant fails to elicit ISR. Furthermore, if the different traits were differentially regulated, it would make a particular PGPB a better biocontrol agent when it is applied in the field (Bakker et al. 2007).

### **1.1.5 IAA Production**

Auxins were the first growth hormone discovered in plants and they play vital roles in cell elongation, plant tropisms, vascular differentiation, floral and fruit development, lateral root formation, and determination of root and shoot architecture (Teale et al. 2006; Taiz and Zeiger 2010; Pliego et al. 2011). Examples of naturally occurring auxins in plants are indole-3-acetic acid (IAA), 4-chloroindole-3-acetic acid (4-Cl-IAA), and indole-3-butyric acid (IBA). Because IAA has been found in all plants, it is considered the most abundant and physiologically important auxin.

Many bacteria can synthesize IAA and it was estimated that 80% of rhizosphere bacteria are capable of producing this hormone (Patten and Glick 1996). Those microorganisms include pathogens such as *A. tumefaciens* (Liu et al. 1982), *Agrobacterium rhizogenes* (Costacura and Vanderleyden 1995), *Erwinia herbicola* (Manulis et al. 1991), and *Pseudomonas syringae* (Mazzola and White 1994), as well as PGPB such as *Azospirillum brasilense* (Barbieri et al.

1986), *P. putida* (Patten and Glick 2002), *Rhizobium* (Badenochjones et al. 1983), *Enterobacter cloacae* (Saleh and Glick 2001), *Acetobacter diazotrophicus* (Fuentes-Ramirez et al. 1993), and *Bradyrhizobium japonicum* (Sekine et al. 1988). IAA secreted by microorganisms can contribute to a plant's endogenous auxin pool, thereby altering the plant growth and development processes related to auxin signaling.

Biosynthesis of IAA has been extensively studied in bacteria and five tryptophan dependent pathways were identified. They are the indole-3-acetamide (IAM) pathway, the indole-3-pyruvate (IPyA) pathway, the tryptamine (TAM) pathway, the tryptophan side-chain oxidase (TSO) pathway, and the indole-3-acetonitrile (IAN) pathway (Patten and Glick 1996; Spaepen et al. 2007). The enzymes and steps involved are as follows:

- In the IAM pathway, tryptophan-2-monooxygenase (IaaM) catalyzes the conversion of tryptophan to IAM. Then IAM is converted to IAA by IAM hydrolase (IaaH).
- In the IPyA pathway, tryptophan is first converted to IPyA by an aminotransferase. Next, IPyA is decarboxylated to indole-3-acetaldehyde (IAAld) by indole-3-pyruvate decarboxylase (IpdC). Then IAA is produced by oxidation of IAAld.
- Three steps are involved in the TAM pathway. First, tryptophan is converted to TAM by tryptophan decarboxylase. Next, TAM is converted to IAAld by amine oxidase. In the last step, IAAld is converted to IAA by IAAld dehydrogenase.
- Different from the IPyA and TAM pathways, in the TSO pathway, tryptophan is directly converted to IAAld, which is subsequently oxidized to IAA.

- In the IAN pathway, the precursor of IAA, IAN can be produced either through tryptophan via indole-3-acetaldoxime then by glucobrassicin or directly from anthranilic acid independent of tryptophan. Next, IAN is converted to IAA by nitrilase, or first to IAM by nitrile hydratase then to IAA by amidase.

Both phytopathogenic bacteria and PGPB are capable of synthesizing IAA. The impact of the exogenous IAA on plants can be deleterious or beneficial, and the effect greatly depends on the IAA concentration (Evans et al. 1994; Xie et al. 1996; Persello-Cartieaux et al. 2001; 2003). In addition, the sensitivity of the host plants may play a role in the effect of IAA as well (Kucey 1988; Dubeikovsky et al. 1993; Persello-Cartieaux et al. 2001). Although it was suggested that pathogenic organisms synthesize IAA through the IAM pathway and beneficial bacteria produce IAA via the IPyA pathway, several studies have proved this is not necessarily the case (Vasanthakumar and McManus 2004; Theunis et al. 2004). Many PGPB have shown plant growth enhancement by producing IAA. They facilitate lateral and adventitious root growth, thereby enhancing plant mineral uptake and root exudation (Dobbelaere et al. 1999; Lambrecht et al. 2000; Steenhoudt and Vanderleyden 2000). However, the observed beneficial effects on plant growth are usually the consequences of multiple mechanisms used by PGPB rather than the impact of IAA production alone (Xie et al. 1996; Bashan and Holguin 1997).

### ***1.1.6 Cytokinins and Gibberellins Production***

Cytokinins are regulators of plant cell division and are present in roots, stems, leaves, flowers, fruits and seeds (Garcia de Salamone et al. 2006). The first natural cytokinin, discovered in the immature endosperm of *Zea mays*, was called zeatin (*trans*-6-(4-hydroxy-3-methylbut-2-enylamino)purine). Zeatin represents the major active compound of cytokinin molecules in higher plants (Letham 1963 and 1973). Zeatin can exist in *cis* or *trans* configuration due to a

double bond in the side chain, and the two forms can be interconverted by zeatin isomerase found in some plants (Taiz and Zeiger 2010). The configuration of zeatin is species specific and both forms can play important roles in plant cell division and development (Taiz and Zeiger 2010).

Cytokinins are able to induce plant cell division when auxin is present and the molar ratios of auxin and cytokinin determine whether roots or shoots differentiate from callus cultures. For instance, high auxin concentrations promote root formation whereas high cytokinin concentrations promote shoot differentiation (Pliego et al. 2011). Root tips and developing seeds are the main sites of cytokinin biosynthesis. Cytokinins are transported from these sites to other plant tissues where they control diverse aspects of development such as senescence inhibition, cell growth, secondary metabolism, and chloroplast development (Neuman et al. 1990; Nooden and Letham 1993; Garcia de Salamone et al. 2006; Pliego et al. 2011).

Many bacteria are able to produce cytokinins, thereby contributing to the plant cytokinin pool. These bacteria include *Agrobacterium*, *Erwinia*, *Pseudomonas*, *Paenobacillus*, *Azotobacter*, *Azospirillum*, *Bacillus*, and *Rhizobium* (Phillips and Torrey 1972; Gonzalez-Lopez et al. 1986; Akiyoshi et al. 1987; Nieto and Frankenberger 1989; Strzelczyk et al. 1994; Lichter et al. 1995; Timmusk et al. 1999; Ortíz-Castro et al. 2008). Many reports have demonstrated that cytokinin-producing PGPB can promote plant growth. For example, mutants in cytokinin biosynthesis were unable to promote growth of wheat and radish compared to the wild type strain *P. fluorescens* G20-18 (García de Salome 2000). In another study, the levels of different cytokinins in *Glycine max* were compared after inoculation with *Bradyrhizobium* and it was observed that a high ratio of cytokinin:auxin in the soybean roots facilitated nodule initiation (Caba et al. 2000). Recently, it was shown that a strain of *B. megaterium* that was able to promote growth of *Arabidopsis thaliana* and *Phaseolus vulgaris* seedlings had a much lower activity when inoculated to *A.*

*thaliana* mutants lacking cytokinin receptors, indicating that the plant growth promotion by this strain is dependent on cytokinin signaling (Ortíz-Castro et al. 2008).

Gibberellins (GAs) are plant phytohormones that affect many developmental and physiological processes in higher plants, including seed germination, cell division, stem elongation, root growth, root hair abundance, pollen tube growth, and flower induction (Pharis and King 1985; Tanimoto 1987; Fulchieri et al. 1993; Hedden and Kamiya 1997; King and Evans 2003). They are diterpenoids that are formed from four isoprenoid units each consisting of five carbons. GAs can be produced by plants, fungi, and bacteria (Taiz and Zeiger 2010). To date, at least 136 GAs were identified from higher plants, although not all of them have biological activity because some of them are either metabolic precursors or deactivation products (MacMillan 2002; Taiz and Zeiger 2010). A couple of GAs such as GA<sub>1</sub>, GA<sub>3</sub>, GA<sub>4</sub>, GA<sub>9</sub>, GA<sub>19</sub> and GA<sub>20</sub> were found in various PGPB including *B. japonicum*, *Azospirillum lipoferum*, *Rhizobium phaseoli*, and *Bacillus* spp. (Bottini et al. 1989; Manero et al. 1996; Piccoli et al. 1996 and 1997; Atzorn et al. 1988; Gutiérrez-Mañero et al. 2001; Boiero et al. 2007).

Biosynthesis of gibberellin in higher plants starts with the cyclization of a C<sub>20</sub> precursor, geranyl geranyl diphosphate (GGPP), yielding *ent*-kaurene (*ent*-K) catalyzed by copalyl diphosphate synthase (E.C. 5.5.1.13) and *ent*-kaurene synthase (E.C. 4.2.3.19). Then *ent*-K is converted to gibberellin by a set of oxidative reactions by membrane-related cytochrome P450 monooxygenases and 2-oxoglutarate dependent dioxygenases (2ODDs) (Bottini et al. 2004). In fungi, the production of *ent*-K is catalyzed by a bifunctional diterpene synthase instead of by two enzymes as in plants. Also, different enzymatic steps are involved after *ent*-kaurenoic acid is formed (Morrone et al. 2009). A cluster of six genes including gibberellin-specific GGPP synthase gene, *ent*-K synthase gene, and four cytochrome P450 monooxygenase genes are involved in gibberellin biosynthesis in *Gibberella fujikuroi* (Mende et al. 1997; Linnemannstöns

et al. 1999; Tudzynski et al. 2003). In bacteria, the first potential gibberellin biosynthetic pathway was reported in *B. japonicum* USDA110 (Morrone et al. 2009). An operon consisting of genes encoding a GGPP synthase, two diterpene synthases, and three P450 monooxygenases are involved in GA biosynthesis in this strain. Sequence comparisons of the diterpene synthases from plants, fungi and bacteria suggest they might have a common origin, although the diterpene synthases in *B. japonicum* are significantly smaller than the ones found in plants and fungi. In addition, bacteria only require three P450 monooxygenases instead of four in fungi and GA biosynthesis is 2ODDs independent. Therefore, GA biosynthesis in bacteria represents a third independent pathway (Morrone et al. 2009).

It was demonstrated that GA-producing *Bacillus* sp. were able to promote growth of red peppers (Joo et al. 2004) and the promoting effect greatly depended on high efficiency of bacterial root colonization (Joo et al. 2005). In another study, *Acinetobacter calcoaceticus* SE370 was isolated from soil based on GA production and its culture filtrates promoted growth of several plants such as cucumber, Chinese cabbage and crown daisy (Kang et al. 2009). Currently, the mechanism of GA involved in plant growth stimulation is still obscure, probably due to the relative paucity of information on the biosynthesis of GA in bacteria. However, this situation will change significantly as more GA biosynthesis genes are identified from completely sequenced bacterial genomes.

### **1.1.7 ACC Deaminase**

ACC deaminase (E.C. 3.5.99.7) is an enzyme that is capable of hydrolyzing ACC, the immediate precursor of ethylene in plants, to  $\alpha$ -ketobutyrate and ammonia. It was first isolated from *Pseudomonas* sp. strain ACP in 1978 (Honma and Shimomura 1978). Since then, ACC

deaminase has been found in a yeast (*Hansenula saturnus*) (Minami et al. 1998), a fungus (*Penicillium citrinum*) (Honma 1993), and in a number of bacterial strains (Klee et al. 1991; Sheehy et al. 1991; Jacobson et al. 1994; Glick et al. 1995; Campbell and Thomson 1996; Burd et al. 1998; Jia et al. 1999; Belimov et al. 2001 and 2005; Mayak et al. 2001; Babalola et al. 2003; Ghosh et al. 2003; Ma et al. 2003; Dey et al. 2004; Uchiumi et al. 2004; Hontzeas et al. 2005; Blaha et al. 2006; Madhaiyan et al. 2006). Many of these bacteria were identified by their ability to grow on minimal medium containing ACC as the sole nitrogen source.

ACC deaminase is an important trait in a PGPB because bacteria containing ACC deaminase can lower plant ethylene levels when the plants are under different environmental stresses, thereby facilitating plant growth (Glick et al. 1998). Ethylene has long been recognized as a hormone that controls plant responses under growth-limiting conditions (Abeles et al. 1992; Morgan and Drew 1997). It has been proposed that ethylene is produced in two peaks in response to environmental stresses (Stearns and Glick 2003; Pierik et al. 2006). The first peak is small and usually occurs a few hours after the stress. This small peak is thought to be beneficial because it may turn on the transcription of genes that are responsible for plant protection (Van Loon and Glick 2004; Van Loon et al. 1997; Glick 2004). The second peak is much larger and occurs one to three days after the stress. This second peak typically causes visible damage to the plant, such as senescence, chlorosis and abscission (Glick et al. 2007a).

A model has been proposed to explain the mechanism of action of ACC deaminase in plant growth promotion (Glick et al. 1998 and 2007a). Generally, rhizobacteria attached to the seeds or roots of a developing plant can synthesize and secrete IAA in response to tryptophan and other small molecules in plant exudates (Patten and Glick 1996; 2002; Bayliss et al. 1997; Penrose and Glick 2001). The IAA produced by the PGPB and plant can stimulate plant growth and development or induce the synthesis of ACC synthase, which converts S-adenosyl-

methionine (SAM) to ACC. This ACC can be subsequently converted to ethylene by ACC oxidase but it can be also exuded to the rhizosphere and taken up by PGPB. Bacteria containing ACC deaminase can hydrolyze the ACC to  $\alpha$ -ketobutyrate and ammonia. Because the amount of ACC outside the roots or seeds is reduced, to maintain the equilibrium between internal and external levels, more ACC will be secreted, resulting in a lower amount of ACC available for oxidation to ethylene. By decreasing the level of ACC inside the plant cells, the inhibitory effect of stress ethylene on root elongation is reduced (Glick 1995). Additionally, IAA can activate the transcription of ACC synthase, leading to more ethylene production (Kim et al. 1992; Kende 1993; Kende and Zeevaart 1997). On the contrary, ethylene inhibits IAA transport and signal transduction; therefore this feedback loop will eventually limit the amount of ethylene levels when the plant is under environmental stress (Burg and Burg 1966; Morgan and Gausman 1966; Suttle 1988; Prayitno et al. 2006). With the presence of ACC deaminase-containing PGPB, the feedback loop breaks because the ACC produced is degraded in the bacteria, thus the repression of auxin response factor synthesis by ethylene is relieved, resulting in more IAA production and plant growth promotion (Glick et al. 2007b).

ACC deaminase is usually present in bacteria at a low level before it is induced and the induction of enzyme activity is a slow and complex process. Therefore, the small peak of ethylene still occurs right after environmental stresses, which will induce a defense response in the plant. As the concentration of ACC increases, bacterial ACC deaminase is induced, thus the second deleterious ethylene peak may be reduced significantly by ACC deaminase activity (Glick et al. 2007a).

ACC deaminase genes have been cloned and characterized from a number of different soil bacteria, such as *Pseudomonas* sp. strains 6G5 and 3F2 (Klee et al. 1991; Klee and Kishore

1992), *Pseudomonas* sp. strain 17 (Campbell and Thomson 1996), *Pseudomonas* sp. strain ACP (Sheehy et al. 1991), *E. cloacae* CAL2 and *P. putida* UW4 (Glick et al. 1995; Shah et al. 1998), *Pseudomonas brassicacearum* Am3 (Belimov et al. 2007), *R. leguminosarum* bv. *viciae* (Ma et al. 2003; Duan et al. 2009), as well as from the yeast, *H. saturnus* (Minami et al. 1998), and fungus, *P. citrinum* (Jia et al. 1999). In all of the above-mentioned strains, the ACC deaminase structural gene (*acdS*) was shown to have the enzyme activity. Putative *acdS* were identified in *Escherichia coli*, various Archaeobacteria and numerous other bacteria. However, none of these organisms has been demonstrated to have ACC deaminase activity. With more complete sequenced genomes available, more putative *acdS* genes may be discovered, but their function has to be confirmed by biochemical assays as other enzymes have similar sequences and may be mistaken for ACC deaminase (Todorovic and Glick 2008).

The regulation of *acdS* has been studied in several PGPB. In *P. putida* UW4, many genetic elements are involved in this complex regulation process (Grichko and Glick 2000; Li and Glick 2001; Cheng et al. 2008). The *acdR* gene encoding a leucine-responsive regulatory protein (Lrp) is the regulatory gene located 5' upstream of the *acdS* and transcribed in the opposite direction of the *acdS*. In between them, there is a 165 bp intergenic region containing possible binding sites for Lrp (an LRP box), which overlaps with the *acdR* promoter region, a fumarate and nitrate reduction protein (an FNR box), which is located within one of the *acdS* promoters, and a cAMP receptor protein (a CRP box) overlapped with the second *acdS* promoter. Based on what is known about other LRP proteins, AcdR (=LRP) functions as an octamer (Leonard et al. 2001) and when it binds to an LRP box, it either activates or represses the transcription of the target gene. Studies have shown that AcdR can bind to a complex of ACC bound to AcdB, a glycerophosphoryl diester phosphodiesterase. The complex may bind to either an FNR or CRP box under anaerobic or aerobic conditions, respectively. The binding to the FNR

or CRP boxes facilitates *acdS* transcription. One of the products generated by ACC deaminase reaction is  $\alpha$ -ketobutyrate, which is a metabolic precursor of leucine. With increasing amounts of leucine levels in the cell, the Lrp octamer is dissociated, resulting in a silencing of *acdS* gene expression.

Many studies have shown that bacteria containing ACC deaminase promote plant growth under different environmental stresses including at the presence of phytopathogens, high concentration of salt, drought, flooding, heavy metals, organic compounds (Burd et al. 1998; 2000; Belimov et al. 2001; 2005; Wang et al. 2000; Grichko and Glick 2001; Nie et al. 2002; Glick 2003; Huang et al. 2004a; 2004b; Mayak et al. 2004a; 2004b; Saleh et al. 2004; Reed and Glick 2005; Farwell et al. 2006; Saravanakumar and Samiyappan 2006; Cheng et al. 2007; Farwell et al. 2007; Hao et al. 2007; Rodriguez et al. 2008; Gamalero et al. 2009). Moreover, the *acdS* gene has become a widely used genetic marker for screening PGPB functions.

## **1.2 *Pseudomonas putida* UW4**

*P. putida* UW4 is a well-studied PGPB that was isolated from the rhizosphere of reeds in Waterloo, Ontario (Glick et al. 1995). This strain has the ability of utilizing ACC as a sole source of nitrogen and promoting canola seedling root elongation in growth pouches under gnotobiotic conditions (Glick et al. 1995).

At the beginning, UW4 was designated *Pseudomonas* sp. on the basis of growing on *Pseudomonas* Agar F (PAF) selective medium and siderophore production. Later on, the name of the bacterium was changed to *Enterobacter cloacae* UW4 based on the results of fatty acid analysis (Shah et al. 1998). However, after sequencing a partial 16S ribosomal RNA gene from UW4, the results indicated that it is *Pseudomonas putida* (Hontzeas et al. 2005), and the genus

and species were further confirmed by thorough metabolic profiling (MicroLog System, Release 4.0).

Under the microscope, *P. putida* UW4 cells are straight rods that are 2-4  $\mu\text{m}$  in length and 2  $\mu\text{m}$  in width. They are Gram-negative, motile, non-spore forming and are routinely grown in Tryptic Soy broth (TSB) at 30°C in the lab. UW4 is a psychrotroph that can proliferate at 4°C. On PAF medium plates, the colonies are round in shape with smooth regular borders and green-yellowish color after 24 hours incubation at 25°C or 30°C. Antibiotic resistance tests showed that this strain can grow in the presence of ampicillin (128  $\mu\text{g/ml}$ ), erythromycin (64  $\mu\text{g/ml}$ ), novobiocin (256  $\mu\text{g/ml}$ ), but is sensitive to kanamycin (4  $\mu\text{g/ml}$ ) and tetracycline (8  $\mu\text{g/ml}$ ) (unpublished data).

In 1998, the gene encoding ACC deaminase was isolated from UW4 and a putative ribosomal binding site (AAGGA) at -13 as well as a potential transcription termination site (GTAGGAGCGGGC) at 37 nucleotides downstream from the stop codon were found. Southern hybridization results indicated that there is only one copy of *acdS* gene in UW4. When the ACC deaminase gene of UW4 was introduced into *E. coli* DH5 $\alpha$ , *P. putida* ATCC 17399 and *P. fluorescens* ATCC 17400, the gene was expressed and the transformed strains were able to promote root elongation of canola seedlings (Shah et al. 1998). Furthermore, when the *acdS* gene in UW4 was disrupted, the strain lost its capability to promote root elongation (Li et al. 2000). In order to understand how the *acdS* gene is regulated, the gene upstream of the *acdS* and the intergenic region between the two genes was studied (Grichko and Glick 2000; Li and Glick 2001; Cheng et al. 2008). The results revealed a fairly complicated transcriptional regulatory network involving several components; the regulation of the *acdS* gene in UW4 has been discussed in the previous section.

From then on, a number of studies focused on the impact of the *acdS* gene of UW4 on plant growth under different environmental stresses. For example, when the *acdS* gene and its regulatory region was introduced into a biocontrol strain, *P. fluorescens* strain CHA0, the transformed strain showed improved ability to protect cucumber against *Pythium* damping-off, and potato tubers against *Erwinia* soft rot under the conditions used (Wang et al. 2000). Furthermore, transgenic tomato plants expressing the UW4 ACC deaminase showed reduced symptoms of *Verticillium* wilt (Robison et al. 2001). And, in the presence of heavy metals such as Cd, Co, Cu, Mg, Ni, Pb, or Zn, ACC deaminase-producing tomato and canola plants showed less deleterious effects of the metals on plant growth compared to the non-transgenic plants (Grichko et al. 2000; Nie et al. 2002; Stearns et al. 2005). In another study, under flood conditions, tomato plants inoculated with UW4 *acdS*-containing bacterial strains showed a significant tolerance to flooding stress (Grichko and Glick 2001).

In addition to the studies on *acdS*-transformed bacteria and transgenic plants, the influence of the strain *P. putida* UW4 on the plant growth under various stresses was also examined. For instance, UW4 has been shown to enhance plant growth in the presence of flooding (Farwell et al. 2007), heavy metals (Farwell et al. 2006), cold (Cheng et al. 2007), high concentration of salt (Cheng et al. 2007), and phytopathogens (Hao et al. 2007; Toklikishvili et al. 2010).

In an effort to better understand the interaction between plants and free-living PGPB, the proteomes of wild type UW4 and its *acdS* minus mutant were investigated upon treatment with canola root exudates (Cheng et al. 2009a). Levels of 72 proteins changed significantly, and putative predictions of many of the proteins are related to nutrient transport, cell envelope synthesis, and transcriptional/translational regulation. Functional analysis of four proteins, outer membrane protein F, peptide deformylase, transcription regulator Fis family protein, and an

uncharacterized protein, confirmed that they are key proteins involved in plant-microbe interactions (Cheng et al. 2009a). Furthermore, when UW4 was exposed to 2 mM Ni, bacterial proteins involved in heavy metal detoxification such as stress adaptation, anti-oxidative stress, and heavy metal efflux proteins were up-regulated significantly (Cheng et al. 2009b). More recently, Cheng et al. (2011) analyzed the protein expression profile of canola plants inoculated with UW4 or its *acdS* minus mutant under salinity stress. As expected, many of the differentially expressed proteins in the plants are related to salt stress tolerance. Moreover, it was observed that the enzyme ACC deaminase played an important role in the salt response of canola plants. For example, the expression of proteins involved in photosynthesis decreased to a lesser extent if the plants were treated with wild type UW4 prior to salt exposure, and the plants were healthier due to the lowered stress ethylene levels (Cheng et al. 2011).

In 2009, a proteome reference map of *P. putida* UW4 was published (Cheng et al. 2009c). The map represents 275 different proteins of UW4 and the bacterium was cultured under the typical conditions used in the lab (aerobically to late-log phase in TSB at 30 °C). Although this map only represents ~ 5% of the total number of proteins synthesized by UW4, it should facilitate future proteomic studies with this bacterium.

### **1.3 Bacterial Genome Sequencing**

Since the first bacterial genome, *Haemophilus influenzae*, was fully sequenced in 1995, over 1700 complete bacterial genome sequences have been determined. DNA sequencing technology has dramatically improved from the first generation, automated Sanger DNA sequencing, which dominated this field for almost two decades, to the current NGS platforms. The newer technology dramatically reduces both the time and cost of DNA sequencing, making it

possible for a small lab to completely sequence the genome of their favorite bacterium. With the enormous amount of information obtained from whole genome sequencing, scientists can readily address a wide range of biological questions that were hitherto beyond their capabilities (Duan et al. 2010).

### **1.3.1 Sanger DNA Sequencing**

The Sanger DNA sequencing technique has been an important method in sequencing bacterial genomes. The sequencing chemistry is based on the use of the DNA chain terminator dideoxynucleotide (ddNTP), which is a molecule lacking a hydroxyl group at the 3' carbon of the deoxyribose sugar (Sanger et al. 1977). During DNA synthesis, an incoming deoxyribonucleotide (dNTP) can form a phosphodiester bond between its 5'  $\alpha$ -phosphate group and the 3' hydroxyl group of the last nucleotide. However, if a dideoxynucleotide is incorporated at the end of the growing strand, DNA chain growth terminates. The four dideoxynucleotides used in Sanger sequencing are labeled with four fluorescent dyes with each dye representing a particular nucleotide (Smith et al. 1986). The dye-labeled ddNTPs are added into the reaction mixture containing single-stranded DNA template, primer, DNA polymerase and all four dNTPs. The polymerase chain reaction (PCR) products are separated by capillary electrophoresis according to their masses. Each fluorescent dye emits light following its activation by a laser at the end of the capillary. Therefore, the DNA sequence can be determined by the order of the fluorescent signals (Swerdlow et al. 1990).

Using Sanger DNA sequencing chemistry to sequence the entire genome of an organism is comprised of three major steps: DNA library preparation, template purification and DNA sequencing. For shotgun de novo sequencing, DNA is randomly fragmented to generate small (2

Kb) and large (15~20 Kb) fragments, which are subsequently cloned into a high-copy-number plasmid. The plasmids are then used to transform *E. coli* cells. After transformation, resultant colonies are transferred into either 96- or 384-well plates. Plasmid purification occurs directly on the plates. Then the PCR-based DNA sequencing is performed (Fleischmann et al. 1995). Because the genomic DNA is randomly broke down, longer DNA sequences can be obtained by aligning and assembling short sequence fragments based on partial sequence overlaps.

After decades of improvement of this methodology, sequences of up to ~1 Kb DNA can be obtained by Sanger sequencing with an accuracy as high as 99.999% (Shendure and Ji 2008). Nevertheless, an individual lab may encounter substantial expense and a number of months of work to complete the sequencing of a microbial genome. These limitations have encouraged scientists to develop and utilize a variety of new sequencing technologies.

### **1.3.2 Roche 454 GS FLX Genome Sequencer**

Roche/454 commercialized Pyrosequencing in 2005 and the sequencer was the first NGS platform on the market (Margulies et al. 2005). The basis of the technology is the measurement of the release of inorganic pyrophosphate by converting it into visible light during DNA synthesis (Ronaghi et al. 1996 and 1998). The sequencing chemistry consists of a series of enzymatic reactions. First, pyrophosphate, released from the growing DNA strand, combines with adenosine-5'-phosphosulfate catalyzed by ATP sulfurylase to form ATP. Then, the ATP is used by luciferase to convert luciferin to oxyluciferin to generate light. Before the next nucleotide is added, it is necessary to remove the unused ATP and the unincorporated deoxynucleoside triphosphate; this is done by the enzyme apyrase. In addition, a thio-modified dATP,

deoxyadenosine  $\alpha$ -thiotriphosphate (dATP $\alpha$ S), is used as a substitute for natural dATP to avoid creating a false positive signal (Ronaghi et al. 1996; 1998).

The general pyrosequencing workflow includes DNA library preparation, emulsion PCR, DNA sequencing and data analysis (Margulies et al. 2005). Briefly, bacterial genomic DNA is fractionated by nebulization, in which the DNA is forced through a small hole, and DNA fragments in the range of 300- to 800-bp are selected. After DNA repair and end polishing to generate blunt ends, short 3' and 5' DNA adapters are added to each fragment. In the next step, each fragment is immobilized onto a 28  $\mu$ m bead, which has sulfurylase and luciferase attached to it. PCR amplification is performed within droplets of an oil-water emulsion. As a result, several thousand copies of the same template sequence are generated on each bead. The beads are then deposited into titanium-coated PicoTiterPlate wells. The diameter of the PicoTiterPlate wells is designed to allow for only one bead per well. During pyrosequencing, individual dNTPs are added sequentially in a predetermined order. The amount of light generated is proportional to the number of dNTPs added. The bioluminescent images are recorded by a charge-coupled device (CCD) camera. Because the linear relationship between light intensity and the number of dNTP incorporated can only hold up to 6 nucleotides, pyrosequencing has high error rates (insertions and deletions) when dealing with homopolymer repeats (Margulies et al. 2005; Metzker 2010).

The aforementioned DNA library preparation method generates sequences that can be assembled into a number of unordered and unoriented contigs. In order to close the gaps in bacterial genome sequences between those contigs, construction of a paired-end DNA library is usually recommended (Jarvie and Harkins 2008). To construct a paired-end library, bacterial genomic DNA is first sheared randomly and certain sized fragments are selected such as 3, 8 or 20 Kb. The fragments are methylated to avoid EcoRI cleavage, and hairpin adapters are ligated

onto both ends of the DNA. Subsequent exonuclease digestion removes all of the DNA fragments that are not protected by hairpins. In addition, the hairpin adapters are biotinylated and contain EcoRI recognition sites that are not methylated. Therefore, after digestion with EcoRI, the DNA can be circularized by self-ligation. The EcoRI digestion step also removes the terminal hairpin structures from the DNA. Second, the circularized DNA is fragmented by nebulization and fragments containing the added adapters are selected using streptavidin, which has very tight biotin-binding capability (Chalet and Wolf 1964). Eventually, a DNA library consisting of true paired end reads is generated, with a 44-mer adapter sequence in the middle, flanked with ~150 bp sequences on average. The two flanking 150 bp sequences are fragments of DNA that were originally located approximately 3, 8 or 20 Kb apart in the genome of interest. This library is now ready for emulsion PCR and DNA sequencing. Using paired-end reads, scaffolds can be obtained from the ordered and oriented contigs and this greatly facilitates the complete sequencing of the genome.

Pyrosequencing has been used to sequence a wide variety of genomes including the second complete human genome (Wheeler et al. 2008). Currently, the latest sequencer GS FLX+ System can produce read lengths up to 1,000 bp, which has reached the current Sanger sequencing capacities. Because the data output per run almost doubled compared to the previous versions of the sequencer, it makes pyrosequencing more attractive again. However, the error-prone raw data caused by insertions and deletions as well as low throughput compared to other NGS platforms may limit its application in certain research fields.

### **1.3.3 Illumina/Solexa Genome Analyzer**

Illumina/Solexa's Genome Analyzer is currently the most widely used DNA sequencing platform. It was the second commercialized NGS platform on market. The sequencing chemistry

is based on four-color cyclic reversible terminators that are blocked at the 3' end using modified nucleotides such as 3'-*O*-azidomethyl-dNTPs (Guo et al. 2008). In the procedure, four fluorescently labeled nucleotides are simultaneously added into the reaction mixture and DNA polymerase incorporates the nucleotide that is complementary to the template base. DNA synthesis terminates after the addition of the one nucleotide. The unincorporated nucleotides are washed away and fluorescence is recorded to determine the incorporated nucleotide. Then, the 3' blocking group is cleaved from the incorporated nucleotide to restore a 3'-OH group and the fluorophore is also removed from the base. Therefore, two chemical bonds need to be cleaved before the next cycle begins (Bentley et al. 2008). Also, since the 3' blocked terminators cannot be readily incorporated into the growing DNA strand by the native form of DNA polymerase, a modified form of DNA polymerase, created by site-directed mutagenesis, is used to carry out the reaction (Chen et al. 2010).

The sequencing workflow of the Genome Analyzer includes three steps: DNA library preparation, generation of clonal clusters, and sequencing (Bentley et al. 2008). Genomic DNA is first fragmented by nebulization or sonication. DNA end-repair is performed in order to generate blunt ended DNA. Following phosphorylation of the 5' and 3' ends, an adenosine overhang is added to each end. This facilitates the ligation between the sequencing adapters and the DNA fragments. Next, a flow cell is used to capture template molecules to generate clonal clusters, which are identical copies of each single DNA template within the diameter of one micron. The flow cell is a silica slide with eight channels and each channel can hold up to 12 samples. Different from emulsion PCR on small beads, in solid-phase amplification denatured DNA templates are covalently attached to a lawn of oligonucleotides immobilized on the flow cell surface. Templates bound to the primers are 3'-extended using a high-fidelity DNA polymerase. After denaturation, the original templates are washed off and the amplified copies are left on the

flow cell surface. Because there are adapter oligonucleotides on the free ends of the bound templates, this adapter may hybridize to adjacent lawn primer, which is immobilized on the flow cell surface, to form a bridge. DNA polymerase copies the template from the primer to form a double-stranded DNA bridge, which is subsequently denatured and two single-stranded DNAs may hybridize to adjacent lawn primers to form new bridges. This process is repeated to create millions of dense clonal clusters, each containing about 2,000 molecules. Following denaturation of the double-stranded DNA bridges, the reverse strand is removed by cleavage at the reverse strand-specific lawn primers. The 3'-OH ends are blocked to avoid nonspecific priming and sequencing primers are hybridized to the adapter attached to the unbound ends of the DNA templates. Now the flow cell, which contains clusters of ~1,000 copies of single-stranded DNA molecules, is ready for transfer to the Genome Analyzer for sequencing (Bentley et al. 2008).

Illumina/Solexa offers three strategies to prepare a DNA library, including single-read, paired-end and mate-pair. The single-read method is described in the sequencing workflow section. Mate-pair library preparation is essentially the same as paired-end library preparation in pyrosequencing, except that in mate-pair protocols the templates to be sequenced are separated by 2-5 Kb inserts instead of 3, 8 or 20 Kb. The protocol for paired-end library preparation for the Illumina Genome Analyzer is completely different from the mate-pair DNA library preparation. It generates twice the amount of sequencing data compared with single-read and it also requires twice the run time. In addition, another instrument, a Paired-End Module, needs to be attached to the Genome Analyzer. End users may choose the length of the sequencing insert ranging from 200 to 500 bp. During paired-end library sequencing, the forward strand of the DNA template is sequenced in the same way as in single-read sequencing. After denaturation, the newly synthesized partial strand is removed and the 3' ends are unblocked. The free ends can bind to lawn primers to reform bridges and double-stranded DNA clusters are regenerated. This time, the

forward strands are cleaved leaving only the newly synthesized reverse strands attached to the flow cell. Subsequent sequencing is performed on the reverse strands to produce paired-end data.

Currently, there are four sequencing instruments provided by Illumina including HiSeq 2000, HiSeq 1000, Genome Analyzer IIX, and MiSeq. The new HiSeq 2000 system is able to produce read length of up to  $2 \times 100$  bp (paired-end reads) with a maximum output of 600 Gbp per run. It also has a dual flow cell design, which enables simultaneous run applications that requires different read lengths. With Illumina technology, the most common error during sequencing is substitutions, especially after a 'G' base (Dohm et al. 2008). Amplification bias during template preparation could also cause underrepresentation of AT-rich and GC-rich regions (Dohm et al. 2008; Hillier et al. 2008; Harismendy et al. 2009; Metzker 2010).

#### **1.3.4 ABI SOLiD System**

In contrast to Roche/454 and Illumina/Solexa, which use DNA polymerase as the core biochemistry, ABI developed its NGS platform based on DNA ligase. DNA ligase can join two DNA strands that have a double-strand break. It can also link the ends on only one of the two strands, providing that the incoming single strand nucleotides are perfectly complementary to the reverse strand (Tomkinson et al. 2006).

The NGS platform commercialized by ABI is named Support Oligonucleotide Ligation Detection (SOLiD) (Valouev et al. 2008). DNA templates are prepared in a manner similar to pyrosequencing technology. The DNA is sheared by nebulization or sonication, and ligated to oligonucleotide adapters. One of the adapters is then hybridized to another adapter, called the P1 adapter, which is immobilized onto one-micron diameter paramagnetic beads. The DNA library is diluted before the hybridization between the two adapters to ensure that only one DNA

template attaches to each bead. In the next step, DNA templates are clonally amplified by emulsion PCR, followed by bead purification. In contrast to pyrosequencing where a PicoTiterPlate is used to catch the beads, in the present technique a flow cell glass slide is used. Before bead deposition, the 3' end of the DNA template on the beads is modified to allow its covalent attachment to the slide. Because there is another adapter attached to the free end of the DNA template, the modification can be made by attaching a polystyrene bead, which has complementary adapter sequences on its surface. Currently, three types of slides are available for SOLiD 4 system, which allow for the analysis of 1, 4, or 8 samples on a single slide. In order to achieve multiplexing capability, barcoding is introduced and up to 96 DNA libraries can be loaded on one region of an 8-region glass slide. The SOLiD 4 system can hold two independent flow cell slides at once. Therefore, up to 1,536 samples can be analyzed in a single run. The newer models 5500 series SOLiD sequencers use microfluidic FlowChips, which has either 6 or 12 independent lanes on the surface, allowing flexibility to run a single lane without paying for the whole run.

The 5500 series and SOLiD 4 system are able to handle three types of DNA libraries including a fragment library, a mate-pair library with insert size from 600 bp – 10 Kb, and a paired-end library. The strategies that are used to construct fragment and mate-paired libraries are similar to those are used in the Illumina/Solexa technology. SOLiD 4 system can produce up to  $2 \times 50$  bp for a mate-paired library and 50 bp for a fragment library, whereas 5500 series can generate  $2 \times 60$  bp and 75 bp, respectively. Paired-end library sequencing involves sequencing of both the forward and reverse direction of DNA templates using DNA ligase. With SOLiD 4, 50 bp of forward strand and 25 bp of reverse strand can be obtained. With the newer 5500 series sequencers, an improvement of read length up to 75 bp for the forward and 35 bp for the reverse

strand have achieved. The maximum output per run for SOLiD 4 and 5500 series systems are 80-100 GB and 300 GB, respectively.

The first step of Sequencing by Ligation (SBL) is to ligate a probe to a sequencing primer (Valouev et al. 2008). Each probe consists of eight base pairs (octamer), of which the first two at the 3' end are the ones providing the measuring information. The remaining six nucleotides are degenerate nucleotides with one of four fluorescent labels linked to the 5' end. Since dinucleotides can generate sixteen different combinations, and only four colors are used for measurement, a two-base encoding strategy is employed. For instance, blue represents the combination of AA, CC, GG or TT; green represents the combination of CA, AC, TG or GT; yellow corresponds to the combination of GA, AG, TC or CT; and red corresponds to the combination of TA, AT, GC or CG. As long as the first nucleotide is known, the second nucleotide can be determined based on the color observed. In the first sequencing step, the probes representing all 16 possible two-base combinations are added into the reaction mixture. Annealing only occurs when the probe is complementary to the sequences immediately adjacent to the sequencing primer. Then ligation is performed and the unbound probes are washed away. In the next step, unextended reactions are capped by dephosphorylation, making them unavailable to participate the future reactions. The last three bases and the fluorescent moiety from the probes are then cleaved with  $\text{AgNO}_3$ . Now the probe is reduced to 5 nucleotides with a free phosphate group. Ligation is repeated up to 15 cycles to obtain a sequence of 75 base pairs. Fifteen cycles of ligation is referred to as a "round". Primer reset is carried out where the extended sequences melt off the template and a new primer, which is one base inset closer to the bead than the starting primer, is hybridized to the adapter. Then the same set of probes is used to measure different pairs of dinucleotides. Primers are reset for five rounds in total and each new primer has a successive offset, i.e. n-1, n-2 and so on. Using this approach, each nucleotide on the

template is sequenced twice by different dye-labeled probes, thereby reducing sequencing errors dramatically. Eventually, 75 color space sequence information is collected which will be taken forward to obtain a 75 nucleotides sequence. In order to convert the color space sequence to a base pair sequence, the first nucleotide has to be known. This information can be easily obtained from the first cycle of the second round of sequencing because the first base pair of probe is the complement of the last nucleotide of the sequencing primer.

In order to obtain paired-end or mate-pair sequences, sequencing of the reverse direction of the template is performed (Valouev et al. 2008). First, 3'-hydroxylated primer is annealed to the adapter region of the templates and probes that are 5' phosphorylated are ligated to the primer. To prevent dephasing, primer that is unextended is capped by a ddNTP that is introduced by polymerase. After cleavage by  $\text{AgNO}_3$ , only 5 nucleotides from the probe will remain. The 3' phosphate is removed and the cycles are repeated until the desired read length is obtained.

Similar to Genome Analyzers of Illumina/Solexa, the most common error type created by SOLiD systems is substitution (Metzker 2010). Secondly, the beads may carry a mixture of sequences, creating false reads and low quality bases. This could be the reason of obtaining large amount of "junk" data in several studies (Harismendy et al. 2009; Suzuki et al. 2011). In addition, SOLiD data also reveals an underrepresentation of AT-rich and GC-rich regions (Harismendy et al. 2009).

### **1.3.5 Helicos HeliScope Sequencer**

HeliScope developed by Helicos BioSciences was the first commercialized single-molecule sequencer (Braslavsky et al. 2003). It has the advantage of not requiring amplification of the templates by PCR before sequencing, since clonal amplification of templates may

introduce errors. This technology significantly increased the speed of DNA sequencing, while decreasing the cost.

The HeliScope uses Virtual Terminators, which are 3'-unblocked cyclic reversible terminators (Bowers et al. 2009). The inhibiting group is a nucleoside analogue that is directly attached to the fluorophore. Using a 3'-unblocked terminator is highly efficient because removal of the fluorophore and terminating group is combined into one step. Furthermore, it is no longer necessary to screen large libraries of mutant DNA polymerase since 3'-unblocked terminators can be incorporated into the growing strand DNA effectively by wild-type DNA polymerase.

The workflow of single-molecule sequencing may be summarized as following (Harris et al. 2008). A DNA sample is sheared into short strands of about 100-200 nucleotides in length before a poly-A universal priming sequence is added to the 3' end of each DNA strand, which is then labeled with a fluorescent adenosine nucleotide. The labeled strands serve as templates for the single molecule sequencing chemistry. The DNA strands are hybridized to the Helico's flow cell that contains billions of oligo-T universal capture sites that are immobilized on the flow cell surface. Because the HeliScope sequencer detects single molecules, the templates can be packed at very high density, i.e. billions of templates per run. After the DNA sequences have been hybridized to the flow cell surface, they are loaded into the Heliscope instrument. A laser illuminates the surface of the flow cell, highlighting the location of each fluorescently labeled template. A CCD camera then produces a map of the template on the flow cell surface. After the template has been imaged, the template label is cleaved and washed away. The sequencing reaction begins by introducing a DNA polymerase and a fluorescently (Cy5) labeled nucleotide with the oligo-T capture sites serving as sequencing primers. DNA polymerase catalyzes the addition of Cy5-labeled nucleotides to the primers in a template directed manner. A washing step then removes the polymerase and any unincorporated nucleotides. The billions of single molecule

templates that have incorporated a particular nucleotide are then visualized by illuminating and imaging the entire flow cell surface. After imaging, the fluorescent labels are cleaved and removed. The process continues with each of the remaining bases and repeats until the desired read length has been achieved. Unlike amplification-based sequencing technologies, the single-molecule sequencing process is asynchronous. Every strand is unique and is sequenced independently.

Paired-end reads can be obtained from individual single molecules as well. Unlike the traditional paired-end library preparation, the procedure does not involve cloning, circularization and digestion of the sheared genomic DNA sample. Briefly, after fragmentation of genomic DNA, an adapter sequence is ligated to the 5' end of the fragments. Then poly-A tails that will hybridize to the poly-T immobilized on the flow cell surface are created on the 3' end of the fragments. Following the completion of sequencing the 3' end of the template, the template is copied to the end by DNA polymerase and all four natural nucleotides. The template DNA is removed by denaturation, leaving only the reverse strand bound to the flow cell surface. A universal primer then hybridize to the adapter sequence, which is at the free end of the DNA fragment, and sequencing can be performed from the free end of DNA template, i.e. the 3' end of the reverse strand.

Currently, HeliScope can generate an average single read length of 35 bp with 21 to 35 Gbp output per run. According to the company's report, the raw error rate for substitution, insertion and deletion are 0.2%, 1.5%, and 3%, respectively. Since the first HeliScope sequencer was sold in March 2008, only a few machines have been installed globally. This is probably due to the high cost of the sequencer (~ one million dollars) and short reads compared to other NGS platforms (Kircher and Kelso, 2010; Glenn 2011).

### **1.3.6 Pacific BioSciences PACBIO RS**

Single molecule real time (SMRT) sequencing is a Sequencing by Synthesis (SBS) technology that involves monitoring the incorporation of fluorescent dye-labeled nucleotides continuously during DNA synthesis (Eid et al. 2009). The SMRT DNA sequencing system was commercialized by Pacific Biosciences in May 2011 and is named PacBio RS. The most attractive feature of this sequencing system is that it can generate more than a thousand base pairs of sequence information in fast cycle times, since DNA polymerase synthesized DNA continuously without termination (Metzker 2009 and 2010; Eid et al. 2009).

During the template preparation, the DNA sample is first sheared to varied lengths (250 bp – 6 Kb); the ends of the fragments are then repaired so that a hairpin structure can be ligated to each end. Following purification, the fragments with hairpin adapters are selected and ready for sequencing. The company provides Template Preparation Kit to convert the DNA sample into SMRTbell library format, which will eventually generate structurally linear and topologically circular DNA morphology. This way, both the sense and antisense DNA strands can be sequenced. In contrast to other SBS approaches in which templates are attached to a solid surface, in SMRT sequencing, single DNA polymerase molecule is anchored to the bottom of a nanoscale well called zero-mode waveguide (ZMW) (Levene et al. 2003; Foquet et al. 2008). Parallel sequencing is achieved using a chip containing thousands of ZMWs to capture individual DNA polymerase molecules (Eid et al. 2009). Because the DNA polymerase and DNA template are diffused into the ZMWs, not all the wells are active for sequencing. It has been estimated that about one third of ZMWs of a chip are active for a given run (Schadt et al. 2010). A modified  $\phi$ 29 DNA polymerase is chosen for this sequencing platform because it can incorporate phospholinked dNTPs efficiently into the growing strand. In addition,  $\phi$ 29 polymerase is able to synthesize DNA in a strand-displacement manner so that the template can be sequenced multiple

times to ensure accuracy. With this approach, read accuracy was improved from <80% for a 499 bp template to >99% by circular consensus sequencing for 15 times or more (Eid et al. 2009). The nucleotides used in SMRT are phospholinked hexaphosphate nucleotides and have a distinct character in that a fluorophore is linked to the terminal phosphate rather than to the base. The formation of phosphodiester bond leads to the release of the dye-labeled pentaphosphate and the signal is recorded immediately before it diffuses away (Eid et al. 2009).

SMRT sequencing can be used to identify DNA modification because the polymerase kinetics changes at and around the modified base position in the DNA template, which is typically slower compared to unmodified DNA. For instance, DNA methylation such as N<sup>6</sup>-methyladenine, 5-methylcytosine, 5-hydroxymethylcytosine, and N<sup>4</sup>-methylcytosine have been detected by SMRT sequencing (Flusberg et al. 2010; Song et al. 2011; Clark et al. 2011). In addition, various forms of DNA damage can be revealed directly by SMRT sequencing as well. It has been demonstrated that DNA templates containing 8-oxoguanine, 8-oxoadenine, O<sup>6</sup>-methylguanine, 1-methyladenine, O<sup>4</sup>-methylthymine, 5-hydroxycytosine, 5-hydroxyuracil, 5-hydroxymethyluracil, or thymine dimers can be identified readily with single-modification resolution and DNA strand specificity (Clark et al. 2011). These applications will advance our understanding DNA-lesion-related diseases, aging, and DNA polymerase enzymology (Clark et al. 2011).

It has been reported that using the SMRT system to sequence an *E. coli* genome, it is possible to achieve 99.3% genome coverage with average read lengths of 964 bp and at high accuracy, i.e. >99.999%. However, it was necessary to have a 38-fold base coverage to obtain this high quality data (Metzker 2010). The error rates of the raw reads are still pretty high (~5%) for this platform, and the dominant forms of errors are insertions and deletions. Furthermore, the throughput of SMRT sequencing depends on the number of ZMWs that are active for a given

run. At present, one SMRT cell contains up to  $2 \times 75,000$  ZMWs. Even if all the wells are active, the maximum throughput may not be comparable with other NGS platforms (Schadt et al. 2010).

### **1.3.7 Polonator G.007**

The Polonator is a platform that may serve as an alternative to ABI SOLiD systems and Illumina Genome Analyzers. The developers of the platform aim to lower the cost of sequencing instruments and reagents dramatically while maintain high throughput, accuracy, and reliability of the data.

The Polonator was initially developed by George Church's lab at Harvard Medical School in collaboration with Dover, and has been sold by Azco Biotech since 2009. The system is an open platform with free operating software and protocols available for public download. Users may employ the standard settings or they can innovate freely.

Two protocols are available to construct DNA libraries, which will generate either a Polony or Rolony library. To construct a Polony library, the genomic DNA is sheared into 1 Kb fragments by a Hydroshear and purified. The DNA segments are then circularized using a T30 Universal Linker followed by a restriction digest to create two mate-paired genomic DNA tags. Next, forward and reverse sequencing primers are ligated to the tags. Then the DNA library is used to attach to magnetic beads, where emulsion PCR occurs. After separating the amplified beads from the unamplified beads, the tags are ready for sequencing (Shendure et al. 2005; Azco Biotech). To construct a Rolony library, the genomic DNA is first sheared into 500 bp – 1 Kb segments and purified. Secondly,  $\frac{1}{2}$  of the Universal Adapter is added to each end of the DNA fragment. This fragment is circularized followed by cleavage with restriction endonuclease to generate the tag. Then a primer is ligated to the end of the tag. Next, another tag is added by

repeating the procedure from the second step. After adding the second tag, the DNA is ligated on a T30 linker to become a circle. Then, rolling circle amplification is performed to create DNA balls, which will subsequently attach to a slide for sequencing. With this protocol, expensive magnetic beads are no longer used (Azco Biotech; Drmanac et al. 2009).

Currently, two sequencing techniques, SBL using DNA ligase and SBS using DNA polymerase, can be performed on Polonator system, making it more flexible. At present, the read length of Polonator is 18 – 22 bp, and about 16 Gbp can be generated per run. The mean accuracy for mappable reads is greater than 99.7% and the current run time for SBL is 2.5 days for a mate-paired run (Azco Biotech). The major drawback of Polonator is that it generates the shortest read lengths compared with other NGS platforms, making genome assembly a big challenge. Before Polonator can push the read length to a competitive position, certain application such as targeted sequencing that does not require long reads will probably be more suitable for this system.

### **1.3.8 Ion Torrent Personal Genome Machine**

Personal Genome Machine (PGM) developed by Ion Torrent is based on an entirely new method known as semiconductor sequencing. The machine was launched at the end of 2010 and it measures the hydrogen ion released when a base is incorporated into a template DNA by polymerase.

To perform this biochemical process, Ion Torrent designed a high-density ion chip with millions of wells, which allow parallel detection of individual sequencing reactions (Rothberg et al. 2011). Beneath each well is an ion sensitive layer and below that is an ion sensor. The sequencer sequentially floods the chip with one natural nucleotide after another. If the nucleotide is complementary to the template base, the pH in the solution will change due to the release of hydrogen ion, which will be converted to voltage and recorded by the semiconductor sensor. In

the case of two identical bases on the DNA strand, the voltage change will double and the signal will be recorded as two identical bases call. This direct detection requires no scanning, no cameras, and no light, which saves a lot of time (Rothberg et al. 2011).

To prepare a DNA library for semiconductor sequencing, genomic DNA is sheared randomly into fragments by mechanical or enzymatic approaches. Then the DNA fragments are ligated to Ion-specific forward and reverse adapters on each end. The DNA with the adapters will be subsequently amplified on beads. Next, the beads containing DNA template are selected and DNA polymerase and sequencing primers are added, and individual beads are loaded onto an Ion chip followed by sequencing (Rothberg et al. 2011). Recently, a protocol for paired-end sequencing became available. During the library preparation, a modified adapter that contains site-specific nicking site is used, which will facilitate the reverse sequencing. After the forward sequencing run, the ion chip is removed from PGM and a series of enzymatic steps are performed. The forward primer is extended fully to the end of the template. Then, the original template is nicked and degraded by an enzyme to produce the primer for the reverse sequencing. Next, the chip is loaded back to PGM, and the sequencing is performed in the reverse direction (Ion Torrent, 2011b).

Ion Torrent offers three kinds of chips, 314, 316, and 318. The number of wells on each type of chip are 1.2, 6.2, and 11.1 million. For single-end sequencing, the three chips can generate >10 Mbp, >100 Mbp, and >1 Gbp output, respectively. Currently, the read length of single-end is around 200 bp, but it will reach 400 bp in 2012 (Ion Torrent, 2011a). With paired-end sequencing on an *E. coli* strain,  $2 \times 100$  bp can be produced using a 314 chip, yielding 69.7 Mbp of data with a quality score greater than 20. Compared with a 314 chip, a  $2 \times 100$  bp paired-end sequencing on a 316 chip can produce a total of 826 Mbp of data, increasing the output

significantly. At present, no paired-end sequencing datasets have been released for the 318 chips (Heger, 2011).

The semiconductor sequencing technology provides even coverage of an entire genome regardless of the changes in G+C content, demonstrating minimum sequence bias in the reads (Rothberg et al. 2011). In terms of data accuracy, semiconductor sequencing is able to achieve >99.5% raw accuracy and >99.99% consensus accuracy according to the company's reports (Ion Torrent, 2011a). However, similar to pyrosequencing, the most common error types are insertions and deletions caused by homopolymers in the DNA template. A 6-mer or longer stretches makes statistical analysis problematic and difficult (Ion Torrent, 2011a).

### **1.3.9 Summary**

Due to the intrinsic limitations and biases of each of the currently available sequencing technologies, it has been suggested that using a combined sequencing strategies is more practical, considering both the quality of sequencing results and cost (Goldberg et al. 2006; McCutcheon and Moran 2007; Aury et al. 2008; Lam et al. 2011). Also, some platform may have a better performance on certain applications compared with other techniques. In addition, for a genome having a large number of repetitive regions, the use of paired-ends or mate-pairs DNA library is necessary since the addition of a large amount of relatively short reads won't help much to reduce the gaps between sequencing regions.

At present, cost and data quality are still the main concerns of bacterial genome sequencing. However, with constantly developing sequencing technology as well as bioinformatics analysis tools, it is possible that in the near future, bacterial genome sequencing and analysis will become a routine procedure in every microbial laboratory.

## 2 OBJECTIVES

Since *P. putida* UW4 is a well-studied PGPB that has great potential to promote plant growth in the field, knowing its genome sequence will be beneficial in understanding the mechanisms of plant-microbe interactions. Moreover, bacterial genome sequencing is now affordable owing to the rapid development of NGS technologies, making the genome sequencing of UW4 feasible. Therefore the objectives of my research project were:

- I. Sequence *P. putida* UW4 genome using one of the NGS platforms namely pyrosequencing, with two DNA libraries, single-read and paired-end read
- II. Fill the gaps between contigs by directed-PCR and primer-walking
- III. Annotate and analyze *P. putida* UW4 genome sequence:
  - a. General features of the genome
  - b. Genes involved in plant growth promotion and UW4 lifestyle
  - c. Metabolic pathways and protein secretion systems
  - d. Whole genome phylogeny and comparisons
  - e. Genome comparison among complete sequenced *Pseudomonas* genomes
  - f. 16S rRNA genes phylogenetic analysis of *Pseudomonas* genomes
  - g. Heat shock protein genes phylogeny and  $\sigma^{32}$  promoter analyses

### 3 MATERIALS AND METHODS

#### 3.1 Bacterial Growth and DNA Extraction

A single colony of *P. putida* UW4 grown on Tryptic Soy agar (Difco Laboratories, Detroit, MI) was inoculated into 5 mL of TSB (Difco Laboratories, Detroit, MI) and grown overnight with shaking at 30°C. Bacterial cells were collected by centrifugation and the genomic DNA was extracted with a Wizard<sup>®</sup> Genomic DNA purification kit (Promega, Madison, WI, USA) according to the manufacturer's instructions. *E. coli* DH5 $\alpha$  (Hanahan, 1983) was used as a recipient for recombinant plasmids. This strain and its transformants with different plasmids were grown at 37°C in Luria-Bertani (LB) broth medium (Difco Laboratories, Detroit, MI), with appropriate antibiotics. Antibiotics were added at the following concentrations for *E. coli*: ampicillin, 100  $\mu$ g/mL; tetracycline, 15  $\mu$ g/mL.

#### 3.2 Whole Genome Pyrosequencing and Sequence Assembly

The complete genomic sequencing was carried out at The McGill University and Genome Quebec Innovation Center where they used the current Roche GS-FLX Titanium chemistry protocols in place to sequence the genomic DNA. First, a shotgun library was prepared from 5  $\mu$ g of DNA and subsequently sequenced, generating 203,178 reads in 74,063,913 bp of sequencing data. The average final read length for the run was 365 bp. For the assembly, 93% of the reads were fully assembled into 312 large contigs, ranging from 518 – 197,691 bp. The sum of the large contigs' size is 6,049,654 bp and about 12 $\times$  of the sequencing coverage was achieved. In order to facilitate gap closure in the genome sequence, an 8 Kb paired-end library was then constructed using 15  $\mu$ g of DNA to re-sequence the entire genome. After sequencing, 186,877 reads were generated in 73,775,344 bp of sequencing data, with an average final read length of

395 bp. Combining the results of shotgun and paired-end sequencing, 96% of the reads were fully assembled into 122 large contigs ranging from 500 – 356,439 bp. Ten ordered and oriented scaffolds with a genome size of 6.22 Mb were obtained. Using paired-end sequencing, another 12× genome coverage was achieved. *De novo* sequence assembly was completed using Roche's Newbler assembler v.2.0.01.14 at The McGill University and Genome Quebec Innovation Center. Gaps between the contigs were filled in by sequencing the PCR products using Applied Biosystems 3730xl DNA Analyzers at The McGill University and Genome Quebec Innovation Center, University of Guelph's Advanced Analysis Centre, and York University Core Molecular Biology and DNA Sequencing Facility. Initially, 100 pairs of primers were designed to fill in the 100 gaps. Then, primer walking was used to close the gaps that were greater than 1.5 Kb. KOD Hot Start DNA Polymerase (EMD Millipore, MA, United States) and GoTaq® Hot Start Polymerase (Promega, WI, United States) were used for PCR amplification.

### **3.3 Genome Annotation and Analysis**

The *P. putida* UW4 genome sequence was first annotated using web-based automated pipelines including Bacterial Annotation System (BASys) v1.0 (Van Domselaar et al. 2005) and Integrative Services for Genomic Analysis (ISGA) v1.2 (Hemmerich et al. 2010). Putative coding sequences (CDS) were identified by Glimmer v3.02 (Delcher et al. 2007) and Prokaryotic Dynamic Programming Genefinding Algorithm (Prodigal) v2.5 (Hyatt et al. 2010). The results from the two programs were combined and manually reviewed. Ribosomal RNA and transfer RNA genes were predicted by RNAmmer v1.2 (Lagesen et al. 2007) and tRNAScan-SE (Lowe and Eddy 1997), which are embedded in the ISGA annotation pipeline. Next, functional annotation of the identified genes was conducted by a sequence similarity search against non-

redundant (NR) protein database at the GenBank by BLAST, and putative function was assigned to each gene with a cutoff E-value of  $\leq 1 \text{ E}^{-05}$ . Cluster of Orthologous Group (COG) and enzyme-coding genes were predicted by COG Finder 1.0 and ECNumber Finder in BASys. With the ISGA pipeline, COG was searched against the NCBI COG database (Tatusov et al. 1997 and 2003) and an E.C. number was assigned by PRIAM (Claudel-Renard et al. 2003). The results from both pipelines were compared and manually corrected based on the current COG database and Enzyme nomenclature database (Bairoch 2000). Protein localization was predicted by PSORTb v3.0.2 (Yu et al. 2010) and genomic islands (GIs) were detected using IslandViewer (Langille and Brinkman 2009). Repeat sequences were examined by Tandem Repeats Finder v4.04 (Benson 1999). The metabolic pathways were constructed using Pathway Tools v15.5 (Karp et al. 2009) and the KEGG database (Kanehisa et al. 2004). Genome comparisons among 10 completely sequenced *P. putida* and *P. fluorescens* genomes were carried out using TBLASTX (Camacho et al. 2009) and displayed by the Artemis Comparison Tool (ACT) (Carver et al. 2005). Orthologs in the 21 *Pseudomonas* genomes were identified using Roundup (DeLuca et al. 2006) with the most stringent blast E-value ( $1 \text{ E}^{-20}$ ) and divergence thresholds (0.2). Then the amino acid sequences of the core genes were aligned using the MUSCLE program in SeaView v4.3.2 (Edgar 2004; Gouy et al. 2010), and poorly aligned regions were removed manually using Geneious Pro 5.4.6 (Drummond et al. 2011). Before constructing a maximum likelihood (ML) tree for each alignment, the model of protein evolution was selected using PROTTEST v2.4 (Abascal et al. 2005). Next, a ML tree was built using PHYML v3.0 (Guindon and Gascuel 2003) embedded in SeaView v4.3.2 with the appropriate model for each alignment. Nodal support was evaluated by the approximate likelihood ratio test (aLRT) (Anisimova and Gascuel 2006). Based on all the orthologs that were identified, a phylogenetic tree of 21 different *Pseudomonas* species was constructed using the consensus tree program of Geneious Pro 5.4.6

(Drummond et al. 2011). DNAPlotter (Carver et al. 2009) was used to draw a *P. putida* UW4 genome atlas. The analysis of DnaA boxes consensus sequences of UW4 was performed using MEME (Multiple Em for Motif Elicitation) v4.8.1 (Bailey and Elkan, 1994).

### **3.4 Phylogeny of 16S rRNA genes of *Pseudomonas* genomes**

The 16S rRNA gene sequences of *P. putida* UW4 were aligned with those of the publicly available *Pseudomonas* genome sequences or the type strain of *P. fluorescens* and *P. putida* using the MUSCLE program in SeaView v4.3.2 (Edgar 2004; Gouy et al. 2010) and refined manually using Geneious Pro v5.4.6 (Drummond et al. 2011). All the 21 *Pseudomonas* genomes have multiple copies of 16S rRNA genes, and only unique sequences were included in this analysis. The substitution model was selected using jModeltest v0.1.1 (Posada 2008) and a ML tree was built by PHYML v3.0 (Guindon and Gascuel 2003) in SeaView v4.3.2 with a general time-reversible model (GTR), with the nodal support assessed by aLRT. Analysis of 16S rRNA sequence identities was performed by BioEdit v7.1.3 (Hall 1999).

### **3.5 Analysis of Heat Shock Protein Genes and $\sigma^{32}$ promoters**

Nucleotide sequences of heat shock protein genes were aligned using Seaview v4.3.2 and poorly aligned regions were eliminated using Geneious Pro v5.4.6. jModelTest v0.1.1 was used to determine the best substitution model for nucleotide sequence alignments and GTR model was chosen for constructing the ML phylogenetic trees. aLRT was used to evaluate the reliability of the nodal support. The promoter sequences of the heat shock sigma factor,  $\sigma^{32}$ , were determined visually through sequence similarity with the consensus sequence.

## 4 Results

### 4.1 General Genome Features

The genome of *P. putida* UW4 has a single circular chromosome of 6,183,388 bp (Fig. 4-1) and an average G+C content of 60.05% (Table 4-1). The genome contains 5,431 predicted CDSs with an average length of 995 bp. Among these CDSs, 4379 (80.7%) genes could be classified into COG families composed of 22 categories (Table 4-2). Seventeen CDSs were assigned pseudogenes due to missing either an N- and/or C-terminus (Table 4-3). Coding regions cover 87.4% of the whole genome. Biological roles were assigned to 4,154 (76.4%) genes of the predicted coding sequences based on similarity searches and experimental evidence. The remaining coding sequences were classified as proteins with unknown function. Among the 1277 (23.5%) CDSs with unknown function, 132 hypothetical proteins have no hit when searched against protein databases using a cutoff *E* value of  $10^{-5}$ , indicating putative unique genes present only in UW4. A total of seven rRNA operons including eight 5S rRNAs, seven 16S rRNAs, and seven 23S rRNAs are present on the chromosome. In addition, 72 tRNA genes that represent all 20 amino acids, and a tRNA for selenocysteine, were identified (Table 4-4).

For circular bacterial chromosomes, replication starts at the *oriC* (replication origin) region and proceeds bidirectionally to the terminus. It has been observed that most circular bacterial genomes display asymmetry in nucleotide composition in the leading strand and lagging strand, due to the strand-biased spontaneous mutation from C to T during replication (Marín and Xia 2008). Therefore GC skew ( $G-C/G+C$ ) has been widely used to identify the origin and terminus of replication in many bacteria. The chromosome of *P. putida* UW4 displays two clear GC skew transitions, which corresponds with its *oriC* and terminus (Fig. 4-1). The *oriC* site contains nine conserved DnaA-binding boxes (TTATCCACA and closely related sequences)

Fig. 4-1. Circular genome map of *P. putida* UW4. From the outside in, the outer black circle shows the scale line in Mbps; circles 2 and 3 represent the coding region with the colors of the COG categories; circle 4 and 5 show tRNA (green) and rRNA (red), respectively; circle 6 displays the IS elements (blue); circle 7 shows the genomic islands (orange); circle 8 represents mean centered G+C content (bars facing outside-above mean, bars facing inside-below mean); circle 9 shows GC skew  $(G-C)/(G+C)$ . GC content and GC skew were calculated using a 10-kb window in steps of 200 bp.



Table 4-1. General Features of *P. putida* UW4 Genome

Features	Chromosome
Size (bp)	6,183,388
G+C content (%)	60.05
Number of CDSs	5431
Total CDSs size (bp)	5,406,063
Coding %	87.4
Average CDS length (nt)	995
Pseudogenes	17
tRNAs	72
rRNA genes (clusters)	22 (7)
Number of genes with assigned function	4154 (76%)
Number of genes without assigned function	1277 (24%)
Number of predicted enzymes	1551 (29%)

Table 4-2. COG Functional Categories of *P. putida* UW4

	Functional Category	UW4
A	RNA processing and modification	1
B	Chromatin structure and dynamics	3
C	Energy production and conversion	288
D	Cell cycle control, cell division, chromosome partitioning	37
E	Amino acid transport and metabolism	500
F	Nucleotide transport and metabolism	90
G	Carbohydrate transport and metabolism	232
H	Coenzyme transport and metabolism	156
I	Lipid transport and metabolism	255
J	Translation, ribosomal structure and biogenesis	174
K	Transcription	431
L	Replication, recombination and repair	177
M	Cell wall/membrane/envelope biogenesis	262
N	Cell motility	114
O	Posttranslational modification, protein turnover, chaperones	179
P	Inorganic ion transport and metabolism	229
Q	Secondary metabolites biosynthesis, transport and catabolism	102
R	General function prediction only	440
S	Function unknown	373
T	Signal transduction mechanisms	231
U	Intracellular trafficking, secretion, and vesicular transport	43
V	Defense mechanisms	62
-	Not in COGs	1052
Total		5431

Table 4-3. *P. putida* UW4 Pseudogenes

locus_tag	product	pseudo
PputUW4_2113	long-chain fatty acid transporter	C-terminus missing fragment
PputUW4_2521	AraC family transcriptional regulator	C-terminus missing fragment
PputUW4_3551	IS630 family transposase, truncated	C-terminus missing fragment
PputUW4_5207	glutamine amidotransferase	C-terminus missing fragment
PputUW4_0114	von Willebrand factor, type A	N-terminus missing fragment
PputUW4_1252	methyltransferase	N-terminus missing fragment
PputUW4_2026	ATP-dependent DNA ligase LigD	N-terminus missing fragment
PputUW4_2127	IS30 family transposase	N-terminus missing fragment
PputUW4_2373	IclR family transcriptional regulator	N-terminus missing fragment
PputUW4_2519	xylulose kinase	N-terminus missing fragment
PputUW4_3361	phage integrase	N-terminus missing fragment
PputUW4_3383	LuxR family transcriptional regulator	N-terminus missing fragment
PputUW4_4664	type IV pilus-associated protein	N-terminus missing fragment
PputUW4_1487	transcriptional regulator	N- and C-terminus missing fragment
PputUW4_2143	hypothetical protein	N- and C-terminus missing fragment
PputUW4_2427	LysR family transcriptional regulator	N- and C-terminus missing fragment
PputUW4_4694	glutamate dehydrogenase	N- and C-terminus missing fragment

Table 4-4. *P. putida* UW4 tRNAs

tRNA	Number
Ala	9
Arg	5
Asn	2
Asp	4
Cys	1
Gln	2
Glu	4
Gly	5
His	2
Ile	7
Leu	6
Lys	2
Met	5
Phe	1
Pro	3
Sec	1
Ser	4
Thr	3
Trp	1
Tyr	1
Val	4

(Fujita et al. 1989; Yee and Smith, 1990) and is located between the *rpmH* and the *dnaA* genes (Fig. 4-2).

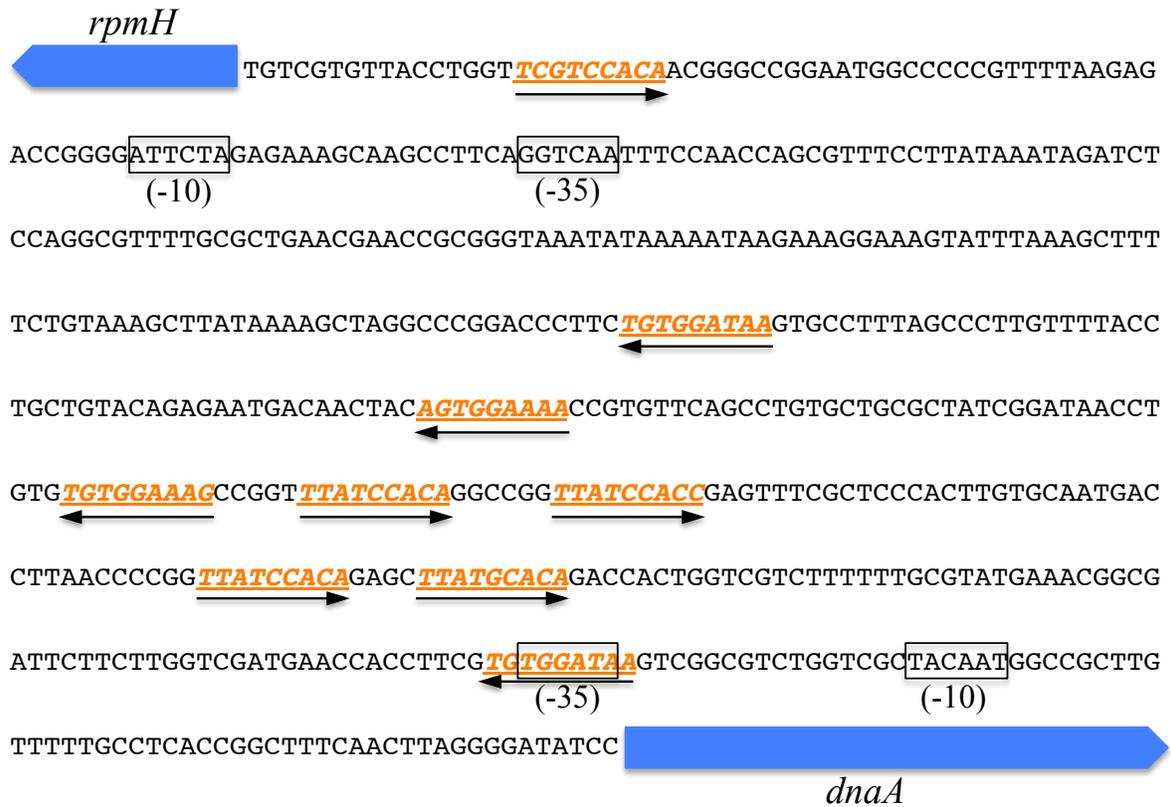
Nineteen putative GIs were identified by IslandViewer, which integrates two prediction methods IslandPath (DNA composition comparison) (Hsiao et al. 2003) and SIGI-HMM (codon usage) (Waack et al. 2006) (Fig. 4-3 and Table 4-5). The size of the 19 islands ranged from 4,143 bp (GI 15) to 25,664 bp (GI 7). The largest GI 7 contains 24 genes, whereas the smallest GI 15 has 6 genes (Table 4-5). Eighteen GIs have a lower GC content ranging from 40.33% to 58.82% compared with the average GC content of the UW4 genome. GI 11 has a GC content of 63.92%, which is higher than the average GC content of the UW4 genome. It contains 5 genes and two of them (PputUW4\_2603 and PputUW4\_2604) showed high similarities (88% and 84% at an amino acid level) with those in the predicted GIs of *P. fluorescens* Pf0-1. Among the 19 GIs, six contain mobile genetic elements, such as integrase and transposase genes, suggesting that these GIs can self-mobilize (Langille et al. 2010). The 3' ends of tRNAs have been suggested to be hot spots for foreign DNA integration (Hacker et al. 1997). In UW4, GI 4 and GI 15 are inserted adjacent to the 3' ends of tRNA-Leu and tRNA-Val, respectively, which support the identification of these two GIs.

The genome of *P. putida* UW4 has 31 complete putative Insertion Sequence (IS) elements and 5 truncated remnants of IS elements (Table 4-6). Among the complete IS elements, sixteen belong to the IS110 family, seven from the IS1182 family and eight from the IS3 family. No intact prophages were observed in the genome of UW4, nevertheless, UW4 carries 19 phage related genes (Table 4-7).

One hundred and eighty two tandem repeats were identified in the *P. putida* UW4 genome (Table 4-8). Among the 182 repeats, 122 were found in the coding region, which may cause changes in protein sequence during replication. Sixty repeats were observed in the non-

Fig. 4-2. (A) Nucleotide sequence of the *dnaA* upstream region of *P. putida* UW4. Putative promoters of *rpmH* and *dnaA* are boxed. DnaA boxes are underlined in orange and the arrows indicate the orientation. (B) Frequencies of nucleotides at each position of the nine UW4 DnaA boxes.

(A)



(B)

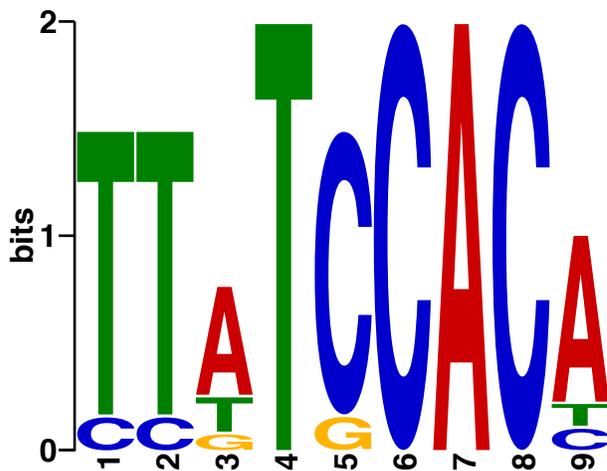


Fig. 4-3. Genomic islands of *P. putida* UW4 predicted by IslandViewer. The outer black circle shows the scale line in Mbps. Predicted genomic islands are colored based on the following methods: SIGI-HMM, orange; IslandPath-DIMOB, blue; Integrated detection, red. Black plot represents the GC content (%).

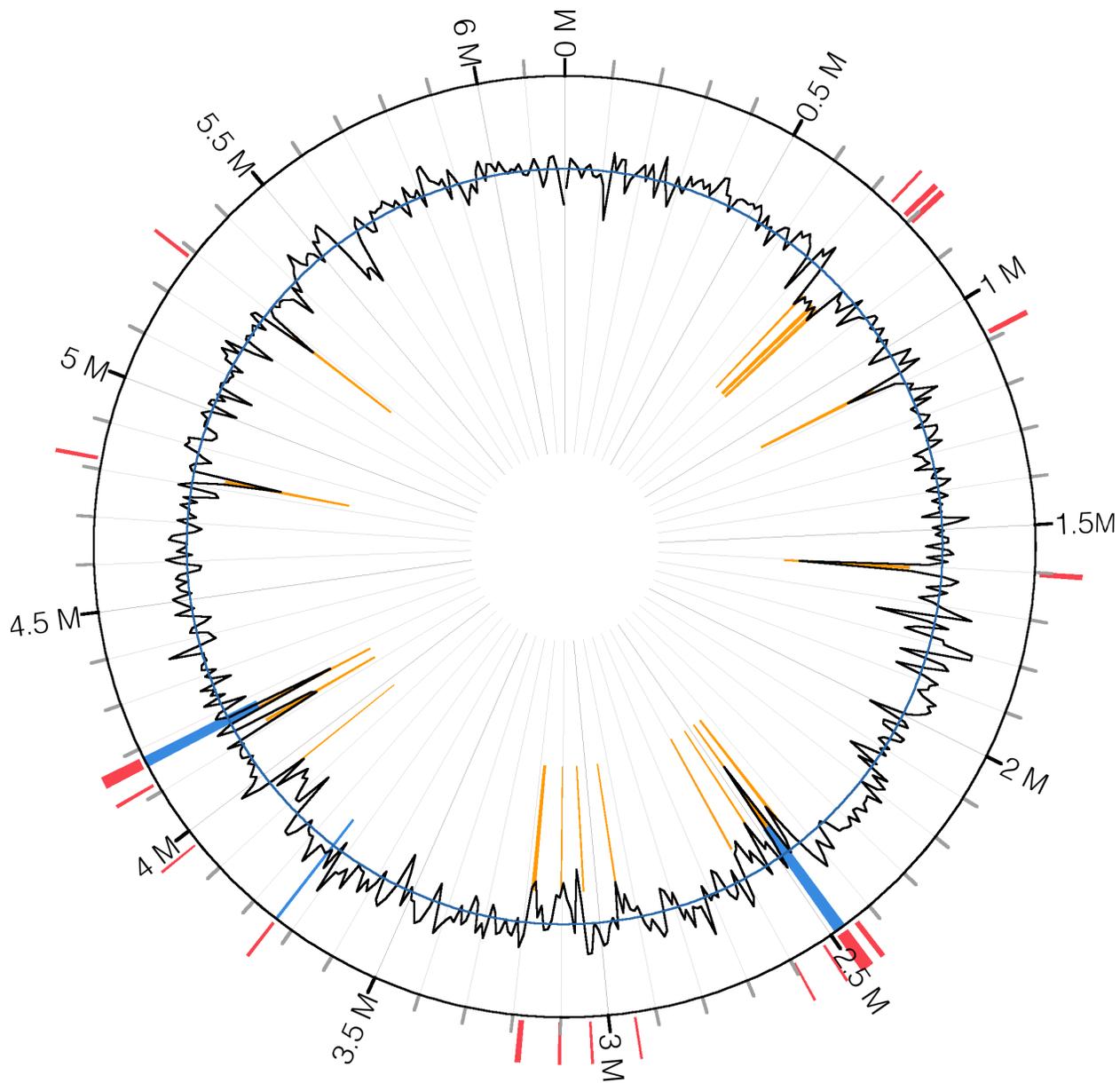


Table 4-5. Genomic Islands of *P. putida* UW4 Predicted by IslandViewer

Genomic Islands			Genes		GC content	
Start	End	Size (bp)	Locus ID	Product		
1	743,662	749,401	5,739	PputUW4_0641	ATPase	53.94
				PputUW4_0642	McrBC 5-methylcytosine restriction system component-like protein	
				PputUW4_0643	hypothetical protein	
				PputUW4_0644	hypothetical protein	
				PputUW4_0645	hypothetical protein	
				PputUW4_0646	hypothetical protein	
				PputUW4_0647	hypothetical protein	
2	780,455	791,108	10,653	PputUW4_0648	hypothetical protein	55.06
				PputUW4_0666	hypothetical protein	
				PputUW4_0667	putative helicase	
				PputUW4_0668	DNA helicase	
3	797,778	809,172	11,394	PputUW4_0669	hypothetical protein	54.74
				PputUW4_0670	phage integrase	
				PputUW4_0673	chemotaxis protein MotB-related protein	
				PputUW4_0674	hypothetical protein	
				PputUW4_0675	SNF2 family DNA/RNA helicase	
4	1,078,659	1,088,425	9,766	PputUW4_0676	5-methylcytosine-specific restriction protein C	52.94
				PputUW4_0677	hypothetical protein	
				PputUW4_0678	Phage integrase	
5	1,598,659	1,609,449	10,790	PputUW4_0917	NAD-dependent formate dehydrogenase gamma subunit	44.61
				PputUW4_0918	type III restriction protein res subunit	
				PputUW4_0919	hypothetical protein	
5	1,598,659	1,609,449	10,790	PputUW4_1368	putative lipoprotein	44.61
				PputUW4_1369	Integration host factor subunit beta	
				PputUW4_1370	chain length determinant family protein	
				PputUW4_1371	acetyltransferase	
				PputUW4_1372	lipopolysaccharide biosynthesis protein RffA	

				PputUW4_1373	hypothetical protein	
				PputUW4_1374	hypothetical protein	
				PputUW4_1375	hypothetical protein	
				PputUW4_1376	hypothetical protein	
				PputUW4_1377	glycosyl transferases group 1 family protein	
				PputUW4_1378	dTDP-4-dehydrorhamnose reductase	
				PputUW4_2095	hypothetical protein	
				PputUW4_2096	hypothetical protein	
				PputUW4_2097	hypothetical protein	
				PputUW4_2098	hypothetical protein	
				PputUW4_2099	diguanylate cyclase	
6	2,436,768	2,446,210	9,442	PputUW4_2100	hypothetical protein	54.74
				PputUW4_2101	acetyltransferase	
				PputUW4_2102	MarR family transcriptional regulator	
				PputUW4_2103	AraC family transcriptional regulator	
				PputUW4_2118	quinone oxidoreductase	
				PputUW4_2119	methionine sulfoxide reductase A	
				PputUW4_2120	RND family efflux transporter, MFP subunit	
				PputUW4_2121	ABC transporter family protein	
				PputUW4_2122	peptide ABC transporter permease	
				PputUW4_2123	peptide ABC transporter permease	
7	2,464,903	2,490,567	25,664	PputUW4_2124	IS1182 family transposase	55.11
				PputUW4_2125	ThiF family protein	
				PputUW4_2126	hypothetical protein	
				PputUW4_2128	hypothetical protein	
				PputUW4_2129	hypothetical protein	
				PputUW4_2130	hypothetical protein	
				PputUW4_2131	hypothetical protein	
				PputUW4_2132	IS1182 family transposase, truncated	
				PputUW4_2133	IS1182 family transposase, truncated	
				PputUW4_2134	hypothetical protein	

				PputUW4_2135	hypothetical protein	
				PputUW4_2136	IS1182 family transposase	
				PputUW4_2137	polysaccharide deacetylase family protein	
				PputUW4_2138	GCN5-related N-acetyltransferase	
				PputUW4_2139	hypothetical protein	
				PputUW4_2140	HxlR family transcriptional regulator	
				PputUW4_2141	nitrilase	
				PputUW4_2142	glutathione-dependent formaldehyde-activatingGFA	
8	2,522,087	2,526,716	4,629	PputUW4_2167	hypothetical protein	48.66
				PputUW4_2168	integrase family protein	
				PputUW4_2169	hypothetical protein	
				PputUW4_2170	hypothetical protein	
				PputUW4_2171	hypothetical protein	
9	2,593,656	2,597,985	4,329	PputUW4_2219	cytosine/purines uracil thiamine allantoinpermease	55.66
				PputUW4_2220	histone deacetylase	
				PputUW4_2221	LysR family transcriptional regulator	
				PputUW4_2222	lysine exporter protein LysE/YggA	
10	2,943,190	2,947,513	4,323	PputUW4_2534	hypothetical protein	49.35
				PputUW4_2535	hypothetical protein	
11	3,036,561	3,041,305	4,744	PputUW4_2600	2,4-dihydroxyacetophenone dioxygenase	63.92
				PputUW4_2601	hypothetical protein	
				PputUW4_2602	short chain dehydrogenase family protein	
				PputUW4_2603	thiolase	
				PputUW4_2604	AMP-dependent synthetase and ligase	
12	3,099,312	3,104,459	5,147	PputUW4_2652	LysR family transcriptional regulator	53.17
				PputUW4_2653	amino acid transporter	
				PputUW4_2654	cystathionine gamma-synthase	
				PputUW4_2655	AraC family transcriptional regulator	

				PputUW4_2727	hypothetical protein	
				PputUW4_2728	CadC family transcriptional regulator	
				PputUW4_2729	hypothetical protein	
				PputUW4_2730	hypothetical protein	
13	3,176,210	3,186,955	10,745	PputUW4_2731	AraC family transcriptional regulator	53.82
				PputUW4_2732	hypothetical protein	
				PputUW4_2733	OmpA-like transmembrane domain protein	
				PputUW4_2734	hypothetical protein	
				PputUW4_3195	two component LuxR family transcriptional regulator	
				PputUW4_3196	hypothetical protein	
				PputUW4_3197	IS110 family transposase	
14	3,737,337	3,743,674	6,337	PputUW4_3198	XRE family transcriptional regulator	58.82
				PputUW4_3199	hypothetical protein	
				PputUW4_3200	hypothetical protein	
				PputUW4_3201	3-oxoadipate enol-lactonase	
				PputUW4_3202	aldehyde dehydrogenase	
				PputUW4_3416	hypothetical protein	
				PputUW4_3417	hypothetical protein	
				PputUW4_3418	hypothetical protein	
15	3,966,752	3,970,895	4,143	PputUW4_3419	hypothetical protein	52.82
				PputUW4_3420	flagella synthesis regulator FleN	
				PputUW4_3421	hypothetical protein	
				PputUW4_3545	glycerol-3-phosphate cytidyltransferase	
				PputUW4_3546	polysaccharide biosynthesis protein	
16	4,116,806	4,123,634	6,828	PputUW4_3547	hypothetical protein	40.33
				PputUW4_3548	hypothetical protein	
				PputUW4_3549	glucosyltransferase	
				PputUW4_3550	glucosyltransferase	
				PputUW4_3585	integrase family protein	
				PputUW4_3586	hypothetical protein	
17	4,159,348	4,182,589	23,241	PputUW4_3587	hypothetical protein	54.4
				PputUW4_3588	hypothetical protein	
				PputUW4_3589	hypothetical protein	

				PputUW4_3590	hypothetical protein	
				PputUW4_3591	hypothetical protein	
				PputUW4_3592	transposon resolvase	
				PputUW4_3593	hypothetical protein	
				PputUW4_3594	hypothetical protein	
				PputUW4_3595	hypothetical protein	
				PputUW4_3596	hypothetical protein	
				PputUW4_3597	hypothetical protein	
				PputUW4_3598	hypothetical protein	
				PputUW4_3599	hypothetical protein	
				PputUW4_3600	NAD-dependent DNA ligase LigA	
				PputUW4_3601	cell division protein ZipA	
				PputUW4_3602	chromosome partition protein	
				PputUW4_3603	GntR family transcriptional regulator	
				PputUW4_3604	xanthine dehydrogenase small subunit	
				PputUW4_3605	xanthine dehydrogenase, molybdopterin-binding subunit B	
				PputUW4_4192	hypothetical protein	
				PputUW4_4193	cytosine-specific methyltransferase	
				PputUW4_4194	hypothetical protein	
				PputUW4_4195	histidine kinase	
18	4,818,892	4,826,804	7,912	PputUW4_4196	hypothetical protein	44.48
				PputUW4_4197	5-methylcytosine-specific restriction enzyme	
				PputUW4_4198	DNA mismatch endonuclease Vsr	
				PputUW4_4611	paraquat-inducible protein A	
19	5,281,019	5,288,766	7,748	PputUW4_4612	paraquat-inducible protein A	52.04
				PputUW4_4613	sulfite oxidase subunit YedZ	

Table 4-6. *P. putida* UW4 IS Elements

Family/Group	<i>P. putida</i> UW4 CDS ID
IS110	PputUW4_0474 PputUW4_0841 PputUW4_2025 PputUW4_4591 PputUW4_5362
IS110/IS1111	PputUW4_0318 PputUW4_1254 PputUW4_1997 PputUW4_3132 PputUW4_3197 PputUW4_3451 PputUW4_4228 PputUW4_4445 PputUW4_5136 PputUW4_5212 PputUW4_5335
IS1182	PputUW4_1560 PputUW4_2124 PputUW4_2136 PputUW4_2164 PputUW4_2166 PputUW4_2215 PputUW4_2223
IS3/IS3	PputUW4_0151 PputUW4_1502 PputUW4_1775 PputUW4_2015 PputUW4_3143 PputUW4_3925 PputUW4_4241 PputUW4_4349
<i>Truncated</i>	
IS110/IS1111	PputUW4_3001
IS1182	PputUW4_2132 PputUW4_2133
IS30	PputUW4_2127
IS5/IS427	PputUW4_3360
IS630	PputUW4_3551

Table 4-7. *P. putida* UW4 Phage Related CDSs

locus_tag	product
PputUW4_0456	HflK protein
PputUW4_0457	HflC protein
PputUW4_0670	phage integrase
PputUW4_0678	phage integrase
PputUW4_1008	GtrA family protein
PputUW4_1009	bactoprenol glucosyl transferase
PputUW4_1029	hypothetical protein
PputUW4_1125	phage repressor
PputUW4_1558	phage integrase family site specific recombinase
PputUW4_1764	phage repressor
PputUW4_1775	integrase catalytic protein
PputUW4_2419	cointegrate resolution protein T
PputUW4_2420	cointegrate resolution protein S
PputUW4_2781	phage integrase
PputUW4_3269	hypothetical protein
PputUW4_3361	phage integrase
PputUW4_3592	transposon resolvase
PputUW4_4196	hypothetical protein
PputUW4_5301	prophage antirepressor

Table 4-8. Tandem Repeats Identified in *P. putida* UW4

Start	End	Period Size	Consensus Size	Copy Number	Affected Gene or Region	Gene Product	Repeats Relative to CDS
43924	43964	21	22	1.9	41	oligopeptidase A	inside
121176	121255	27	28	3	114 (pseudo)	von Willebrand factor, type A	inside
121830	121859	6	6	5	114 (pseudo)	von Willebrand factor, type A	inside
124695	129038	318	317	13.7	114 (pseudo), 115, 116	von Willebrand factor, type A; conserved repeat domain protein; conserved repeat domain protein	inside
129077	129901	300	300	2.8	116	conserved repeat domain protein	inside
130252	130291	6	6	6.7	116	conserved repeat domain protein	inside
448990	449029	21	21	1.9	396	dihydrolipoamide acetyltransferase	inside
571073	571412	84	84	4	497	hypothetical protein	inside
634896	634947	15	14	3.7	555	acetyl-CoA carboxylase biotin carboxyl carrier protein subunit	inside
634923	634952	6	6	5	555	acetyl-CoA carboxylase biotin carboxyl carrier protein subunit	inside
1058842	1059321	123	123	3.9	903	hypothetical protein	inside
1237244	1237308	21	20	3.1	1043	methyl-accepting chemotaxis protein	inside
1373898	1373930	16	17	2	1170	hypothetical protein	inside
1408421	1408452	12	12	2.7	1199	hypothetical protein	inside
1479472	1479511	15	15	2.7	1266	hypothetical protein	inside
1479931	1480007	36	36	2.1	1267	hypothetical protein	inside

1807605	1807634	15	16	1.9	1550	Non-ribosomal peptide synthetase	inside
1865671	1865696	12	12	2.2	1584	Hpt sensor hybrid histidine kinase	inside
1920541	1921009	39	39	12	1630	hypothetical protein	inside
1939911	1939955	21	21	2.1	1649	hypothetical protein	inside
1940032	1940078	24	24	2	1649	hypothetical protein	inside
2029994	2030045	27	27	1.9	1725	heme peroxidase	inside
2030037	2030826	273	272	2.9	1725	heme peroxidase	inside
2030243	2030299	27	27	2.1	1725	heme peroxidase	inside
2030533	2030592	27	27	2.2	1725	heme peroxidase	inside
2067404	2067444	20	21	2	1756	AFG1-like ATPase	inside
2222958	2223231	39	39	7	1899	hypothetical protein	inside
2248389	2248418	15	15	2	1921	hypothetical protein	inside
2248428	2248469	15	15	2.8	1921	hypothetical protein	inside
2248458	2248521	30	30	2.2	1921	hypothetical protein	inside
2259311	2259560	84	84	3	1932	hypothetical protein	inside
2449938	2450234	33	33	9	2106	hypothetical protein	inside
2450214	2450295	18	18	4.6	2106	hypothetical protein	inside
2450218	2450303	36	36	2.4	2106	hypothetical protein	inside
2532903	2532935	6	6	5.3	2173	cyclic beta 1-2 glucan synthetase	inside
2544850	2544899	18	19	2.8	2181	hypothetical protein	inside
2544895	2544959	9	9	7.2	2181	hypothetical protein	inside
2562691	2563936	276	275	4.5	2190	sensor histidine kinase	inside
2600012	2600043	16	17	1.9	2224	RipR family transcriptional regulator	inside
2665103	2665175	27	27	2.7	2284	hemolysin-type calcium-binding region	inside

2665366	2665503	54	54	2.6	2284	hemolysin-type calcium-binding region	inside
2665925	2665996	27	27	2.7	2284	hemolysin-type calcium-binding region	inside
2682575	2682603	14	15	2	2301	3-oxoacyl-[acyl-carrier-protein] reductase	inside
2801188	2801654	123	123	3.8	2395	hypothetical protein	inside
2975571	2975623	24	24	2.2	2556	hypothetical protein	inside
2975578	2975624	24	24	2	2556	hypothetical protein	inside
3112030	3112056	12	12	2.2	2664	cation efflux family protein	inside
3144781	3144819	18	17	2.4	2692	hypothetical protein	inside
3154032	3154065	15	15	2.3	2701	hypothetical protein	inside
3176471	3176601	67	68	1.9	2727	hypothetical protein	inside
3510124	3510909	300	300	2.6	2993	hypothetical protein	inside
3511360	3512139	375	375	2.1	2993	hypothetical protein	inside
3641058	3641162	48	48	2.2	3101	hypothetical protein	inside
3641142	3641185	18	18	2.4	3101	hypothetical protein	inside
3644694	3645032	126	126	2.7	3106	hypothetical protein	inside
3644757	3644959	63	63	3.2	3106	hypothetical protein	inside
3738308	3738659	78	78	4.5	3196	hypothetical protein	inside
3741377	3741429	27	27	2	3199	hypothetical protein	inside
3741468	3741516	24	24	2	3199	hypothetical protein	inside
3846098	3846150	15	15	3.5	3304	hypothetical protein	inside
3846136	3846204	18	18	3.7	3304	hypothetical protein	inside
3846167	3846212	21	21	2.2	3304	hypothetical protein	inside
3846236	3846265	9	9	3.3	3304	hypothetical protein	inside
3846422	3846586	36	36	4.6	3304	hypothetical protein	inside
4050739	4050793	24	24	2.3	3489	copper-resistance protein CopA	inside
4051392	4051496	39	39	2.7	3490	copper resistance protein B	inside

4155899	4155941	21	21	2	3582	hypothetical protein	inside
4165503	4165535	15	16	2.1	3589	hypothetical protein	inside
4216459	4216692	63	63	3.7	3635	hypothetical protein	inside
4394420	4394470	18	18	2.8	3805	dihydrolipoamide succinyltransferase	inside
4421694	4421724	15	15	2.1	3829	urea transporter	inside
4449929	4449975	15	15	3.1	3860	chemotaxis sensor histidine kinase CheA	inside
4469374	4469413	12	12	3.3	3881	flagellar assembly protein H	inside
4531905	4531945	21	22	1.9	3939	TrkH family potassium uptake protein	inside
4560584	4560618	18	18	2	3957	ABC transporter, transmembrane region	inside
4654006	4654046	18	18	2.3	4047	ribonuclease E	inside
4654015	4654049	18	18	1.9	4047	ribonuclease E	inside
4654106	4654169	27	27	2.4	4047	ribonuclease E	inside
4718832	4718880	21	21	2.2	4110	GTP diphosphokinase	inside
4767521	4769427	462	462	4.1	4154	hypothetical protein	inside
4769292	4769562	135	135	2	4154	hypothetical protein	inside
4769427	4770439	462	463	2.2	4154	hypothetical protein	inside
4785020	4785089	34	34	2.1	4166	hypothetical protein	inside
4850751	4850780	15	15	2	4226	lysine exporter protein LysE/YggA	inside
4930376	4930450	27	24	3.1	4297	TolA colicin import membrane protein	inside
4930413	4930507	24	25	3.8	4297	TolA colicin import membrane protein	inside
4930462	4930537	24	24	3.2	4297	TolA colicin import membrane protein	inside

4930518	4930582	24	24	2.7	4297	TolA colicin import membrane protein	inside
4930537	4930591	24	21	2.5	4297	TolA colicin import membrane protein	inside
4930623	4930735	18	18	6.3	4297	TolA colicin import membrane protein	inside
5038760	5038790	9	9	3.4	4391	chaperonin GroEL	inside
5316884	5316909	12	12	2.2	4639	glucose-6-phosphate isomerase	inside
5439683	5439712	15	15	2	4756	hypothetical protein	inside
5439985	5440013	12	12	2.4	4756	hypothetical protein	inside
5550454	5550510	27	27	2.1	4858	single-stranded DNA-binding protein	inside
5727150	5727183	12	12	2.8	5019	hypothetical protein	inside
5797511	5797551	21	21	2	5080	formate/nitrate transporter	inside
5803893	5803923	12	12	2.6	5085	ATP-dependent RNA helicase	inside
5821054	5821090	12	12	3.1	5104	signal transducer, CheW	inside
5864843	5864867	12	12	2.1	5147	aldehyde dehydrogenase	inside
5914851	5915078	48	48	4.8	5195	S-type Pyocin	inside
5919408	5919444	15	16	2.5	5197	AraC family transcriptional regulator	inside
5971992	5972025	15	16	2.2	5248	TonB domain-containing protein	inside
5997825	5997883	30	30	2	5272	hypothetical protein	inside
5997897	5998087	24	24	7.7	5272	hypothetical protein	inside
6033004	6033039	18	18	2	5308	xanthine phosphoribosyltransferase	inside
6069812	6069848	15	16	2.4	5346	TonB-like protein	inside
1948571	1948892	133	134	2.4	1660	hypothetical protein	inside and downstream

3574756	3575299	260	260	2.1	3048	hypothetical protein	inside and downstream
3768552	3768922	123	123	3	3226	hypothetical protein	inside and downstream
3956354	3956411	17	17	3.5	3404	LacI family transcription regulator	inside and downstream
4134951	4135470	284	284	1.8	3561	aerotaxis receptor	inside and downstream
4785054	4785456	123	123	3.3	4166	hypothetical protein	inside and downstream
1954356	1955116	263	265	2.9	1666	hypothetical protein	inside and upstream
3156644	3157707	123	123	8.7	2705	hypothetical protein	inside and upstream
3176187	3176540	123	123	2.9	2727	hypothetical protein	inside and upstream
3235498	3235719	123	123	1.8	2777	hypothetical protein	inside and upstream
3593725	3594335	136	136	4.5	3064	hypothetical protein	inside and upstream
3680599	3680921	133	133	2.4	3142	hypothetical protein	inside and upstream
4046312	4046894	123	123	4.7	3486	hypothetical protein	inside and upstream
3729837	3730254	133	133	3.1	3189	hypothetical protein	inside, upstream and downstream
4858915	4859409	112	112	4.4	4234	hypothetical protein	inside, upstream and downstream
305869	306071	79	80	2.5		intergenic region	
1002063	1002092	14	14	2.1		intergenic region	

1374095	1374169	19	19	3.9	intergenic region
1824758	1824840	18	18	4.5	intergenic region
1824936	1824975	19	19	2.1	intergenic region
1916677	1916736	6	6	10	intergenic region
2176903	2177229	129	131	2.5	intergenic region
2187851	2187894	17	17	2.6	intergenic region
2196382	2196412	14	14	2.2	intergenic region
2229075	2229120	15	15	3	intergenic region
2354303	2354650	134	134	2.6	intergenic region
2355824	2356096	135	134	2	intergenic region
2375568	2375860	131	132	2.2	intergenic region
2388278	2388645	133	132	2.8	intergenic region
2436141	2436172	15	15	2.1	intergenic region
2450520	2450623	33	33	3.2	intergenic region
2471329	2471355	6	6	4.5	intergenic region
2546335	2546418	18	18	4.7	intergenic region
2546336	2546400	21	21	3.1	intergenic region
2546351	2546388	15	15	2.5	intergenic region
2546683	2546763	9	9	9	intergenic region
2546786	2546882	33	33	2.9	intergenic region
2546929	2547042	33	33	3.5	intergenic region
2546930	2547092	66	66	2.5	intergenic region
2546944	2546985	21	21	2	intergenic region
2547120	2547179	18	18	3.3	intergenic region
2565588	2565646	11	11	5.4	intergenic region
2646854	2647217	123	123	3	intergenic region
2707912	2707959	16	16	2.9	intergenic region
2744769	2744999	123	123	1.9	intergenic region
2759436	2759466	15	14	2.1	intergenic region
2888440	2888575	68	68	2	intergenic region

2994262	2994913	123	123	5.3	intergenic region
3022913	3023148	125	125	1.9	intergenic region
3110014	3110072	8	8	7.4	intergenic region
3129909	3129952	15	15	2.9	intergenic region
3138106	3138445	123	123	2.8	intergenic region
3279148	3279185	19	19	2	intergenic region
3354637	3354674	19	19	2	intergenic region
3408820	3408861	18	18	2.3	intergenic region
3421369	3422102	123	123	5.9	intergenic region
3432149	3432476	123	123	2.7	intergenic region
3577587	3577730	68	68	2.1	intergenic region
3577672	3578137	123	123	3.8	intergenic region
3591390	3591761	125	125	3	intergenic region
3654879	3654911	17	17	1.9	intergenic region
3655013	3655045	17	17	1.9	intergenic region
3671332	3671585	123	123	2.1	intergenic region
3698756	3699197	169	167	2.6	intergenic region
3751285	3751347	19	19	3.4	intergenic region
3756009	3756048	19	19	2.1	intergenic region
3784412	3784469	28	28	2.1	intergenic region
4128189	4128229	18	18	2.3	intergenic region
4731531	4731562	16	16	2	intergenic region
4764716	4764748	17	17	1.9	intergenic region
4772590	4772615	9	9	2.9	intergenic region
5015287	5015322	18	18	2	intergenic region
5262394	5262754	141	141	2.6	intergenic region
5755701	5755743	21	21	2	intergenic region
6127819	6127848	14	14	2.1	intergenic region

coding region, which may act as promoter components of downstream genes or transcription terminators of upstream genes (Yeramian and Buc 1999; Usdin 2008).

In order to elucidate the protein function of the 5431 CDSs, protein localization prediction was performed. The results indicate that UW4 consists of 2509 (46%) cytoplasmic proteins, 1259 (23%) cytoplasmic membrane proteins, 176 (3%) periplasmic proteins, 115 (2%) outer membrane proteins, and 44 (1%) extracellular proteins. The remaining 1328 (25%) CDSs have unknown localization (Fig. 4-4 and Table 4-9).

## **4.2 Genes Involved in Plant Growth Promotion and *P. putida* UW4 Lifestyle**

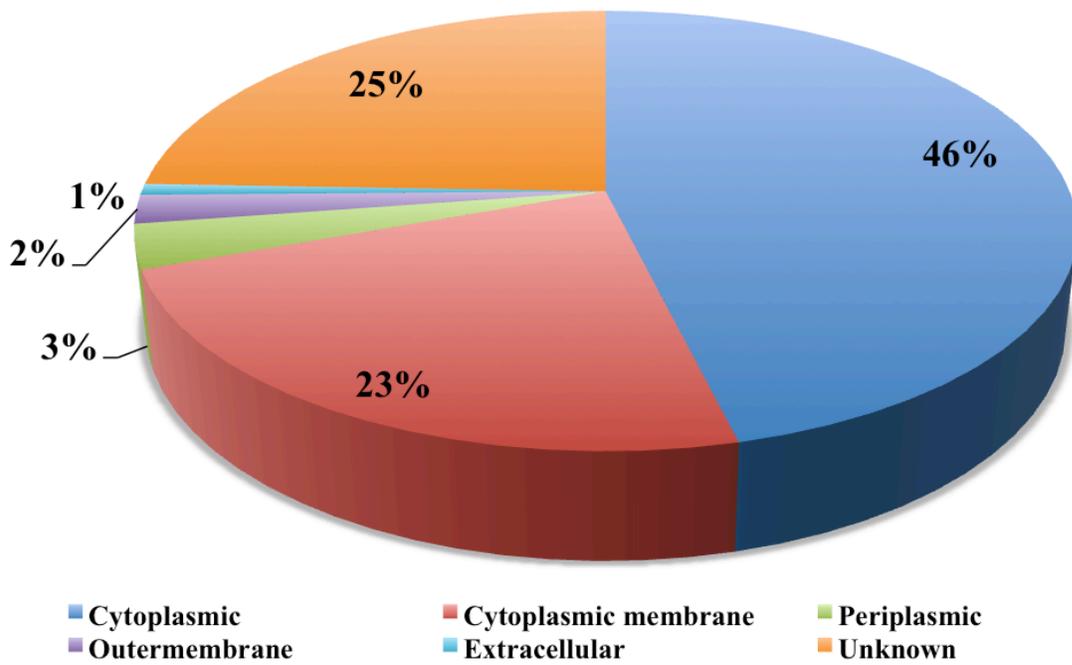
### **4.2.1 ACC deaminase**

The ACC deaminase gene, *acdS* (PputUW4\_4159), and its upstream regulatory gene, *acdR* (PputUW4\_4160), were characterized previously (Shah et al. 1998; Grichko and Glick 2000; Cheng et al. 2008). UW4 genome sequencing confirmed the presence of both genes as well as the intergenic sequences between the two genes. Interestingly, examination of genes downstream of *acdS* revealed that PputUW4\_4153 and 4155 were found only in other genera but not in any pseudomonads, and PputUW4\_4154 and 4157 are genes that are unique to UW4. In addition, a tRNA-Arg gene was found 3,138 bp downstream of *acdR* and tRNA-Arg is one of tRNA genes that are preferentially used for insertion of GI (Langille et al. 2010). Therefore, it is possible that the region of the genome that encodes *acdS* and *acdR* was acquired from other genera by horizontal gene transfer, which is in concert with previous findings (Hontzeas et al. 2005). However, this region was not identified as a GI by automatic prediction using IslandViewer.

Table 4-9. *P. putida* UW4 Predicted Protein Localizations

Protein Localization	Number
Cytoplasmic	2509
Cytoplasmic membrane	1259
Periplasmic	176
Outer membrane	115
Extracellular	44
Unknown	1328
Total	5431

Fig. 4-4. *P. putida* UW4 predicted protein localizations.



### 4.2.2 Siderophores

*P. putida* UW4 fluoresces under UV light, consistent with the production of pyoverdine siderophore by this organism. In the UW4 genome, putative genes associated with pyoverdine synthesis are shown in Table 4-10 and Fig. 4-5. The *pvdF* gene for the type I pyoverdine found in *P. aeruginosa* is absent in UW4. This gene encodes a transformylase responsible for the formation of  $N^5$ -formyl- $N^5$ -hydroxyornithine from  $N^5$ -hydroxyornithine (McMorran et al. 2001). However, in UW4, a gene encoding hydroxyornithine acetylase, *pvdYII*, was found. The gene product of *pvdYII* can convert  $N^5$ -hydroxyornithine to *N*-hydroxy-*cyclo*-ornithine, resulting in the production of type II pyoverdine in *P. aeruginosa* (Lamont et al 2006) (Fig. 4-6). Amino acids sequence alignment of PvdYII from UW4 and *P. aeruginosa* Pa4 (ABC55668) showed that the two proteins share 70% identities and 79% similarities, and that conservation occurs to the greatest extent at the C-terminus (Fig. 4-7).

### 4.2.3 IAA production

It was demonstrated previously that *P. putida* UW4 actively produces the phytohormone IAA (Saleh and Glick 2001). Here, two potential IAA biosynthesis pathways were identified in the genome of UW4; they are the IAM pathway and the IAN pathway (Fig. 4-8). However, the IPyA pathway, which was identified in another PGPB, *Pseudomonas putida* GR12-2 (Patten and Glick, 2002), is absent in UW4. In the IAM pathway, tryptophan is converted to IAM by tryptophan 2-monooxygenase (PputUW4\_4967, 4535) and then to IAA by amidase (PputUW4\_3355). In the IAN pathway, tryptophan is converted to indole-3-acetaldoxime and then to IAN by indoleacetaldoxime dehydratase (PputUW4\_3353). Next, IAA can be produced directly through IAN by nitrilase (PputUW4\_2466). Alternatively, IAN can be first converted to IAM by nitrile hydratase (PputUW4\_3356 and PputUW4\_3357), and then IAM is converted to

Fig. 4-5. Pyoverdine synthesis genes in *P. putida* UW4. Genes are not drawn to scale and are oriented according to the direction of transcription.

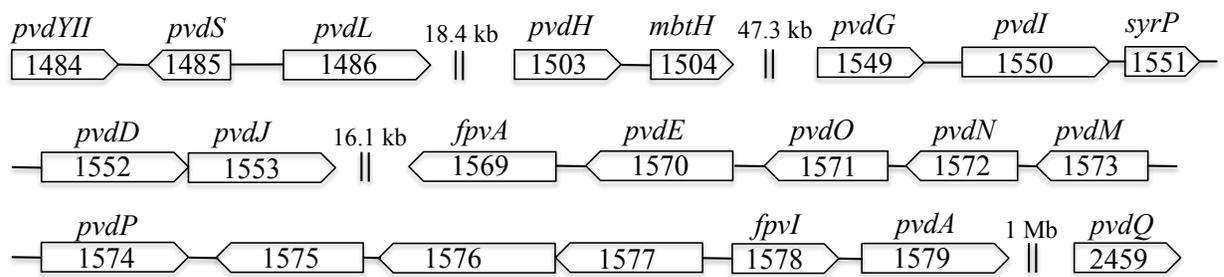


Table 4-10. Genes associated with pyoverdine synthesis in *P. putida* UW4.

CDS ID	Gene	Function
1484	<i>pvdYII</i>	Hydroxyornithine acetylase
1485	<i>pvdS</i>	ECF iron sigma factor
1486	<i>pvdL</i>	Non-ribosomal peptide synthetase
1503	<i>pvdH</i>	Aminotransferase
1504	<i>mbtH</i>	Unknown
1549	<i>pvdG</i>	Thioesterase
1550	<i>pvdI</i>	Non-ribosomal peptide synthetase
1551	<i>syrP</i>	Pyoverdine biosynthesis regulatory protein
1552	<i>pvdD</i>	Non-ribosomal peptide synthetase
1553	<i>pvdJ</i>	Non-ribosomal peptide synthetase
1569	<i>fpvA</i>	TonB-dependent pyoverdine receptor
1570	<i>pvdE</i>	ABC transporter
1571	<i>pvdO</i>	Unknown
1572	<i>pvdN</i>	Aminotransferase
1573	<i>pvdM</i>	Dipeptidase
1574	<i>pvdP</i>	Unknown
1575		Outer membrane efflux protein
1576		ABC transporter
1577		Membrane efflux protein
1578	<i>fpvI</i>	ECF sigma factor required for expression of <i>fpvA</i>
1579	<i>pvdA</i>	L-ornithine hydroxylase
2459	<i>pvdQ</i>	Acylase

Fig. 4-6. *P. putida* UW4 predicted pyoverdine biosynthesis pathway.

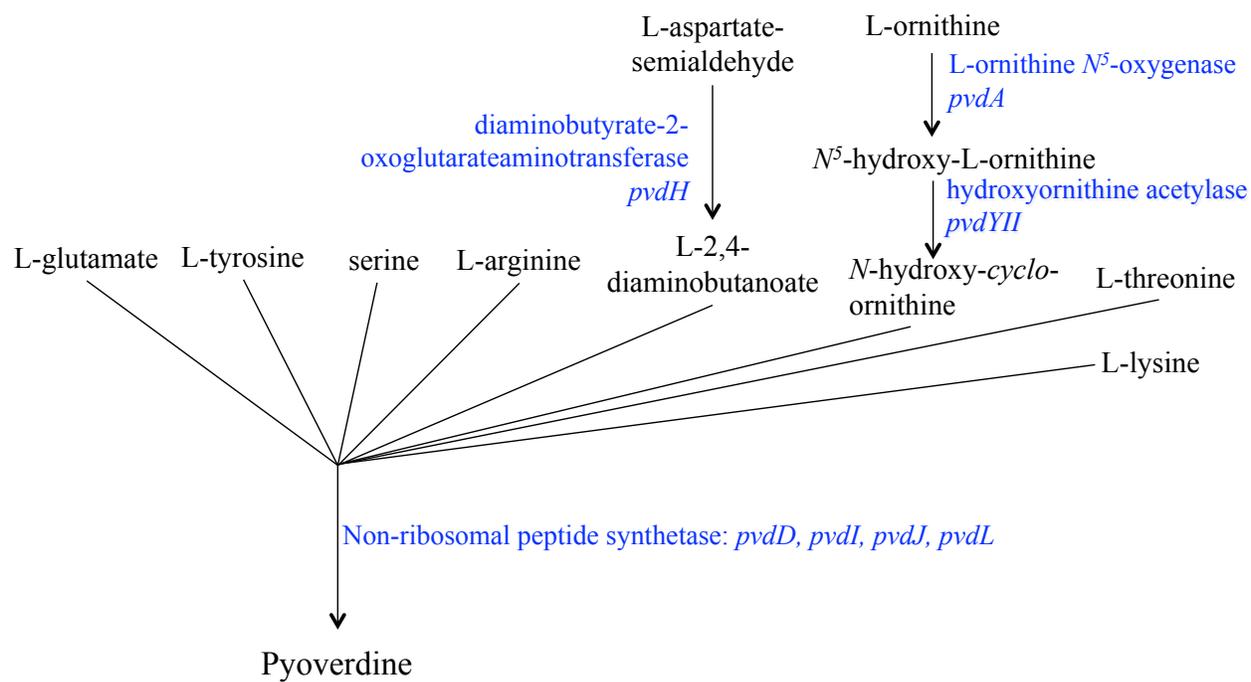


Fig. 4-7. Amino acid sequence alignment of *pvdYII* in *P. putida* UW4 and *P. aeruginosa* (Accession no. ABC55668).

	1	10	20	30																																			
pvdYII_uw4	M	S	N	L	N	D	L	T	A	L	C	L	P	T	G	S	R	L	A	A	D	E	T	O	G	H	L	S	L	S	L	E	G	Q	P	L	L	T	L
pvdYII_Pae	M	P	L	L	N	S	L	T	V	L	P	L	P	G	G	O	R	L	S	G	D	E	T	D	S	R	L	S	L	A	L	D	G	A	L	L	L	R	L
	40	50	60	70																																			
pvdYII_uw4	R	L	N	R	-	G	P	E	L	H	V	Q	L	L	E	H	F	S	H	P	E	G	Q	A	I	W	A	A	C	Y	W	L	F	A	R	N	P	D	C
pvdYII_Pae	R	L	E	R	T	D	T	E	I	R	L	A	V	E	E	L	P	G	G	D	A	E	H	A	L	W	A	A	S	Y	W	L	F	A	R	D	P	D	C
	80	90	100	110																																			
pvdYII_uw4	Q	R	Q	V	W	Q	L	O	D	V	P	H	E	A	V	L	S	G	L	L	I	P	T	S	V	N	G	Q	Y	C	C	E	R	T	L	F	W	Q	L
pvdYII_Pae	Q	R	L	V	W	S	L	S	W	E	P	E	E	S	L	A	K	G	M	L	T	G	A	A	A	G	T	Y	V	C	E	R	G	R	F	W	Q	L	
	120	130	140	150																																			
pvdYII_uw4	P	O	P	W	L	G	R	S	L	T	G	S	Y	P	Q	Q	M	V	M	T	G	K	R	H	P	L	R	A	V	K	P	R	G	E	V	Y	R	R	
pvdYII_Pae	P	W	P	W	L	P	L	A	-	Q	P	C	Y	P	Q	S	M	Q	L	T	D	G	K	R	H	P	L	R	R	P	K	P	R	G	E	V	Y	R	R
	160	170	180	190																																			
pvdYII_uw4	F	D	A	R	L	G	A	W	V	S	L	R	T	L	E	I	E	O	D	L	A	R	F	N	R	W	O	N	S	P	R	V	A	S	F	W	Q	E	E
pvdYII_Pae	F	D	A	R	L	G	C	W	I	S	L	R	T	L	D	I	E	R	D	L	E	R	F	H	R	W	O	N	S	P	R	V	A	E	F	W	Q	E	D
	200	210	220	230																																			
pvdYII_uw4	G	S	L	E	Q	H	R	Q	Y	L	S	K	L	E	A	D	P	H	T	L	T	L	I	G	C	F	D	D	O	P	F	A	Y	F	E	A	Y	W	A
pvdYII_Pae	G	S	L	E	Q	H	R	E	R	L	R	Q	L	E	A	D	P	H	A	L	G	L	I	G	C	F	D	D	E	P	F	A	Y	F	E	A	Y	W	A
	240	250	260	270																																			
pvdYII_uw4	K	E	D	R	I	A	P	F	Y	D	A	G	D	Y	D	R	G	I	H	M	L	V	G	E	E	N	H	R	G	P	H	K	V	A	S	W	L	S	A
pvdYII_Pae	K	E	D	R	I	A	P	F	Y	V	A	E	D	Y	D	R	G	I	H	M	L	V	G	E	E	A	H	R	G	P	H	K	V	A	C	W	L	S	A
	280	290	300	310																																			
pvdYII_uw4	L	V	H	L	F	L	D	D	P	R	T	Q	R	V	V	A	E	P	R	A	D	N	A	K	M	I	G	Y	M	H	N	Q	C	F	H	C	D	K	
pvdYII_Pae	L	V	H	L	F	L	D	D	P	R	S	Q	R	V	V	A	E	P	R	A	D	N	A	R	M	I	G	H	M	H	N	L	C	F	H	C	E	K	
	320	330	338																																				
pvdYII_uw4	E	F	D	F	P	H	K	R	A	A	L	M	V	L	G	R	E	R	E	F	F	D	R	C	A	L	A												
pvdYII_Pae	E	F	D	F	P	H	K	R	A	A	L	M	I	I	G	R	E	R	E	F	F	E	R	C	R	L	A												

IAA by amidase (PputUW4\_3355). The functions of the above mentioned genes in IAA biosynthesis are putative and they need to be confirmed experimentally.

#### **4.2.4 Trehalose**

Trehalose is a non-reducing disaccharide of glucose whose two glucose moieties are linked by an  $\alpha,\alpha$ -1,1-glycosidic bond. It functions as an osmoprotectant in the stabilization of biological structures including dehydrated enzymes, proteins and lipids under environmental stresses such as drought, high salinity and low temperature in a wide range of organisms, i.e. bacteria, archaea, fungi, invertebrates, insects and plants. In transgenic rice, trehalose improves the plant's abiotic stress tolerance (Garg et al. 2002). In another study, when maize plants were inoculated with a strain of *Azospirillum brasilense* transformed to overexpress trehalose, 85% of the plants survived drought stress, whereas only 55% of the plants inoculated with the non-transformed strain survived. Furthermore, a 73% increase in the biomass of maize plants was obtained when the plants were inoculated with the transformed strain (Rodríguez-Salazar et al. 2009). In the genome of UW4, two trehalose synthesis pathways, TreS and TreY-TreZ pathways, were identified (Fig. 4-10). The TreS pathway involves the conversion of maltose to trehalose by trehalose synthase (TreS) (PputUW4\_2805). In the TreY-TreZ pathway, maltodextrin is first converted to maltooligosyltrehalose by maltooligosyltrehalose synthase (TreY) (PputUW4\_2797), and then to trehalose by maltooligosyltrehalose trehalohydrolase (TreZ) (PputUW4\_2795).

#### **4.2.5 Acetoin and 2,3-butanediol**

Fig. 4-8. *P. putida* UW4 IAA biosynthesis pathways.

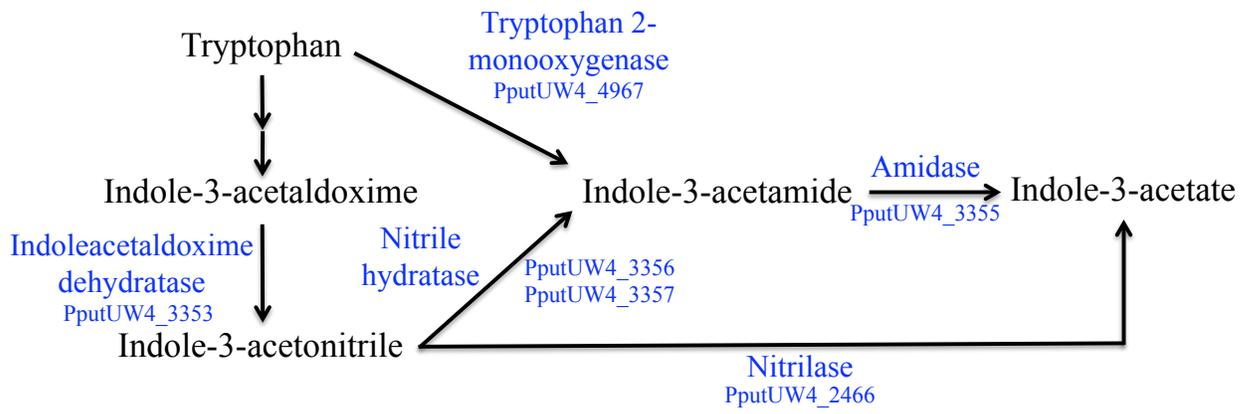
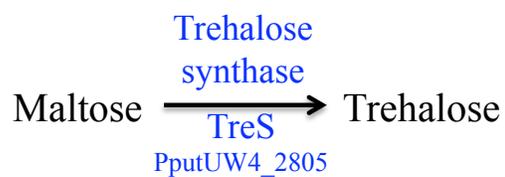
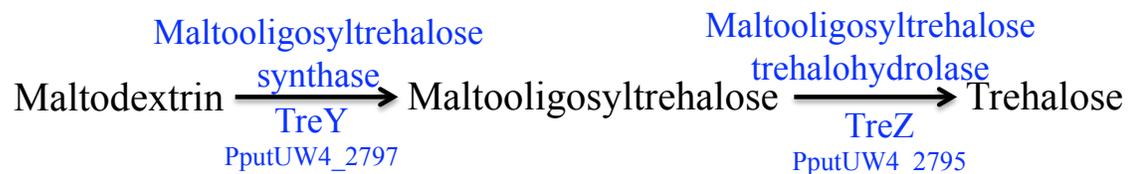


Fig. 4-9. *P. putida* UW4 trehalose biosynthesis pathways.

### TreS pathway



### TreY-TreZ pathway



Acetoin and 2,3-butanediol are volatile compounds released from certain PGPB, which can promote plant growth by stimulating root formation (Ryu et al. 2003). In the genome of UW4, genes involved in acetoin production were identified, including acetolactate synthase (PputUW4\_4617 and PputUW4\_4618) and zinc-containing alcohol dehydrogenase (PputUW4\_3051) (Fig. 4-10). However, enzymes responsible for 2,3-butanediol synthesis are absent from the UW4 genome.

#### 4.2.6 Antimicrobial compounds and antibiotics resistance

It has been reported that 4-hydroxybenzoate has antimicrobial activity and its biosynthesis pathway has been found in the genome of several PGPB, such as *Pseudomonas fluorescens* Pf-5 (Paulsen et al. 2005), *Enterobacter* sp. 638 (Taghavi et al. 2010) and *Mesorhizobium amorphae* (Hao et al. 2012). In addition, a complete pathway of 4-hydroxybenzoate synthesis from chorismate was identified in the genome of UW4 (Fig. 4-11).

Antibiotic susceptibility testing of UW4 has shown that it is resistant to ampicillin (128 µg/ml), erythromycin (64 µg/ml), and novobiocin (256 µg/ml). Two genes that encode β-lactamase were found in the UW4 genome (PputUW4\_1226 and PputUW4\_1639), which may confer the ampicillin resistance of the strain. One gene that encodes macrolide glycosyltransferase (PputUW4\_3151) was identified; the product of this gene can glycosylate and inactivate macrolide antibiotics such as erythromycin (Bolam et al. 2007). Novobiocin is produced by *Streptomyces* and this antibiotic's target is DNA gyrase subunit B (Gellert et al. 1976; Thiara and Cundliffe 1988). There are two mechanisms used by bacteria to inhibit novobiocin activity. One strategy is through the mutation of the gyrase B (*gyrB*) subunit gene (Thiara and Cundliffe 1993). For example, *Streptomyces sphaeroides* has two *gyrB* genes and

Fig. 4-10. *P. putida* UW4 acetoin synthesis pathway.

## Glycolysis I



pyruvate



acetolactate synthase  
PputUW4\_4617 and PputUW4\_4618

2-acetolactate



spontaneous

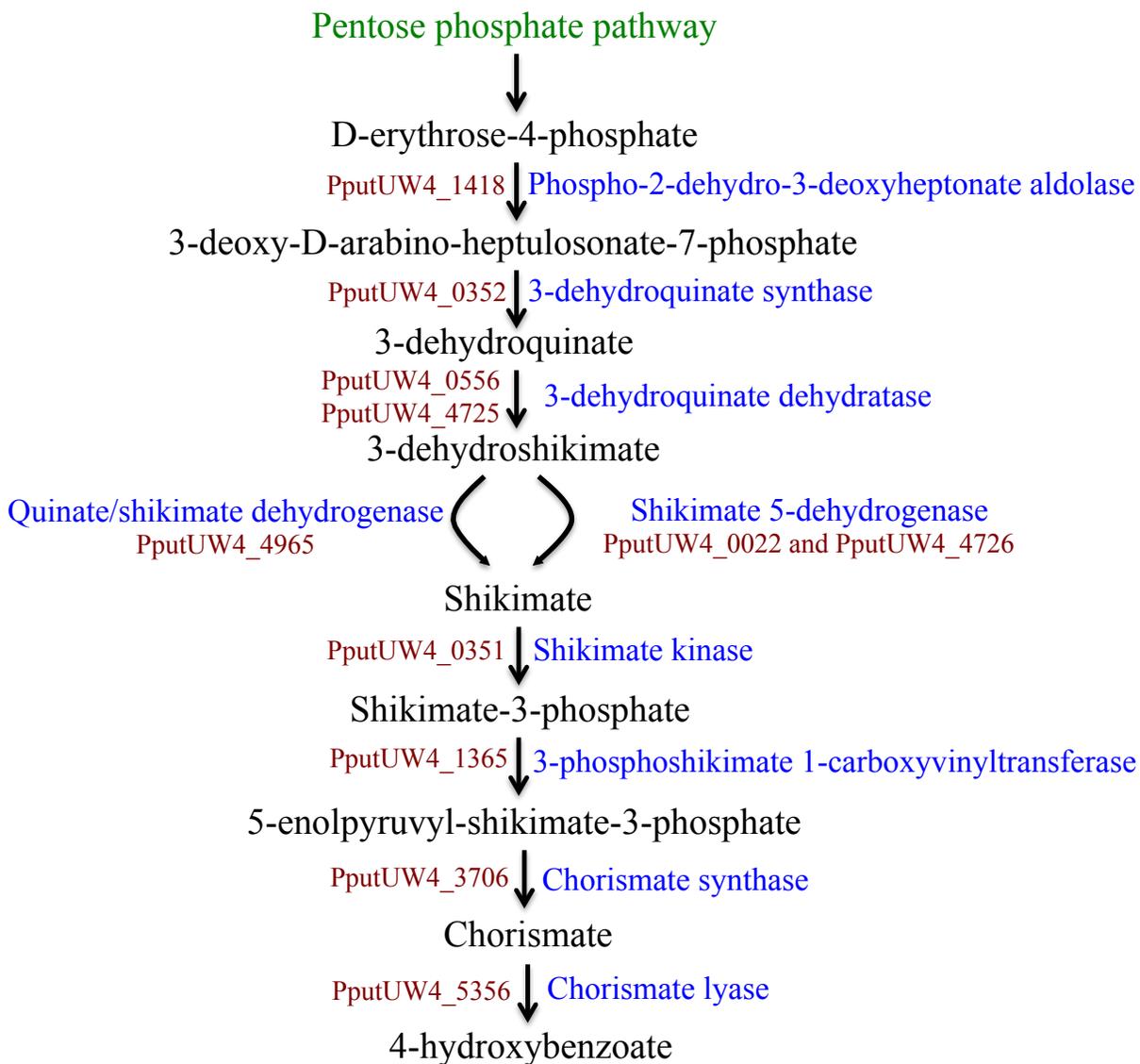
diacetyl



Zinc-containing alcohol  
dehydrogenase  
PputUW4\_3051

acetoin

Fig. 4-11. *P. putida* UW4 4-hydroxybenzoate synthesis pathway.



one is novobiocin sensitive, which is constitutively produced, and the other one is novobiocin resistant, which is induced by the drug (Thiara and Cundliffe 1989). The second strategy of novobiocin resistance is through the use of multidrug efflux pumps (Poole 2001). *P. putida* UW4 has a single *gyrB* gene in its genome. The product of this gene hasn't been characterized to see if it is novobiocin sensitive or resistant. On the other hand, multiple multidrug efflux systems have been identified in the UW4 genome based on sequence similarity search, which may play an important role in novobiocin resistance (Table 4-11).

#### **4.2.7 Polyhydroxyalkanoates biosynthesis**

Polyhydroxyalkanoates (PHAs) are a group of metabolic energy and carbon storage compounds that are deposited as intracellular water-insoluble granules in many living organisms during imbalanced growth conditions (Verlinden et al. 2007). PHAs extracted from bacteria can be used as alternative starting materials to petrochemical in the synthesis of plastics because they are biodegradable and environmentally friendly (Madison and Huisman 1999). Furthermore, bacteria can accumulate PHAs to levels as high as 90% (w/w) of the dry cell mass, making them potential candidates for the large-scale production of PHAs (Steinbüchel and Lütke-Eversloh 2003). Recently, it has been reported that PHA production played an important role in cold adaptation of an Antarctic bacterium *Pseudomonas* sp. 14-3, likely by alleviating the oxidative stress induced by cold environments (Ayub et al. 2009). Thus, the PHA synthase-minus mutant of *Pseudomonas* sp. 14-3 could not grow at 10°C and was more susceptible to freezing than the wild-type strain. In addition, cold shock treatment caused rapid degradation of PHA in the wild-type strain (Ayub et al. 2009).

Table 4-11. *P. putida* UW4 Multidrug Efflux Systems.

PputUW4	Product
0125	multidrug efflux MFS transporter
0126	multidrug efflux MFS membrane fusion protein
0127	multidrug efflux MFS outer membrane protein
0135	fusaric acid resistance protein
0137	multidrug resistance efflux pump protein
0138	multidrug efflux outer membrane protein
0180	multidrug RND transporter, membrane fusion protein
0181	multidrug RND transporter, membrane fusion protein
0182	acriflavin resistance protein
0503	small multidrug resistance protein SugE
1069	multidrug RND transporter, membrane fusion protein
1070	acriflavin resistance protein
1214	multidrug efflux RND outer membrane protein
1215	multidrug efflux RND inner membrane transporter
1216	multidrug efflux RND membrane fusion protein
1594	EmrB/QacA family drug resistance transporter
1595	multidrug resistance protein (HlyD family), EmrA-like secretion
1897	multidrug resistance efflux protein
1898	EmrB/QacA subfamily drug resistance transporter
2114	multidrug efflux system outer membrane protein
2115	RND family efflux transporter MFP subunit
2116	acriflavin resistance protein
2144	acriflavin resistance protein
2145	RND family efflux transporter MFP subunit
2146	RND efflux system outer membrane lipoprotein
2443	multidrug efflux system transmembrane protein
2444	multidrug efflux RND membrane fusion protein
2536	fusaric acid resistance protein
2538	multidrug resistance efflux pump protein
2539	multidrug efflux system outer membrane protein
2540	RND multidrug efflux transporter
2541	RND family efflux transporter MFP subunit
2815	multidrug efflux system outer membrane protein
2816	multidrug efflux system transmembrane protein
2817	multidrug efflux system transmembrane protein
2818	multidrug efflux system inner membrane protein
2954	ABC efflux system outer membrane protein
2955	ABC efflux system ATP-binding protein
2956	ABC efflux system permease

3471	multidrug resistance efflux pump
3472	EmrB/QacA family drug resistance transporter
3533	multidrug ABC transporter ATPase/permease
3536	multidrug efflux transporter
3756	small multidrug resistance protein
3988	multidrug ABC transporter ATP-binding protein/permease
4061	ABC-type multidrug transport system, permease component
4062	ABC-type multidrug transport system, ATPase component
4383	RND family multidrug transporter membrane fusion protein
4384	RND multidrug efflux transporter
4491	fusaric acid resistance protein
4493	fusaric acid resistance protein
4494	RND efflux system, outer membrane lipoprotein
4496	RND family multidrug transporter membrane fusion protein
4497	multidrug efflux system transmembrane protein
5156	multidrug resistance transporter, Bcr/CflA family
5306	multidrug resistance-transport membrane protein
5424	permease
5425	multidrug efflux system outer membrane protein
5426	multidrug resistance efflux pump
5427	membrane protein

In the genome of *P. putida* UW4, Two PHA synthesis and one degradation pathway were identified (Fig. 4-12). In the first synthesis pathway, two molecules of acetyl-CoA are combined by  $\beta$ -ketothiolase (PhaA) to produce acetoacetyl-CoA. In the presence of NADH, acetoacetyl-CoA reductase (PhaB) catalyzes the reduction of acetoacetyl-CoA to  $\beta$ -hydroxybutyryl-CoA. Then, PHA synthase (PhaC) polymerizes  $\beta$ -hydroxybutyryl-CoA to PHA, releasing coenzyme-A. The second PHA synthesis pathway in UW4 is through fatty acid de novo biosynthesis. One intermediate of fatty acid de novo synthesis,  $\beta$ -hydroxyacyl-ACP, is converted to the corresponding CoA-derivative by 3-hydroxydecanoyl-ACP-CoA transacylase (PhaG). Finally, PHA synthase (PhaC) catalyzes the formation of PHA through the substrate,  $\beta$ -hydroxyacyl-CoA. The PHA degradation pathway in UW4 is catalyzed by PHA depolymerase (PhaZ). PhaZ first depolymerizes PHA to  $\beta$ -hydroxybutyrate monomers, which are subsequently converted to acetoacetate by  $\beta$ -hydroxybutyrate dehydrogenase. Then, acetoacetate is recycled by acetoacetyl-CoA synthetase to form acetoacetyl-CoA.

#### **4.2.8 Degradation of aromatic compounds**

In the genome of UW4, a complete degradation pathway of benzoate via the catechol route of the  $\beta$ -keto adipate pathway was identified (Fig. 4-13). In addition, the protocatechuate branch of the  $\beta$ -keto adipate pathway is also present. Protocatechuate is one of the key intermediates during the degradation of various aromatic compounds, including 4-hydroxybenzoate and quinate (Fig. 4-14) (Jiménez et al. 2004). A putative degradation pathway of 3-hydroxyphenylpropionate (3-HPP) was described for *P. putida* W619, and the complete pathway includes the enzymes encoded within the *mhpRABCFET* operon (Wu et al. 2011). When searching these eight genes within the genome sequence of UW4, five putative enzyme-

Fig. 4-12. *P. putida* UW4 PHA biosynthesis and degradation pathways. Black arrows indicate the biosynthesis pathways. Red arrows represent the degradation pathway. PhaA,  $\beta$ -ketothiolase (PputUW4\_1998); PhaB, acetoacetyl-CoA reductase (PputUW4\_4038); PhaC, PHA synthase (PputUW4\_0333, 0335, and 2305); PhaG, 3-hydroxydecanoyl-ACP-CoA transacylase (PputUW4\_1255); PhaZ, PHA depolymerase (PputUW4\_0334, 2306); FabD, malonyl CoA-acyl carrier protein transacylase (PputUW4\_4039); FabB, 3-oxoacyl-(acyl carrier protein) synthase I (PputUW4\_0363 and 4105); FabF, 3-oxoacyl-(acyl carrier protein) synthase II (PputUW4\_0360, 1386 and 4036); FabG, 3-oxoacyl-(acyl carrier protein) reductase (PputUW4\_0361, 0576, 2301, 2241, 3674, 4038, 4217, 4395); FabI, enoyl-(acyl-carrier-protein) reductase (PputUW4\_2893); FabA, 3-hydroxydecanoyl-(acyl carrier protein) dehydratase (PputUW4\_0362 and 4104).  
Acetyl-CoA carboxylase: PputUW4\_0554, 0555, 1107, 3446;  $\beta$ -hydroxybutyrate dehydrogenase: PputUW4\_2378; Acetoacetyl-CoA synthetase: PputUW4\_1999 and 2000.

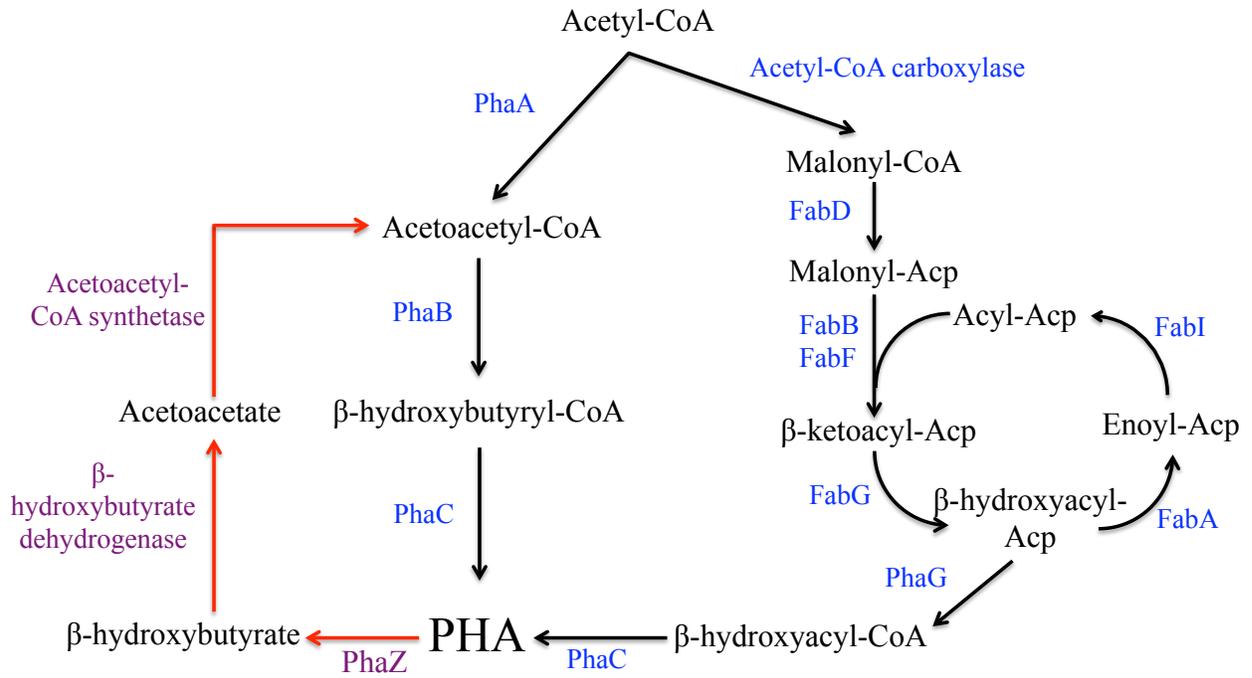


Fig. 4-13. Benzoate degradation pathway in *P. putida* UW4.

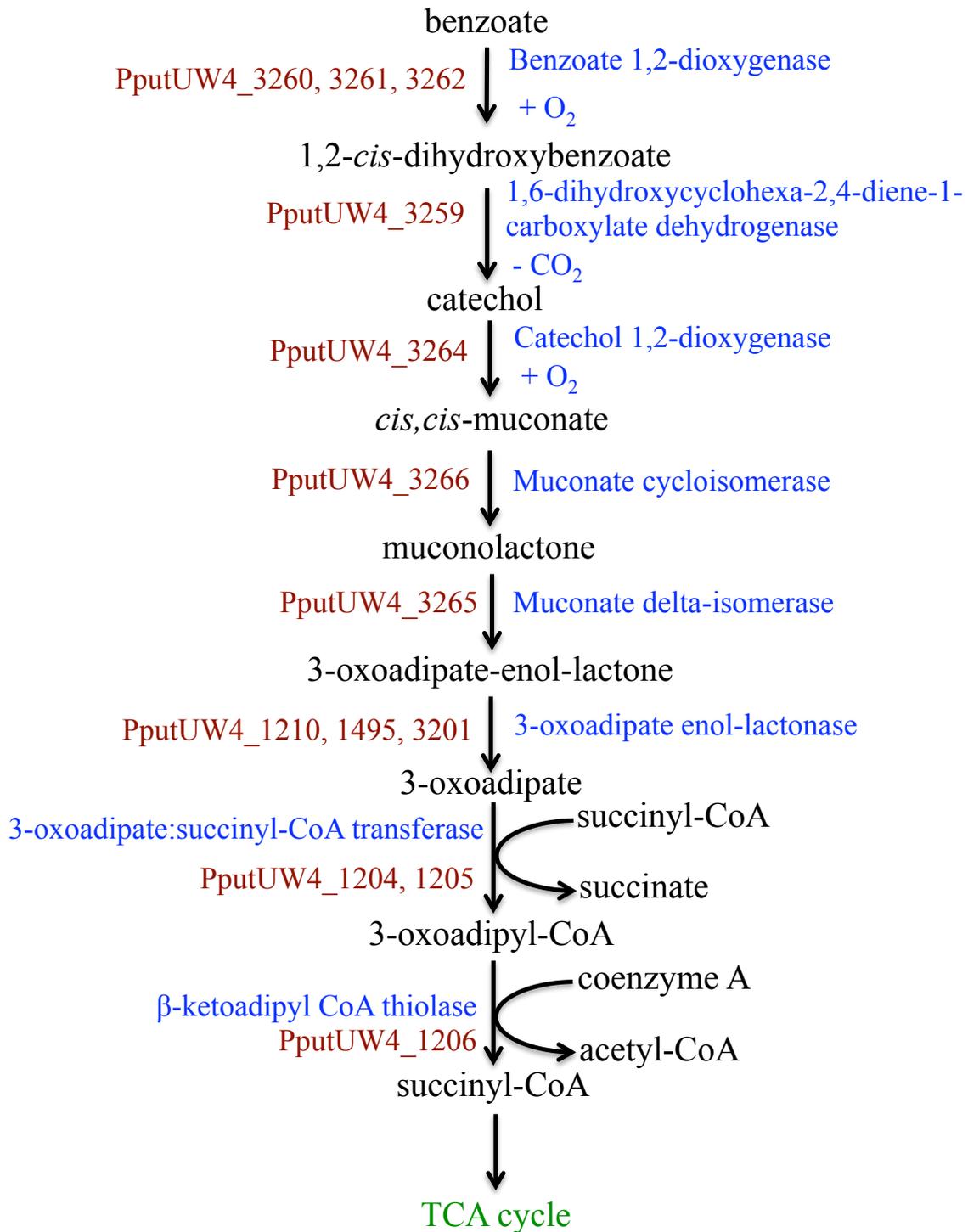
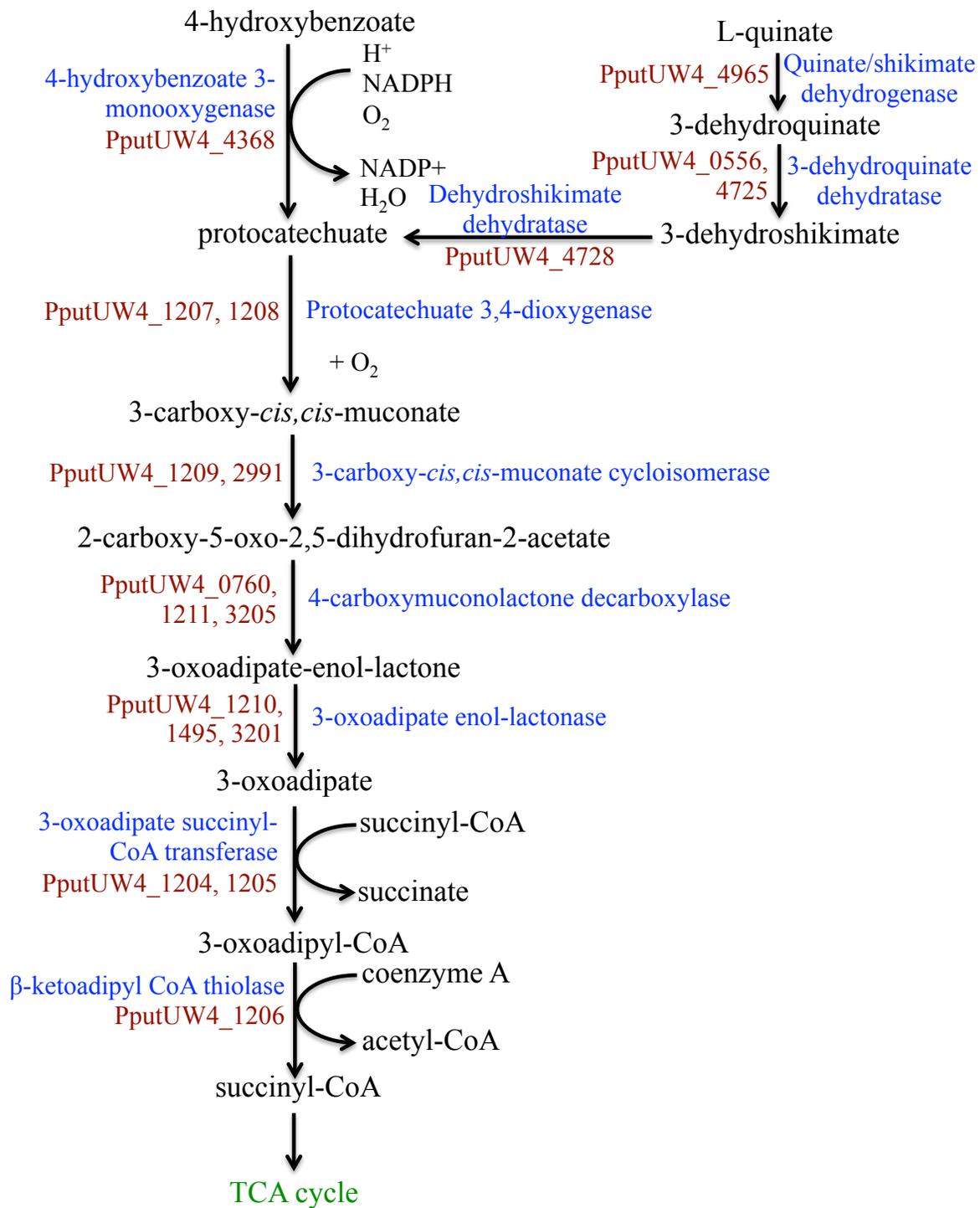


Fig. 4-14. 4-hydroxybenzoate and L-quininate degradation pathways in *P. putida* UW4.



encoding genes, *mhpA CDFE*, were found (Table 4-12), indicating the presence of an incomplete degradation pathway for 3-HPP.

#### 4.2.9 Heavy metal resistance

Based on the genome sequence of *P. putida* UW4, various heavy metal resistance determinants were identified (Table 4-13). It has been shown that UW4 can grow in rich medium containing 2 mM nickel at a growth rate of 0.24 generation/hour (Cheng et al. 2009b). As expected, putative nickel transporters were found in the UW genome. The genes encoding the transporters showed similarities to the Nik system (*nikABCDE*) that was originally identified in *E. coli*. A typical Nik system is comprised of five components including an ABC transporter periplasmic nickel binding protein (NikA), two nickel transporter cytoplasmic permeases (NikB and NikC), and two membrane-associated nickel import ATP-binding proteins (NikD and NikE) (Navarro et al. 1993). In addition, a nickel-responsive regulator NikR is present either downstream of NikE (in *E. coli*) or upstream of NikA (in *P. putida* KT2440 and *P. putida* W619) and acts as a repressor when the nickel concentrations are high. In the genome of UW4, the locus of the Nik system contains three copies of NikA (PputUW4\_0746, 0748, 0749) and a single copy of NikB (PputUW4\_0745), NikC (PputUW4\_0744), NikD (PputUW4\_0743), and NikE (PputUW4\_0742) (Fig. 4-15). At the amino acid level, pairwise sequence identities and similarities between NikA2 and NikA3 were 70% and 81%. Whereas NikA1 showed lower identities and similarities compared with NikA2 (47% identity and 63% similarity) and NikA3 (46% identity and 65% similarity) (Fig. 4-16). Furthermore, located between the first and the second copy of *nikA*, there is a putative gene encoding an outer membrane porin, which is transcribed in the same orientation as *nikA*. However, based on a sequence similarity search, NikR is not encoded in the UW4 genome.

Table 4-12. Genes involved in 3-hydroxyphenylpropionate degradation.

Gene	Product	W619	UW4
		ORF ID	ORF ID
		PputW619	PputUW4
<i>mhpT</i>	3-Hydroxyphenylpropionic acid transporter	1985	
<i>mhpE</i>	4-Hydroxy-2-oxovalerate aldolase	1984 2007	1665
<i>mhpF</i>	Acetaldehyde dehydrogenase	1983 2008	1664
<i>mhpD</i>	2-Hydroxypenta-2,4-dienoate hydratase	1982 2011	1663
<i>mhpC</i>	2-Hydroxy-6-keonona-2,4-dienedioic acid hydrolase	1981	1662
<i>mhpB</i>	2,3-Dihydroxyphenylpropionate 1,2-dioxygenase	1980	
<i>mhpA</i>	3-(3-Hydroxy-phenyl)propionate hydroxylase	1979	2149
<i>mhpR</i>	Mhp operon transcriptional activator	1978	

Sixteen genes that might be involved in the copper resistance of UW4 were identified (Table 4-13). These 16 genes are located at six regions on the chromosome, including three copper resistance systems in UW4, two individual sets of two-component transcriptional regulators, and one gene that might be involved in the bacterium survival in the presence of high bioavailable Cu(II). The first region contains three genes (PputUW4\_0580, 0581, 0583) and the amino acid sequences of all three genes showed high identities/similarities (89%/95%, 93%/96%, 88%/91%, respectively) compared to those of CueAR-CopP in *P. putida* PNL-MK25 (Fig. 4-17BCD), which has been experimentally confirmed to play an important role in copper homeostasis in PNL-MK25 (Adaikkalam and Swarup 2002). Thus, the *cueAR* mutant of PNL-MK25 showed a six-fold reduced tolerance to copper compared to the wild-type strain. CopP (PputUW4\_0583) is located upstream of *cueA* and is transcribed away from the *cueAR* operon (Fig. 4-17A). The *copP* gene encodes a heavy metal transport/detoxification protein containing a heavy-metal-associated (HMA) domain at its N terminus (Fig. 4-17D) and a *cueR* binding site, a *cop* box, in the promoter region (Fig. 4-17E).

The second region related to copper resistance contains four genes (PputUW4\_3489 – 3492) (Fig 4-19A). The homologs for the four genes were characterized in *P. putida* PNL-MK25 as well and they are designated *copABCD* (Adaikkalam and Swarup 2005). CopA is a multi-copper oxidase family protein, and a type-1 copper-binding motif (HCHLLYHM) is present at its carboxyl end (Ouzounis and Sander 1991). Similar to the CopA in PNL-MK25, CopA in UW4 has two HXXMXXM motifs at position 375 (Fig. 4-18B). CopB is a protein involved in copper binding. One MXXM and two HXXMXXM motifs were found at the N terminus of CopB in UW4 (Fig. 4-18C), whereas only one copy of MXXM motif was found in this protein in PNL-MK25 (Adaikkalam and Swarup 2005). The gene encoding CopC is similar to periplasmic proteins involved in copper resistance, and one MXXM metal binding motif is present in the

Table 4-13. Genes potentially involved in metal resistance of *P. putida* UW4.

Metal	Gene	PputUW4_	Product
Nickel	<i>nikA3</i>	0749	nickel ABC transporter periplasmic nickel-binding protein
	<i>nikA2</i>	0748	nickel ABC transporter periplasmic nickel-binding protein
	<i>nikA1</i>	0746	nickel ABC transporter periplasmic nickel-binding protein
	<i>nikB</i>	0745	nickel transporter permease
	<i>nikC</i>	0744	nickel transporter permease
	<i>nikD</i>	0743	nickel import ATP-binding protein
	<i>nikE</i>	0742	nickel import ATP-binding protein
Copper	<i>copP</i>	0583	heavy metal transport/detoxification protein
	<i>cueA</i>	0581	copper-translocating P-type ATPase
	<i>cueR</i>	0580	MerR family transcriptional regulator
	<i>copA</i>	3489	copper resistance protein
	<i>copB</i>	3490	copper resistance protein
	<i>copC</i>	3491	copper resistance protein
	<i>copD</i>	3492	copper resistance protein
	<i>cinQ</i>	3503	7-cyano-7-deazaguanine (pre-Q0) reductase
	<i>cinA</i>	3504	copper-containing azurin-like protein
	<i>cinR</i>	3505	two-component heavy metal response transcriptional regulator
	<i>cinS</i>	3506	heavy metal sensor histidine kinase
		2454	involved in survival in the presence of high bioavailable Cu(II)
		2050	heavy metal sensor signal transduction histidine kinase
	2051	two-component heavy metal response transcriptional regulator	
	4498	two-component heavy metal response transcriptional regulator	
	4499	heavy metal sensor signal transduction histidine kinase	
Cadmium	<i>cadA</i>	5170	cadmium translocating P-type ATPase
	<i>cadR</i>	5171	MerR family transcriptional regulator
	<i>cadA2</i>	5411	cadmium-translocating P-type ATPase
Zinc		1619	metallothionein
	<i>znuA</i>	0067	zinc ABC transporter periplasmic protein
	<i>znuB</i>	0064	zinc ABC transporter permease
	<i>znuC</i>	0065	zinc import ATP-binding protein
	<i>zur</i>	0066	ferric uptake regulator family protein
Molybdate	<i>modA</i>	2404	molybdate-binding periplasmic protein
	<i>modB</i>	2403	molybdate ABC transporter permease
	<i>modC</i>	2402	molybdate ABC transporter ATP-binding protein
		4989	ModE family transcriptional regulator

Cobalt	<i>cbtA</i>	2364	cobalt transporter subunit A
	<i>cbtB</i>	2365	cobalt transporter subunit B
Arsenate	<i>arsR</i>	2256	arsenical resistance operon repressor
	<i>arsB</i>	2255	arsenite efflux transporter
	<i>arsC1</i>	2254	arsenate reductase
	<i>arsH</i>	2253	NADPH-dependent FMN reductase
	<i>arsC2</i>	1085	arsenate reductase
	<i>arsC3</i>	4122	arsenate reductase
Chromate	<i>chrA</i>	3072	chromate transporter
	<i>oscA</i>	0155	hypothetical protein

Fig. 4-15. Genetic organization of nickel transporters located on the genome of *P. putida* UW4.

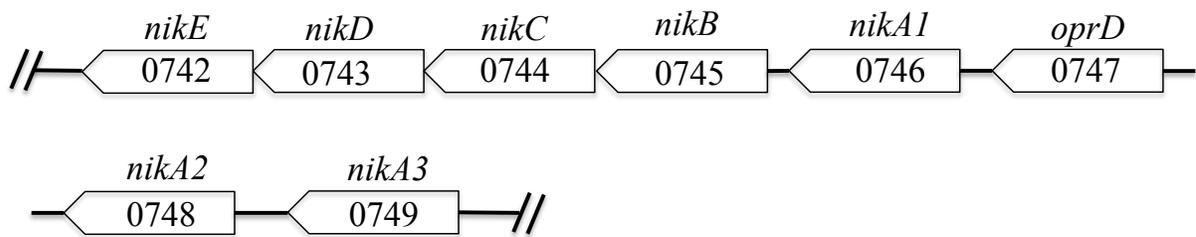


Fig. 4-16. Multiple sequence alignment of NikA of *P. putida* UW4.

```

1          10          20          30          40          50          60          70
UW4_nikA1 MKMLPLRAAATAAALLSVAVGVSAKP--LVVCTEASPEGFDMVOYTTAVTADAV AETIFNRLADFKPGT
UW4_nikA2 MRHTLVLSALLGTGLLAATSISOAANNSLVFCSEGSPAGFDTAQYTTATDNDAAEP LYNRLAEFEKCAT
UW4_nikA3 MLKQAVITPFLV GASLLASAPFAQAATN-LVFCSEGSPAGFDPGQYTTGTDFDAS AETMFNRLTQFERGGT

80          90          100         110         120         130         140
UW4_nikA1 EVIPALADSWDISEDGLSYTFHLRKGVKFHTTEYFKPTRDMNADDVVVWSEQRQLDPNHPWHKLSVSGFPY
UW4_nikA2 NVVPGLATSWDISEDGLK YTFHLREGVKFHTTKYFTPTRDFNADDVLFNRMRLDPQQPFRKAYPTEFPY
UW4_nikA3 AVIPGLATSWDVSDDGLTYTFHLREGVKFHTTPYFKPTREFNADDVLFNRMINKDDPFRKAYPTEFPY

150         160         170         180         190         200         210
UW4_nikA1 FESMGFKELIKSVEKVDDNTVKFETLRRR EAPFLADIAMAFSSISAEYADQLLKANKAGDLNKPVGTGP
UW4_nikA2 FNGMSLNKNILARVEKTGPLTVVFTLNSV DAAFIONIAMSFASILSAEYADKLLAEGKPSDINOKPVGTP
UW4_nikA3 FTDMDGMDTNIITKIDKVDDKTVKFETLKDV DAAFIQNMAMSFASISAEYAAQLLKEGKAADINOKPVGTP

220         230         240         250         260         270         280
UW4_nikA1 FVFQRYNKDAQVRFKANPDYFR-GKPPADALILIAITDNNVRLQKLANECHOVALYPKPDDIPSIKKDSN
UW4_nikA2 FVFKSYOKDSNIRYTGKHYWDP SRVKLDNLIFAINTDASVRVOKLKADECQITLHPRPADVPALKNDPA
UW4_nikA3 FVFKSYOKDSNIRYTGKDYWKPD DDKVIDNLIFAITTDPSVRIQKLLKNECOVTLFPRPADLKALKEDKD

290         300         310         320         330         340         350
UW4_nikA1 LKVDELNAMTVSYIAMNTHK-----YMSDVRVRKAIDIAFDKEAYVNALFGKGNATAGVNPYPDTL
UW4_nikA2 LQLIEKPGFNLGYIAYNVRHK-----PFDQLLEVROALDMAVNKQGIILNAVY-OGAGQLAVNAMPPTQ
UW4_nikA3 LKLPDQAGFNLGYIAYNVMPKVKGQTDANPLAELKVRQALDMAVNKQIIDS VY-OGAGQLAVNAMPPTQ

360         370         380         390         400         410         420
UW4_nikA1 LGYNHDLKNPARDL DKARALLKEAGVPEGTFTLFRNNGGPTNPNPMLG TOMMOADLAKVGIKIDIRVM
UW4_nikA2 WSYDDTIKDAAYNPEKARELLKAGVKEGTEITLWAMPVQRPYNPNAKLMAEMLOADWAKIGLKVKI VSY
UW4_nikA3 WSYDDTIKDAAYNPEKAKELLKEAGVKGTEIVLWAMPVQRPYNPNAKLMAEMLOSDW SKIGLKVKI QSY

430         440         450         460         470         480         490
UW4_nikA1 EWGEMLKRAKNGEHDMSAGWAGDNGDPDNFLTPMLSCEAAKNGENYARWCNDK FQALIDQAREK TDPAE
UW4_nikA2 EWGEYIKRRTKNCEHDISLICWTCGDNQDPDNWLGTLYSCDAI-GGNNYSMWCDPA YDKLIKQAKVV TDRDQ
UW4_nikA3 EWGEYIKRSKCGENQAMLICWSCDNGDPDNWLNVLFGCDSL-EGNNFSKWCDK KFDITVKQAKRTSDQAK

500         510         520         530         540         545
UW4_nikA1 RAALYEQAQVIFNODQPIWISMAHTRMFTAMRNNVEGYHISPLTTNNFATTQVK*
UW4_nikA2 RTLLYKQAQOYLKQOVPIPTVAHSTVNOPLSAKIEGFKVSPFGRNVFSGVSTIEK*
UW4_nikA3 RTELKQAQHV LKDAVPMPTPIAHSTVFOPMRANVQDFKISPFGLNSFYGVSVSK*

```

CopC of UW4 (Fig. 4-18D). The last gene in the operon is *copD*, which encodes a copper transport protein containing eight hydrophobic regions predicted by SPLIT4 server (Juretic et al. 2002) (Fig. 4-18EF). It was demonstrated that the *copABCD* system in PNL-MK25 showed minimal role compared with the same system in other microorganisms, including *P. syringae* pv. *syringae*, *E. coli* and *Xanthomonas campestris*. Furthermore, the *copABCD* system played a minor role compared with the *cueAR* system of the same strain. It was suggested that the reduced function of the *copABCD* operon is likely caused by these genes encoding fewer metal-binding domains compared to the homologs in other microorganisms (Adaikkalam and Swarup 2005). Based on the studies of the *copABCD* system in PNL-MK25, it is possible that this copper resistance system also plays a less important role in UW4 compared with other copper resistance systems because only two more metal binding motifs were observed in one of the four proteins (CopB).

The third copper resistance locus consists of four genes, *cinQARS* (PputUW4\_3503-3506) (Fig. 4-19A). The gene *cinQ* encodes a putative 7-cyano-7-deazaguanine (pre-Q<sub>0</sub>) reductase, and *cinA* encodes a putative copper-containing azurin-like protein. The gene products of the *cinRS* operon are a two-component heavy metal response transcriptional regulator (CinR) and a heavy metal sensor histidine kinase (CinS). It has been shown that *cinA* and *cinQ* of *P. putida* KT2440 were cotranscribed and induced by copper but not by nickel, ferrous iron, or zinc (Quaranta et al. 2007). However, deletion of *cinA* or disruption of *cinQ* in strain KT2440 did not increase the copper sensitivity significantly in disk assays, likely due to the redundancy in copper resistance genes in this strain (Quaranta et al. 2007). Furthermore, the two-component system *cinRS* activated transcription of *cinAQ* in the presence of copper, and two histidines, H37 and H147, in the periplasmic domain of CinS were found to be essential to induce the transcription of *cinAQ* (Quaranta et al. 2009). Amino acid sequence alignments of CinQARS between strains

Fig. 4-17. CueAR copper resistance system in *P. putida* UW4. (A) Genetic organization of *cueAR* system on the genome of UW4. (B) Amino acid sequence alignment of CueA between UW4 and PNL-MK25. (C) Amino acid sequence alignment of CueR between UW4 and PNL-MK25. (D) Amino acid sequence alignment of CopP between UW4 and PNL-MK25. Red bar shows the HMA domain (MSCGHC). (E) Intergenic region between the hypothetical protein and CopP. The predicted -10 and -35 regions of *copP* are in black boxes. Putative *cop* box is underlined in orange.



(C)

PNL-MK25_CueR	1	10	20	30
UW4_CueR	1	10	20	30
PNL-MK25_CueR	40	50	60	70
UW4_CueR	40	50	60	70
PNL-MK25_CueR	80	90	100	
UW4_CueR	80	90	100	
PNL-MK25_CueR	110	120	130	137
UW4_CueR	110	120	130	137

Sequence alignment for CueR protein domains. PNL-MK25\_CueR and UW4\_CueR are compared. Residues are numbered at the top. Differences are indicated by bolded letters in the UW4\_CueR sequence.

(D)

PNL-MK25_CopP	1	10	20	30
UW4_CopP	1	10	20	30
PNL-MK25_CopP	40	50	60	65
UW4_CopP	40	50	60	65

Sequence alignment for CopP protein domains. PNL-MK25\_CopP and UW4\_CopP are compared. Residues are numbered at the top. Differences are indicated by bolded letters in the UW4\_CopP sequence. A red box labeled "HMA domain" is positioned below the UW4\_CopP sequence between residues 10 and 20.

(E)

hypothetical

GGCTGAATCTCCTGAGTTCGAGTCATTGATGG

ACGCAGCTTGGCCGAGCGTAGCGCCGATTGCCAGCCCGGCAG

-35 -10

CAATCTTCA **GAACTTGACCTTGCCATGGTGGCAAGC**TCGAAAC

TGGACTCAACCCTCGGAAAAGGAGTCATCCC *copP*

Fig. 4-18. CopABCD copper resistance system in *P. putida* UW4. (A) Genetic organization of *copABCD* operon on the chromosome of UW4. (B) Amino acids sequence alignment of CopA between PNL-MK25 and UW4. Red bar indicates the type-1 copper-binding motif (HCHLLYHM) and brown bars show the metal binding motif (HXXMXXM). (C) Amino acids sequence alignment of CopB between PNL-MK25 and UW4. Brown bars indicate the metal binding motif (HXXMXXM/MXXM). (D) Amino acids sequence alignment of CopC between PNL-MK25 and UW4. Brown bar indicates the metal binding motif (MXXM). (E) Amino acid sequence alignment of CopD between PNL-MK25 and UW4. Red bars show the transmembrane helix domains of CopD of UW4. (F) Transmembrane analysis of CopD of UW4 by SPLIT 4.0 SERVER. Red line: transmembrane helix preference (THM index); blue line: beta preference (BET index); gray line: modified hydrophobic moment index (INDA index); violet boxes (below abscissa): predicted transmembrane helix position (DIG index).



(C)

PNL-MK25\_CopB  
UW4\_CopB

1 10 20 30 40 50 60  
MKEVHHDS~~THA~~CFPLLLAGT~~S~~A~~MA~~ADA~~M~~H~~GMDHS~~OMP~~GW~~IIR~~Q~~CKAWTT~~G~~MMOP  
MIRFTPVFFAFV~~T~~ST~~A~~MA~~A~~SDG~~M~~OGMDH~~S~~Q~~M~~DHG~~M~~O~~GMDH~~G~~M~~DH---S~~Q~~M~~Q~~SMDE~~G~~Q~~M~~OP

70 80 90 100 110 120  
AAPSESRTPI~~P~~V~~T~~DADRAAV~~F~~T~~S~~PG~~G~~H~~Q~~VHDS~~V~~I~~N~~S~~Y~~FLADKLEW~~Q~~N~~A~~DDGS~~A~~V~~A~~WDL~~S~~GW  
AAPSESRTPI~~P~~V~~T~~DADRAAV~~V~~VD~~P~~A~~G~~H~~H~~VHD~~N~~T~~L~~N~~S~~F~~F~~LADKLEW~~Q~~D~~A~~DDGS~~T~~L~~A~~WDL~~S~~GW

130 140 150 160 170 180  
IGGD~~T~~DRLWLRSEGER~~N~~NG~~K~~TEDA~~I~~QALWGHAI~~S~~PWWD~~V~~V~~G~~VRQDFKPGAP~~Q~~S~~W~~A~~A~~FGV~~Q~~  
VGGD~~T~~DRLWLRSEGER~~T~~NG~~V~~TEDA~~I~~QALWGHAI~~S~~PWWD~~V~~V~~S~~VRQDFKPGAP~~Q~~T~~W~~A~~A~~FGV~~Q~~

190 200 210 220 230 240  
GMALYNFEA~~E~~ATA~~F~~I~~G~~EGGQSAVRLEGDYDILL~~T~~NRLILOPT~~A~~ELN~~I~~YGKNDPORGIG~~S~~GLS  
GMALYNFEA~~Q~~ATA~~F~~I~~G~~EGGQSAVRLEGDYDILL~~T~~NRLILOPT~~V~~ELN~~A~~YGKNDPORGIG~~A~~GLS

250 260 270 280 290 300 303  
NTEAGLRLRYEIRREFAP~~Y~~I~~G~~V~~T~~WNR~~T~~Y~~G~~NTADYA~~R~~E~~E~~DE~~D~~RSEARLVLGVR~~L~~WF  
NTEAGLRLRYEIRREFAP~~Y~~I~~G~~V~~T~~WNR~~T~~Y~~G~~NTADYA~~N~~DE~~G~~DE~~D~~RSEARLVLGVR~~M~~WF

(D)

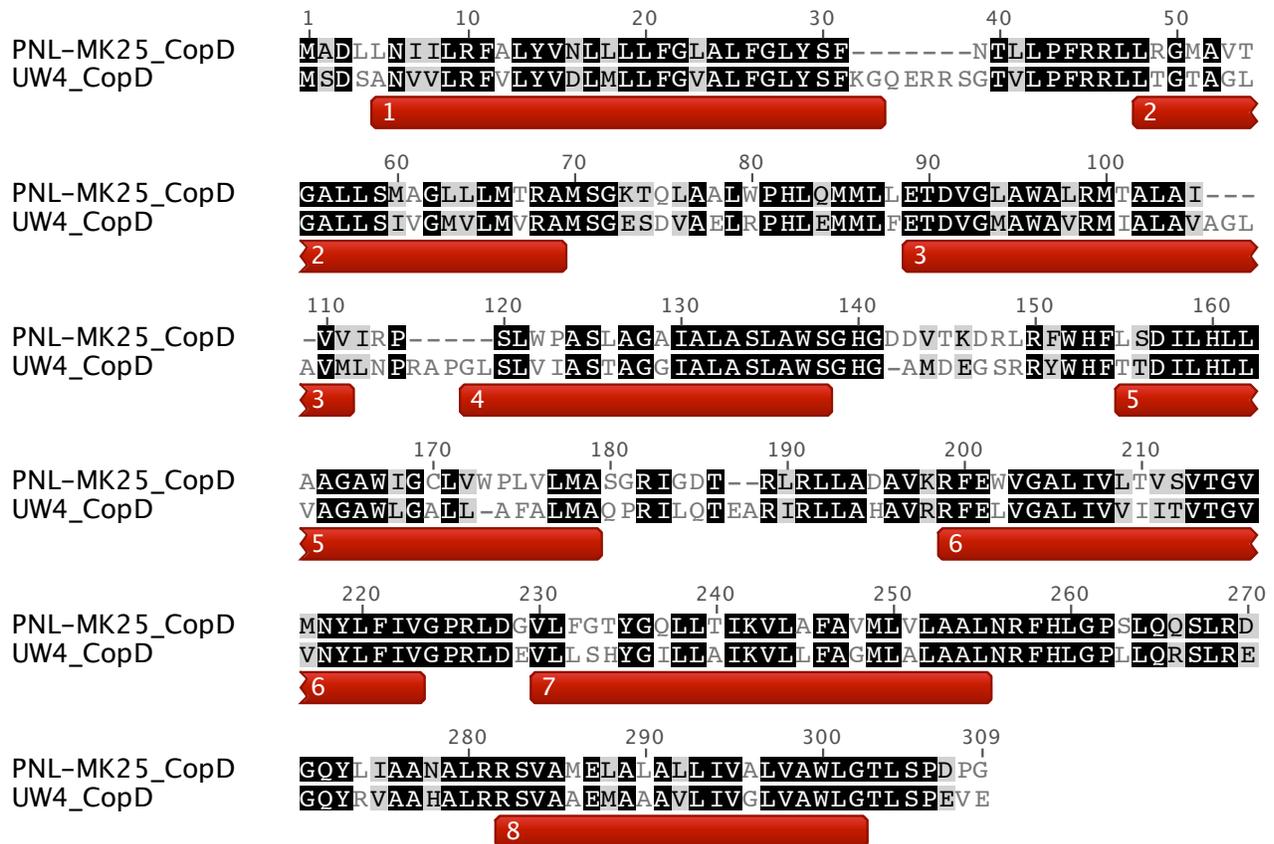
PNL-MK25\_CopC  
UW4\_CopC

1 10 20 30 40  
MR~~T~~L~~T~~L~~P~~L~~V~~FA~~S~~G~~L~~L~~L~~S~~A~~LAQAHPKLL~~S~~ST~~P~~A~~E~~GAEGA~~A~~PNRI~~E~~L  
MRL~~I~~K~~P~~TAV~~V~~IALL~~S~~G~~L~~L~~M~~S~~T~~LAQAHPKLL~~A~~ST~~P~~V~~E~~GAEGA~~P~~PD~~K~~I~~E~~L

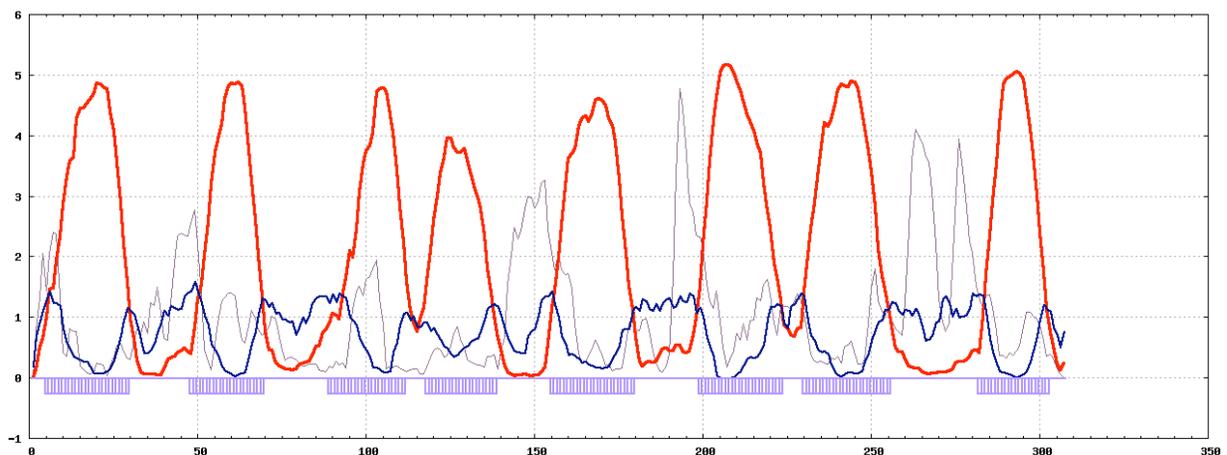
50 60 70 80 90  
HFSEN~~L~~V~~T~~OFSGAKL~~V~~M~~T~~EM~~G~~GRAHS~~P~~MP~~M~~KAKVS~~A~~S~~S~~DPKTMLIT~~P~~L  
HFSEN~~L~~V~~T~~OFSGAKL~~L~~M~~T~~EM~~A~~GMAHA~~P~~MP~~V~~KAKVS~~G~~G~~S~~DPKTMLIT~~P~~T

100 110 120 127  
APL~~P~~AGTY~~Q~~V~~Q~~WRAVS~~A~~D~~T~~HPITGN~~V~~TFK~~V~~K  
APL~~V~~AGTY~~K~~V~~E~~WRAVS~~S~~D~~T~~HPITGS~~V~~TFK~~V~~K

(E)



(F)



UW4 and KT2440 are shown in Fig. 4-19BCDE. The identities and similarities of each pair of sequence are 55%/72% (CinA), 78%/90% (CinQ), 87%/92% (CinR), and 66%/78% (CinS). The copper binding motifs (HXXMXXM/MXXM) are highlighted in the alignment of CinA, and the two conserved histidines required for the transcription of *cinAQ* are shown in the alignment of CinS.

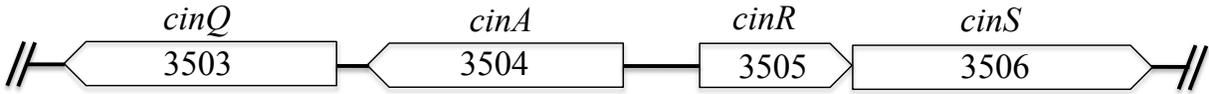
Sequence analysis of the two sets of two-component transcriptional regulators PputUW4\_2050-2051 and PputUW4\_4498-4499 showed similarities compared with CopRS in *P. putida* KT2440. The *copR* gene encodes a two-component heavy metal response transcriptional regulator and *copS* encodes a heavy metal sensor signal transduction histidine kinase. The alignment of CopS, PputUW4\_2050, and PputUW4\_4499 showed low similarities at the N terminus and high similarities at the C terminus (Fig. 4-20A), which is common in two-component systems in bacteria because the N terminus contains the variable sensing domain and the C terminus contains the conserved kinase domain. The alignment of CopR, PputUW4\_2051, and PputUW4\_4498 showed high similarities (Fig. 4-20B) and the pairwise identities/similarities are 60%/75% (2051 and CopR), 55%/74% (4498 and CopR), and 57%/76% (2051 and 4498). A proteomics study on the copper response in *P. putida* KT2440 has previously shown that CopR was up-regulated significantly in response to copper induction (Miller et al. 2009).

Lastly, a protein that might be involved in bacterial survival in the presence of high bioavailable Cu(II) was identified in the genome of UW4 (PputUW4\_2454). A sequence similarity search showed that it has high similarities compared with CopG1 and CopG2 in KT2440 (Fig. 4-21). Both *copG1* and *copG2* are located within copper resistance operons in KT2440. However, this is not observed in the UW4 genome.

Besides nickel and copper, *P. putida* UW4 may possess resistance to other heavy metals, such as cadmium, zinc, cobalt, molybdenum, chromate, and arsenate. Two genes, *cadA*

Fig. 4-19. CinQARS copper resistance system in *P. putida* UW4. (A) Genetic organization of *cinQARS* operon on the chromosome of UW4. (B) Amino acids sequence alignment of CinA between KT2440 and UW4. Brown bars show the metal binding motif (HXXMXXM/MXXM). (C) Amino acids sequence alignment of CinQ between KT2440 and UW4. (D) Amino acids sequence alignment of CinR between KT2440 and UW4. (E) Amino acid sequence alignment of CinS between KT2440 and UW4. Blue bars above the sequence of UW4 CinS show the conserved histidines essential for the transcription of *cinAQ*.

(A)



(B)

1 10 20 30 40 50

UW4\_CinA MFLR**KSL**LAQILCLLAL**SSPV**WADGGHAYD**FGOS**APAAKATRS**VEVVMG**DMSFA  
KT2440\_CinA MKHL**LFIT**ATLAL**GLP**AFAGQAQSF**AFGE**PAPAAKATRT**VNVVLK**DIAFE

60 70 80 90 100

UW4\_CinA **PKAIDIKAGETVRFVLVNKGRIL**HEFN**LGDA**AMHAR**HQOEMLO**MOQS-----  
KT2440\_CinA **PKSLOVKAGETVRFLLVNEGKLP**HEFN**LGDK**AMHAR**HQOEMVAM**OGKMFTEAM

110 120 130 140 150

UW4\_CinA **GMKAMDHGK**MAG**MDHGK**--**MDHG**--MK**HDDPN**SVL**VEPGKT**AELT**WTF**SKATGL  
KT2440\_CinA **NHEGMDHGK**--**MNHGOM****MDHG**DGHG**HGGN**TV**VMVMPG**OR**AELT**WTF**RQ**SAPI

160 170 181

UW4\_CinA **EFACNIPGHYQAGMVGKLT**VSQ  
KT2440\_CinA **EFACNV**PGHYQAGMVG**ALT**IE

(C)

1 10 20 30 40 50

UW4\_CinQ **MHPAAEH**SPLGKSSEYIATY**TPSL**LFPIPR**TAKW**AELGL**TAETLP**YK**GVDFW**  
KT2440\_CinQ **MHPAAEH**SPLGKSSEYIATY**SPEQL**FPPIPR**TAKW**AELGV**TAOTLP**WQ**GV**DYW

60 70 80 90 100

UW4\_CinQ **NCFEL**SWLLPSGKPVVA**IGFES**IPADSPN**IESK**SFKLYLNS**INOT**PE**T**DTA  
KT2440\_CinQ **NCFEL**SWLLPSGKPVVA**IGFEFA**IPADSPN**IESK**SFKLYLNS**INOT**VE**T**SLG

110 120 130 140 150

UW4\_CinQ **SLEATL**AKDLSAAAGK**PVGVRIR**SL**KDVER**EGVVAL**PGVC**IDE**LDI**SVSS**YE**  
KT2440\_CinQ **ALQVCL**EKDLSAAAGK**PVGKVRT**LA**EVEA**OGVVAL**PGQ**CID**ALD**VAIS**NYE**

160 170 180 190 200

UW4\_CinQ **HPRPE**LLRCDD**SRIVEE**SVH**SHLL**KSNC**PVTS**OPDWGS**VAVEYR**GA**ALD**HAS  
KT2440\_CinQ **QPOPE**LLRCN**PERVVEE**TLH**SHLL**KSNC**PVTG**OPDWGS**VVOYK**GR**ALD**HAS

210 220 230 240 250 260

UW4\_CinQ **LLEYIV**SFROHS**DFHEOC**VERI**F**LDL**ORLL**K**PEK**LTVYAR**YVRR**GGLD**INPY**  
KT2440\_CinQ **LLTYLI**SFROHA**DFHEOC**VERI**Y**LDL**KNLL**O**PEH**LTVYAR**YVRR**GGLD**INPY**

270 276

UW4\_CinQ **RSTEE**VOLP**NHRL**VRO  
KT2440\_CinQ **RSTGPI**SPD**NKRL**VRO

(D)

1 10 20 30 40 50  
UW4\_CinR **MKLLIVEDQPKTGEYLRQGLAEAGFNTELVADGSTGOQLALSGEYALLIILDVM**  
KT2440\_CinR **MKLLIVEDQARTGOYLSOGLSEAGFA TELATDGE TGOFLAATGDHDLILLDVM**

60 70 80 90 100  
UW4\_CinR **LPGRDGWQILOAVRGAGLDTPVLFELTARDAVEDRVHGLELGADDYLVKPFAFS**  
KT2440\_CinR **LPGRDGWQILOAVROAGLDTPVLFELTARDAVEDRVHGLELGADDYLVKPFAFS**

110 120 130 140 150  
UW4\_CinR **ELLARVRSLLRRGSATPQETSLOLADLRDLIRRRV ERSGORIDLTAKFEFALL**  
KT2440\_CinR **ELLARVRSLLRRGTSPSQDTILSLADLRDLIRRRAEFAGLRIDLTAKFEFALL**

160 170 180 190 200 210  
UW4\_CinR **EMLLRROGEVLPKSLIASOVWDMNFDSDTNVIEVAIRRLRLKIDDEF PNKLIH**  
KT2440\_CinR **EMLLRROGEVLPKSLIASOVWDMNFDSDTNVIEVAIRRLRLKIDDPH PNKLIH**

220 226  
UW4\_CinR **TVRGMGYVLEER**PA  
KT2440\_CinR **TVRGMGYVLEER**TE

(E)

1 10 20 30 40 50  
UW4\_CinS **MRRVSLG SRLALLFAACTAVVSLFAGVLF SRASEAHFVELDQOOLLGAKITGL**  
KT2440\_CinS **MRRVSLG SRLALLFAACTATVSLGAGL LFSRASEAHFVELDQOOLLDSRISL**

60 70 80 90 100  
UW4\_CinS **RRALHDFOSSDS-EVKLT---EELSRQADLSLRITGSDGQRWYDSSSTNLP LLL**  
KT2440\_CinS **RTQLAGVSTADELOARLPALRDELSHQADLALRISASNGATWFBSRSGLPH-A**

110 120 130 140 150  
UW4\_CinS **PROLGLSTVSHEGMDYRVLNAPLDV NKPDSPOLTLMLDITHHOHFLOMOHLI**  
KT2440\_CinS **AQATGLATLHAPGIDYRSLSVPLTQ GAMOSPRITLMLDITHHOHFLOGMORLI**

160 170 180 190 200 210  
UW4\_CinS **WLTVGLSALATALLGAWAARSGLRPLRRISAVASGVSAOSINMRLPEENMPPE**  
KT2440\_CinS **WLTVGLSALITALLGAWAARSGLRPLROMGVAASVSARSLITRLPVAQMPPE**

220 230 240 250 260  
UW4\_CinS **LAEMAHNFNAMLGRLLDSFORLSAFSADIAHELRTPLSNLLTHTQVTLTRPRP**  
KT2440\_CinS **LAELASSMNAMLORLDDAFORLSAFSADIAHELRTPLSNLLTHTQVTLTRPRS**

270 280 290 300 310  
UW4\_CinS **IEDYREALHSNLEELQWMAQLVNDMLYLAKADHGLLIPKREPLBELGQEADLLT**  
KT2440\_CinS **LEEDYREALHGNLEELQWMAQMINDMLEFLAKADHGLLVPGDAPLALHDEVDAIT**

320 330 340 350 360 370  
UW4\_CinS **EFFAPLAEDAQVTLAREGHGCLLEGDRSMLRRALS NLLDNALRFTPV DGEVRVN**  
KT2440\_CinS **EYYAPLAEDSDVOMLREGEAVLHGDOHMLRRALS NLLDNAMRFTPVACGOIKVT**

380 390 400 410 420  
UW4\_CinS **IVEQANDRLRSVENTCSAIDAQLPRLFDRFYRA DPARREGSSEHAGLGLAIT**  
KT2440\_CinS **L---GPGPTINVANTGLAIDPAALPRLFDRFYRV DPARREGSSEHAGLGLAIT**

430 440 450 454  
UW4\_CinS **QSTIRAHGGQIRCESENGWTRFVIDLPKGD**  
KT2440\_CinS **RSTIVQAHGGCIRAECEGWTRFVIEFTQDR**

(PputUW4\_5171) and *cadR* (PputUW4\_5172), involved in cadmium resistance were identified. The gene *cadA* is known to encode a cadmium-transporting ATPase, and CadR is a MerR family response regulator responsible for cadmium resistance. It has been reported that *cadAR* played a major role in cadmium resistance and also contributed to zinc resistance in *P. putida* 06909 (Lee et al. 2001). In addition, the transcription of *cadR* is induced by cadmium and repressed by high levels of CadR in the absence of cadmium (Lee et al. 2001). Sequence alignments of CadAR from UW4 and 06909 are shown in Fig. 4-22AB. In the genome of UW4, another *cadA* gene (PputUW4\_5412) was identified based on a sequence similarity search. However, when comparing the amino acid sequence of the PputUW4\_5412 to CadA from *P. putida* 06909, the identities and similarities are only 36% and 52%, respectively. Furthermore, PputUW4\_5412 lacks the HMA domain at N terminus. Therefore, the function of this CadA needs to be confirmed experimentally in UW4.

Zinc is an essential trace element that acts as a cofactor for many enzymes. However, high concentrations of zinc are toxic to the cell. Bacteria employ different strategies to control zinc levels, including storage by metallothionein and export from the cell by ABC transporter systems (Hantke 2005). In the genome of UW4, a putative metallothionein was identified (PputUW4\_1619). It has previously been observed that the metallothionein from *P. putida* KT2440 bound multiple  $Zn^{2+}$  molecules via cysteine side chains (Blindauer et al. 2002). A sequence alignment of metallothionein from *P. putida* UW4, *P. putida* KT2440, *P. fluorescens* Pf0-1, and *P. fluorescens* SBW25 shows that the protein contains 10 conserved cysteine residues (Fig. 4-23). In addition, a common zinc transporter system is also present in UW4. The system consists of three genes *znuABC* (PputUW4\_0067, 0064, 0065) and one transcriptional repressor *zur* (PputUW4\_0066). The gene products of *znuABC* are a periplasmic binding protein, a membrane permease, and an ATPase, respectively. The gene *zur* is located between *znuA* and

*znuC*, and is transcribed in the same orientation as *znuBC*, but in the opposite direction from *znuA* (Fig. 4-24A). An amino acid sequence comparison of ZnuA from UW4 and *E. coli* K12 indicated 39% identities and 59% similarities, with the ZnuA from UW4 having the conserved His and Asp residues that have been shown by crystallography to bind to zinc (Hantke 2005) (Fig. 4-24B).

The molybdate transport system in *P. putida* UW4 is comprised of three genes, *modABC* (PputUW4\_2404, 2403, 2402). ModA is a periplasmic binding protein; ModB is an integral membrane protein; and ModC is an ATPase. In *E. coli*, the *modABC* expression is tightly controlled by a repressor protein, ModE, and the gene is located upstream of *modABC* operon (Self et al. 2001). In the genome of UW4, a homolog of ModE is not present upstream of the molybdate transport system. However, a ModE family transcriptional regulator (PputUW4\_4990) is found elsewhere on the chromosome. A sequence comparison of ModE between *E. coli* and UW4 is shown in Fig. 4-25, and the two sequences have 39% identities and 57% similarities.

One cobalt transporter locus comprising two genes, *cbtA* (PputUW4\_2364) and *cbtB* (PputUW4\_2365) was identified in the genome of UW4. This transport system has been found in various bacteria and it is related to vitamin B<sub>12</sub> biosynthesis. Homologs of CbtA usually have five transmembrane segments, and the gene is always co-localized with *cbtB*, which encodes one transmembrane segment and a histidine-rich C terminus likely to be a metal-binding site (Rodionov et al. 2003).

Arsenic ions are very toxic to most microbes and are common environmental pollutants. Arsenic resistance determinants were found in three regions on the chromosome of UW4, including an operon *arsRBCH* (PputUW4\_2256-2253), and two individual *arsC* genes (PputUW4\_1085 and PputUW4\_4122). The gene product of *arsC* is an arsenate reductase that catalyzes the reduction of arsenate to arsenite. ArsB is an arsenite efflux transporter, which can

Fig. 4-20. Multiple sequence alignments of CopRS, PputUW4\_2050-2051, and PputUW4\_4498-4499. (A) Amino acids sequence alignment of CopS, PputUW4\_2050 and 4499. (B) Amino acids sequence alignment of CopR, PputUW4\_2051 and 4498.

(A)

```
1      10      20      30      40      50
UW4_4499  M S S N S T A I R L S G M F T L V A L L V F L L I G W A L Y Q O V D K G L G L L F E A E L D
UW4_2050  M N A M T P L R S A S R L S H S L T L R L A L V F A L L A F V S L A L L G V A L Y R D L E R E L I H R D D I A L I
KT2440_CopS M I L T R S S L V K R L T L M I M F A V I A V L V V A G I S F N M L S Q H H F R M L D E Q A L S

60      70      80      90      100     110
UW4_4499  A R Y S V L E S T V G R Y G T P E H W V K I N N K L - - N I L S E E D K R I S F W I I S G D P H Y - - E Y G N L I T
UW4_2050  T R I D Q L R T F L N D S N T - L D L I K A K P A L F Q N M M G N R E A L L T I G T P G Q P P L L V V N P G N L I T
KT2440_CopS E K L E S T R H I L S I Q A P G T G M F E L R P Q L R A L I G A H O D L T A E I L T S E G E V L F - - - - S D L I -

120     130     140     150     160     170
UW4_4499  - P L V R A F A Q G - P L G M R D M - H L P D - H P Y P L K V L V S E F P A K D Q R P P L R F M I A I D T E T F F
UW4_2050  M P V L I T P V A M D H R L T L D D V Q H L P N V N G V P F S A L A A T I D S G D L G S L O V V T G R I M T E R T A
KT2440_CopS - K A V Q I P D K Y K R A D K E E M W E W Q D - E S H N F R G I T T D V E I Q G Q P A P G T V M L M I D I T S H A

180     190     200     210     220
UW4_4499  H T Q H L L V A L I S L A I V G V L L A S A L G Y W V A R I G L K P L I K L S Q E A R R L A P P L R S G R L R L
UW4_2050  V L A D Y R L N V Y L F A T L A A I L L A L S - C Y V L V H R G L L P V R R L A R H T Q S I D V G N L N E R L I D S
KT2440_CopS H F F E T L Q R W F A I G L V I S A L V S A G I G W L V A K S G L R P V E O I T K V A T S I S A R S L Q E R I P L

230     240     250     260     270     280
UW4_4499  S P L E P E L D Q F V S S F N S T L E R V E Q A Y S R L E S F N A D V A H E L R S P L T N L I G Q T O V A L T R G
UW4_2050  Q G A P L E L L P M I D A F N A M L D R L D K G F V Q L G Q V S T D M A H E L R T P T I N N L L G E T O V A L Q O N
KT2440_CopS E P V P L E L Q T L I L S F N G M L A R L E D A F V R L S N F S A D I A H E L R T P V S N L L T H T E V V L T R K

290     300     310     320     330     340
UW4_4499  R S A E H Y F E V L Q S N L E E L E R L R S I I N D M L F L A S A D Q G S K A T K L T S T S L A D E V A T T L E Y
UW4_2050  R S T D A Y Q O L L A S N V E E L E R L A R M L D N M L F L A R T D P R S A L R Q R O E L D A A D E M O R M A D Y
KT2440_CopS R D V D A Y E E N L Y S N L E D L K R M S R M I D D M L F L A K S D N G L I I P E Q V D I Q L E H D L V S K L F E Y

350     360     370     380     390
UW4_4499  L D F I L E D A Q V O V E V S G D A O V S T E I A H L R R A L I N L L N N A V O H T A P G O V T E F V H I E A A E H
UW4_2050  F E G P A T D V G I C I K A H G S G L I W A E P M L L R R A L A N L C A N A I K Y G A P H T E L C I H A N P T A D
KT2440_CopS Y Q L L A D D R G I R L T L Q G K G V I S G D R L M T D R A V S N I L S N A V R Y T P D G S G T S V E I Q Q A A D

400     410     420     430     440     450
UW4_4499  Q V S T G V A N P G S P I A H E H L P R L F E R F Y R V D A S R S N S G H - N H G L G L A I V K A I A L M H G C D
UW4_2050  G I O V O V R N H G Q T I P A Q H L S R L F E R F Y R V D E S R E R S A Q S N - G L G L S I V A T I M Q L H N G A
KT2440_CopS K V L T I K N G C A T I D P O H I N K I F D R F Y R A D P A R R E G G P S N A G L G L A I T R S I V K A H D G K

460     470     480     485
UW4_4499  V F V R S D H C M N T F G I H L P V
UW4_2050  C H V S S A D G I T C F E L F F P A R Q Q H E
KT2440_CopS V W C T S S E G V T A F N F I F P A A H R T A H I R T R H
```

(B)

```
1      10      20      30      40      50
UW4_4498  M R V L I I E D E K T A D Y L H R G L T E G Y T V D L A O D G V E G L H L A L E S D Y A V I V L D V M L P G
UW4_2051  M N I L V V E D E P K A G N Y L N G L Q E L G Y T V N L A R D G V D G L H L A L E H D F D V I V L D V M M P K
KT2440_CopR M K L L V A E D E P K T G A Y L Q O G L A E A G F T V D R V L T G T D A L Q H A L S E S Y D L L I L D V M M P G

60      70      80      90      100     110
UW4_4498  L D G F G V L R A L R A R - K O T P V I M L T A R E R V E D R I R G L R D G A D D Y L G K P F S F L E L V A R L
UW4_2051  M D G W E V L R R L R K E - A D T P V L E L T A R D D I A D R I K G L E L G A D D Y L I K P F S F A E L V A R L
KT2440_CopR L E G W E V L R M V R A A G K D V P V L E L T A R D C V D D R V K G L E L G A D D Y L I K P F A F S E L L A R V

120     130     140     150     160
UW4_4498  Q A L T R R S - G G H E P V Q V S I A D L W I D L I S R K A T R C G A R L D L T A K E F S L L S V L A R R O G E
UW4_2051  R T L T R R G - P T R E D E H L F V D D L Q I D V L K R R V S R A G N R I T L T N K E F A L L H L F A T H O G E
KT2440_CopR R T L L R R G T S T A A Q T T M K M A D I E V D L L K R R A T R S G K R I D L T A K E F S L L E L L M R R R G E

170     180     190     200     210     220
UW4_4498  I L S K T A I A E M V W D I N F D S D A N V V E V A I K R L R A K L D G P F E E K L L H T I R G M G Y V L E S R
UW4_2051  V L S R S M I A S R V W D M N F D S D T N V V D V A V R R L R L K I D D P F O L K L I H S V R G I G Y R F D T O
KT2440_CopR V L P K S L I A S O V W D M N F D S D T N V I E V A I R R L R A K I D D D F A P K L I H T A R G M G Y M M D A P

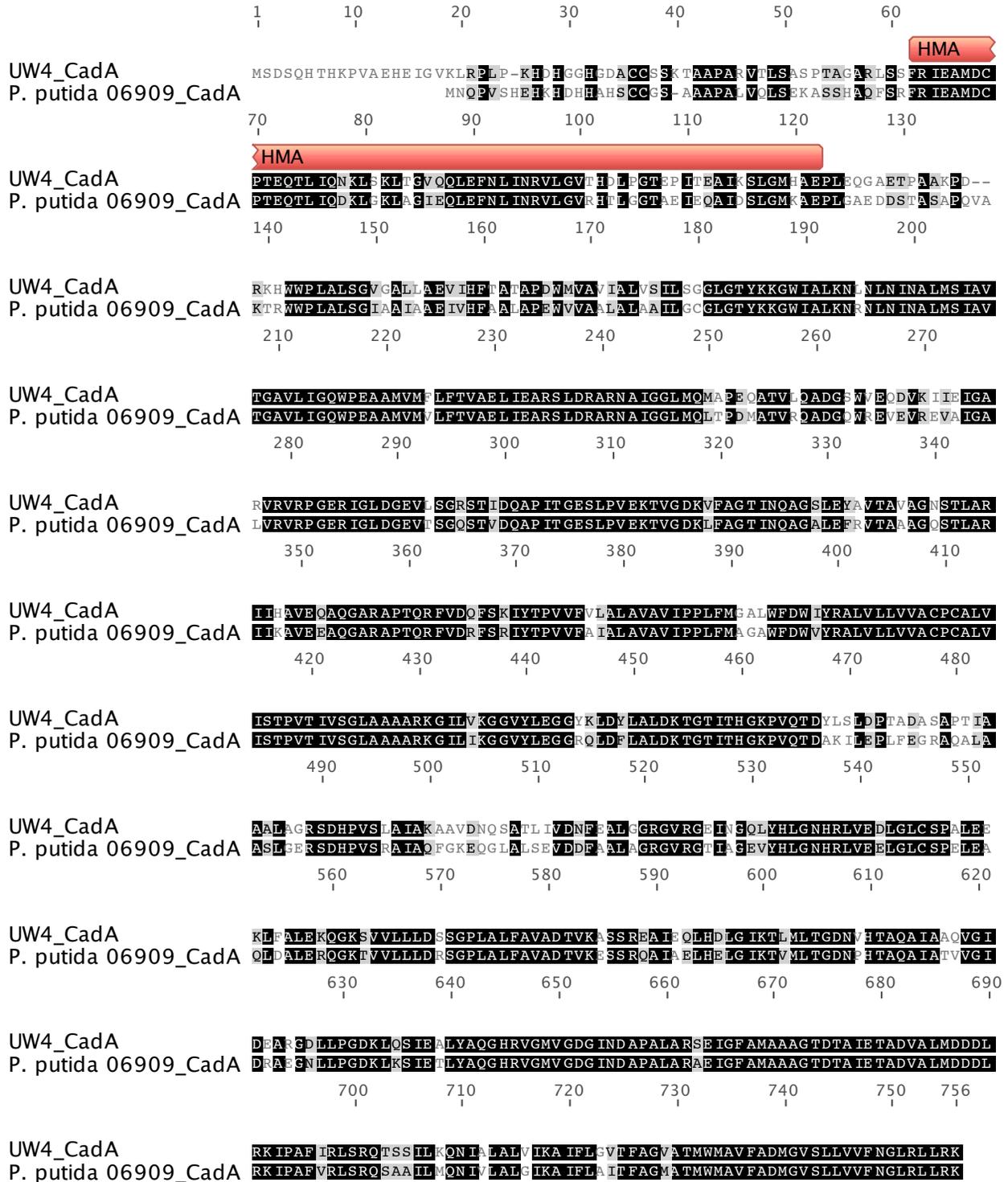
227
UW4_4498  G V O
UW4_2051  P
KT2440_CopR  E
```

Fig. 4-21. Amino acid sequence alignment of PputUW4\_2454 and CopG of KT2440.

	1	10	20	30	40	50
KT2440_2203_CopG		M L R K H L L T L S L L A V T G L A Q A A E - - T	I D V Y R D P N C G C C K A W I S H L R D N G			
KT2440_5377_CopG	M K A N S R F L R - - L A A F A A F I A L P P V Y A A V P T T	I D V H R D A N C G C C K K W I A H L E Q N G				
UW4_2454	M R N P L R - - L I A L S A L F M S S L V O A A D L I P	I E V H R D A N C G C C K K W I S H L E A N G				
	60	70	80	90	100	
KT2440_2203_CopG	F T V N D H V E P N M S A V K Q R L G V A P R L A S C H T G V I G	G K F V E G H V P A E Q V R L L A K R N D				
KT2440_5377_CopG	F K V I D H V E T D M S A V K Q S L G V A P R L A S C H T A V I D	G K F V E G H V P A E Q V R E L A N R D D				
UW4_2454	F K V E D H V E A N M S E F K Q Q H G V P P R L A S C H T A L I N	G K F V E G H V P A D Q V L A L S K R D D				
	110	120	130	140	150	153
KT2440_2203_CopG	L K G L A V P G M P M G S P G M E M G D H K D A Y Q V I G V T O D G O D T V V A N Y					
KT2440_5377_CopG	L I G V A V P G M P A G S P G M E V D G V S H A Y Q V I G L N K S G R D E V V A E Y P G H					
UW4_2454	L L G V A A P G M P M G S P G M E M D G M S D A Y Q V I G L K K D G T D L V V A D Y P A H					

Fig. 4-22. Amino acid sequence alignments of CadAR from *P. putida* UW4 and *P. putida* 06909. (A) Alignment of CadA (identities/similarities: 77%/86%). Heavy-metal-associated (HMA) domain is shown in red. Metal binding sites are shown in blue. (B) Alignment of CadR (identities/similarities: 77%/84%). Histidine rich region at C terminus is highlighted in green.

(A)



(B)

	1	10	20	30	40	50				
P. putida 06909_CadR	<b>MKIGELAKA</b>	<b>TDCAVET</b>	<b>IRYYERE</b>	<b>OLLPEPARSDGNYR</b>	<b>LYTOAHV</b>	<b>ERLTFIRNC</b>				
UW4_CadR	<b>MKIGELAKL</b>	<b>TDCAVET</b>	<b>IRYYEREN</b>	<b>LLPEPARSDGNYR</b>	<b>VYTOTHA</b>	<b>ERLTFIRNC</b>				
	60	70	80	90	100					
P. putida 06909_CadR	<b>RTLDMTLD</b>	<b>EIRSLRL</b>	<b>RDSPD</b>	<b>DSG</b>	<b>SVNALIDEHT</b>	<b>EHVQARIDGLVALQ</b>	<b>EQLV</b>			
UW4_CadR	<b>RTLDMTLE</b>	<b>EIRSLLA</b>	<b>FRDSPD</b>	<b>DC</b>	<b>SVNALIDEHT</b>	<b>EHVKARIDGLLALQA</b>	<b>QLL</b>			
	110	120	130	140	149					
P. putida 06909_CadR	<b>ELRRC</b>	<b>N</b>	<b>AOGA</b>	<b>--E</b>	<b>CA</b>	<b>ILQOLE</b>	<b>TNGAV</b>	<b>SVPE</b>	<b>EHSHVGRSHGH</b>	
UW4_CadR	<b>DLRRC</b>	<b>G</b>	<b>ERPG</b>	<b>ID</b>	<b>QC</b>	<b>ILQORLE</b>	<b>VSGV</b>	<b>VAT</b>	<b>EA</b>	<b>EHSHVGRSHGH</b>



extrude arsenite out of the cell. ArsR functions as an arsenical resistance operon repressor that responds to arsenate (Cánovas et al. 2003). ArsH is a NADPH-dependent FMN reductase and its role in arsenic resistance is not clear (Ye et al. 2007). The other two individual ArsC proteins showed far less similarity compared with the ArsC in the operon, implying that they belong to different families of arsenate reductase.

The mechanism used by various bacteria to extrude toxic chromate is through a chromate transporter, ChrA (Nies et al. 1990; Cervantes et al. 1990; Alvarez et al. 1999; Pimentel et al. 2002; He et al. 2010). A gene encoding ChrA was identified in the UW4 genome (PputUW4\_3072). The deduced amino acids sequence showed 31.2% identities compared with a previously characterized ChrA in *P. aeruginosa* (accession no. AAA88432) (Fig. 4-25A). However, the protein sequence of ChrA in UW4 showed much higher identities (75.8%) when compared with the gene from strain KT2440 (Fig. 4-25B). In another study, a small protein, OscA, was found to be responsible for chromate resistance in *Pseudomonas corrugata* 28 (Viti et al. 2009). The gene, *oscA*, was identified by characterizing a chromate-sensitive mutant created by transposon mutagenesis. In the genome of UW4, an *oscA* homolog (PputUW4\_0155) was found upstream of a sulfate-binding protein gene (*cysP*), which has been demonstrated to form a transcriptional unit with *oscA* (Viti et al. 2009). The genetic organization of the *oscA* region (Fig. 4-26) is exactly the same as this region in *P. corrugata* 28, indicating that the *oscA* gene from UW4 may also play an important role in chromate resistance.

#### **4.3 *P. putida* UW4 Central Metabolic Pathways and Protein Secretion Systems**

A schematic summary of the metabolic strategies in *P. putida* UW4 is shown in Fig. 4-27. The genome of *P. putida* UW4 contains a complete central carbon metabolism pathway including

Fig. 4-23. Amino acid sequence alignment of metallothionein in *P. putida* KT2440, *P. fluorescens* Pf0-1, *P. fluorescens* SBW25, and *P. putida* UW4. Pink bars indicate the conserved cysteine residues.

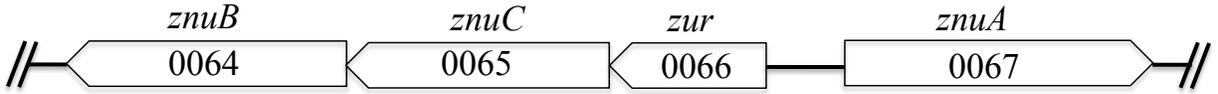
	1		10		20		30		40																																			
KT2440	<b>M</b>	<b>N</b>	<b>D</b>	<b>Q</b>	<b>R</b>	-	<b>C</b>	<b>A</b>	<b>C</b>	<b>T</b>	<b>H</b>	<b>C</b>	<b>S</b>	<b>C</b>	<b>T</b>	<b>V</b>	-	<b>D</b>	<b>A</b>	<b>N</b>	<b>A</b>	<b>L</b>	<b>Q</b>	<b>R</b>	<b>D</b>	<b>G</b>	<b>K</b>	<b>A</b>	<b>Y</b>	<b>C</b>	<b>C</b>	<b>E</b>	<b>A</b>	<b>C</b>	<b>A</b>	<b>S</b>	<b>G</b>	<b>H</b>	<b>R</b>	<b>K</b>				
Pf0-1	<b>M</b>	<b>A</b>	<b>D</b>	<b>R</b>	<b>K</b>	-	<b>C</b>	<b>D</b>	<b>C</b>	<b>P</b>	<b>G</b>	<b>C</b>	<b>R</b>	<b>C</b>	<b>T</b>	<b>I</b>	<b>K</b>	<b>E</b>	<b>G</b>	<b>E</b>	<b>H</b>	<b>S</b>	<b>Y</b>	<b>V</b>	<b>A</b>	<b>H</b>	<b>G</b>	<b>K</b>	<b>H</b>	<b>Y</b>	<b>C</b>	<b>C</b>	<b>E</b>	<b>A</b>	<b>C</b>	<b>A</b>	<b>H</b>	<b>H</b>	<b>K</b>	<b>S</b>				
SBW25	<b>M</b>	<b>P</b>	<b>D</b>	<b>E</b>	<b>T</b>	-	<b>C</b>	<b>A</b>	<b>C</b>	<b>P</b>	<b>H</b>	<b>C</b>	<b>K</b>	<b>C</b>	<b>V</b>	<b>L</b>	-	<b>G</b>	<b>V</b>	<b>D</b>	<b>A</b>	<b>V</b>	<b>M</b>	<b>K</b>	<b>D</b>	<b>G</b>	<b>K</b>	<b>G</b>	<b>Y</b>	<b>C</b>	<b>C</b>	<b>Q</b>	<b>G</b>	<b>C</b>	<b>A</b>	<b>E</b>	<b>H</b>	<b>H</b>	<b>A</b>	<b>H</b>				
UW4	<b>M</b>	<b>I</b>	<b>D</b>	<b>R</b>	<b>E</b>	<b>S</b>	<b>I</b>	<b>C</b>	<b>D</b>	<b>C</b>	<b>P</b>	<b>K</b>	<b>C</b>	<b>S</b>	<b>C</b>	<b>K</b>	<b>L</b>	-	<b>G</b>	<b>E</b>	<b>H</b>	<b>P</b>	<b>I</b>	<b>A</b>	<b>R</b>	<b>H</b>	<b>G</b>	<b>K</b>	<b>H</b>	<b>Y</b>	<b>C</b>	<b>C</b>	<b>E</b>	<b>A</b>	<b>C</b>	<b>A</b>	<b>K</b>	<b>H</b>	<b>H</b>	<b>E</b>	<b>H</b>			

			50		60		70		79																												
KT2440	<b>G</b>	<b>E</b>	<b>P</b>	<b>C</b>	<b>R</b>	-	<b>M</b>	<b>Q</b>	<b>D</b>	<b>C</b>	<b>H</b>	<b>C</b>	<b>G</b>	<b>E</b>	<b>K</b>	<b>P</b>	<b>G</b>	<b>E</b>	<b>S</b>	<b>A</b>	<b>V</b>	<b>D</b>	<b>N</b>	<b>A</b>	<b>L</b>	<b>D</b>	<b>E</b>	<b>T</b>	<b>F</b>	<b>P</b>	<b>A</b>	<b>S</b>	<b>D</b>	<b>P</b>	<b>I</b>	<b>S</b>	<b>P</b>
Pf0-1	<b>G</b>	<b>E</b>	<b>E</b>	<b>C</b>	<b>S</b>	-	<b>G</b>	<b>K</b>	<b>G</b>	<b>C</b>	<b>Q</b>	<b>C</b>	<b>A</b>	<b>H</b>	<b>P</b>	<b>E</b>																					
SBW25	<b>G</b>	<b>E</b>	<b>P</b>	<b>C</b>	<b>S</b>	<b>A</b>	<b>A</b>	<b>T</b>	<b>G</b>	<b>C</b>	<b>E</b>	<b>C</b>	<b>A</b>	<b>K</b>	<b>S</b>	<b>A</b>	<b>K</b>	<b>G</b>																			
UW4	<b>G</b>	<b>E</b>	<b>E</b>	<b>C</b>	<b>A</b>	-	<b>H</b>	<b>K</b>	<b>G</b>	<b>C</b>	<b>K</b>	<b>C</b>	<b>S</b>	<b>H</b>	<b>G</b>																						

Fig. 4-24. (A) Genetic organization of zinc transporter system in UW4. (B) Amino acid sequence alignment of ZnuA between UW4 and *E. coli* K12. Pink bars indicate the conserved His or Asp residues that bind zinc. Residues underlined by green indicate the His-Asp loop usually found in zinc transporters.

(A)



(B)

Sequence alignment of UW4\_ZnuA and E. coli K12\_ZnuA. The alignment is shown in blocks of 10 amino acids, with positions 1-50, 60-110, 120-170, 180-220, 230-280, and 290-313 indicated. Conserved residues are highlighted in black. A green bar highlights a region in E. coli K12\_ZnuA between positions 125 and 150. Pink ovals highlight specific residues in E. coli K12\_ZnuA at positions 105, 145, 210, and 280.

```
1      10     20     30     40     50
UW4_ZnuA  MVIVSRLFSIFVAFFTSFLLIGPAQAEVAVLASVKPLQLTAAAVQDGVAIPEVLLPP
E. coli K12_ZnuA MLHKKTLLFAALSAALWGGATQAADAAVVASLKPVGFIASAIADGVTETEVLLPP

60     70     80     90     100    110
UW4_ZnuA  GASPEHNYALRPSDVRKVSVDLLYWICPDMEGFLPRV---LNGRTLPSVA-VODLPG
E. coli K12_ZnuA GASPEHDYSLRPSDVKRLQNALDLVWVWVGPMEAFMQKPVSKLPGAKQVTIAOLEDKP

120    130    140    150    160    170
UW4_ZnuA  MKLRRFAEDSHSHAEEADDDHDDHDFGSLDAHLWLSFVNARVIATKMAADLSAVDIPD
E. coli K12_ZnuA LLMKSIHGDDDDH-DHAEKSDDDHHDGDFNMHLWLSPEIARATAVAIHGKIVELMPQ

180    190    200    210    220
UW4_ZnuA  NAVRYQSNLKAFDERLDALDARLKKRLADVOGKPYFVFHEAFDYFEDAYGLKHVGVF
E. coli K12_ZnuA SRAKLDANLKDFEAQLASTETQVGNELAPLKGKGYFVFHDAYGYFEKQFGLTPLGHF

230    240    250    260    270    280
UW4_ZnuA  SVAAEVOPGAQHVAAMRRLQEVGKTCVFSEPPPLRPLAETLVAGLPVKLAELDALG
E. coli K12_ZnuA TVNPEIOPGAQRLHEIRLQLVQKATCVFAEPQFRPAVVESVARGT SVRMGTLDPLG

290    300    310 313
UW4_ZnuA  GYTPATAQGYEQVLEKLGNDLAGCLESL
E. coli K12_ZnuA TNIKLGKTSYSEFLSOLANQYASCLKGD
```

Fig. 4-25. Sequence comparison of ChrA between (A) *P. putida* UW4 and *P. aeruginosa* (B) *P. putida* UW4 and *P. putida* KT2440.

(A)

```
1          10          20          30          40          50
P. aeruginosa ChrA MSVANEESYRPSKATDATT EAVPPP--MSY POLFARFLKFCGLLAWGPPVAQID
UW4_ChrA           MSSTPSPITQOAVPQPRRTISLREAF LFWLKLGFI SFGGPAGQIS

60          70          80          90          100
P. aeruginosa ChrA MLRRELVD EERWIS SKRFNKL LAVMQV L PGPEA HEICVHLG IRAKGRL GGVLA
UW4_ChrA           IMHQELV EERRWIS ERFLL HALNY CMLL PGPEA QOLATY IGWLMHRTW GGVIA

110         120         130         140         150
P. aeruginosa ChrA GLGFMLPG FL LMFAL SWLV FQ IEFVGTAL GAAFL GVOAAV IAL IVRAV HRIG E
UW4_ChrA           GVL FVLP SL FILL IAL SWVY IA FG EL PVVAG -L FY GIK PAVT AIVVOA HRIG S

160         170         180         190         200         210
P. aeruginosa ChrA H I L L D R W L W V I A ----- I V C A L A A I G R V D F W I T --- L P A G G -
UW4_ChrA           R A L K N G W L W G I A A A S F V A I F V I N L P F P L I V L G A A V I G Y F G G R L A P E K F R T G G H

220         230         240         250         260
P. aeruginosa ChrA ----- L V Y A L L V L N ----- H R A S A L L V T L A A V A L A A A V A L W A A P T A K L V E A V V
UW4_ChrA           S A A K K S Y G P A V I D D D T P S P E H A R F S G L K V A R L A V I G A I - L W T L P M A M L T - A L F

270         280         290         300         310
P. aeruginosa ChrA Q G Q A S V L L I F A S G L K A G L L T F G G A Y T A I P F V R N D A V G R - G W M T D G Q F L D G L A L
UW4_ChrA           G W E G T L T O M A W F F T K A A L L T F G G A Y A V L P Y V Y O G A V G H Y G W L T P T O M I D G L A L

320         330         340         350         360         370
P. aeruginosa ChrA S G V L E A P L I I F A T F V G Y V A ----- G G P I G A V A M T V G V F L P A F A F
UW4_ChrA           G E T T P G P L I M V V A F V G F I G A Y V Q O V F G A D Q A F V A G A V A A T L V T W F T F L P S F L F

380         390         400         410         420
P. aeruginosa ChrA S L I F Y D R L E A V V E N K R L H A F L D G V A A G V V G L I G A T T I D L A Q -- V T A E R V P S L T
UW4_ChrA           I L A G G P L V E S T H N E L T F T A P L T A I T A A V V G V I L N L A C F F A Y H V L W P K G F E G T L

430         440         450         460         465
P. aeruginosa ChrA V G M S I F A A G L A F L Y A W K N K L N V V V I L A A C L A C W L V F P N Q G
UW4_ChrA           D W P S L L I A I V A G I A L L R F K R G V I E V L I G C G L I G L V V H L M F
```

(B)

```
1          10          20          30          40          50
KT2440_ChrA MAEHLQETAGOT PASVGLGOAL LFWLKLGFEV SFGG PAGQ IA IMHQELV E R R R W I S E
UW4_ChrA       MSSTPSPITQOAVPQPRRTISLREAF LFWLKLGFI SFGGPAGQIS IMHQELV E R R R W I S E

60          70          80          90          100         110
KT2440_ChrA RRFLHALNY CMLL PGPEAQOLATY IGWLMHRTWGGV IAGAL FVLP S L F I L I G L S W V Y L T
UW4_ChrA     RRFLHALNY CMLL PGPEAQOLATY IGWLMHRTWGGV IAGV L F V L P S L F I L I A L S W V Y I A

120         130         140         150         160         170
KT2440_ChrA W S E V P V V A G I F E G I K P A V T A I V V H A A H R I G T R A L K N G W L W A I A A A S F V A I F A L D V P F P L
UW4_ChrA     F G E L P V V A G L F Y G I K P A V T A I V V O A A H R I G S R A L K N G W L W G I A A A S F V A I F V I N L P F P L

180         190         200         210         220         230
KT2440_ChrA I V L A A A L L G Y V C G R L R P A O F A P G G H S A V R A R Q R P A L I D D D T P I P A H A C F S W G R L V T L L
UW4_ChrA     I V L G A A V I G Y F C G R L A P E K F R T G - G H S A A K K S Y G P A V I D D D T P S P E H A R F S G L K V A R L A

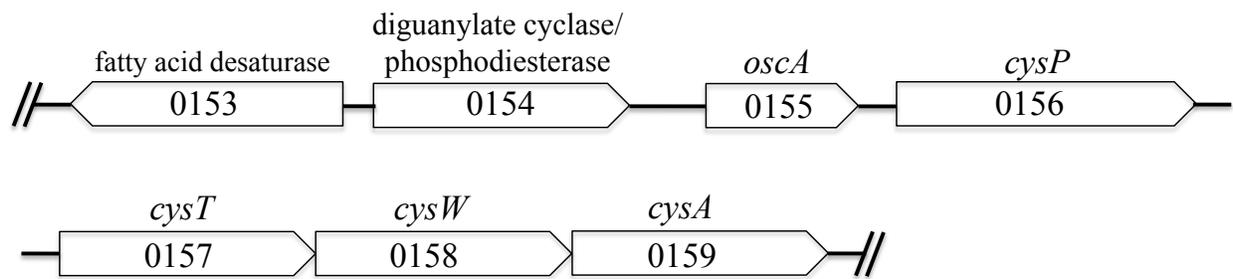
240         250         260         270         280         290
KT2440_ChrA L T V A S L W L L P M G L L M A T F G W E G T L T O M G W F F T K A A L L T F G G A Y A V L P Y V Y O G A V G H Y G W
UW4_ChrA     V I G A I L W T L P M A M L T A L F G W E G T L T O M A W F F T K A A L L T F G G A Y A V L P Y V Y O G A V G H Y G W

300         310         320         330         340         350
KT2440_ChrA L S P T O M I D G L A L G E T T P G P L I M V V A F V G F V G Y L Q P M F G A H Q A F L A G A V A A G L V T W F T F
UW4_ChrA     L T P T O M I D G L A L G E T T P G P L I M V V A F V G F I G A Y V Q O V F G A D Q A F V A G A V A A T L V T W F T F

360         370         380         390         400         410
KT2440_ChrA L P S F L F I L A G G P L V E S T H N E L K F T A P L T G I T A A V V G V I L N L A L F F G Y H V L W P D G F D G A F
UW4_ChrA     L P S F L F I L A G G P L V E S T H N E L T F T A P L T A I T A A V V G V I L N L A C F F A Y H V L W P K G F E G T L

420         430         440         450         453
KT2440_ChrA H W S A V T I A A A V A L F R V K R G V I H V L I G C A V A G L A L H L F R
UW4_ChrA     D W P S L L I A I V A G I A L L R F K R G V I E V L I G C G L I G L V V H L M F
```

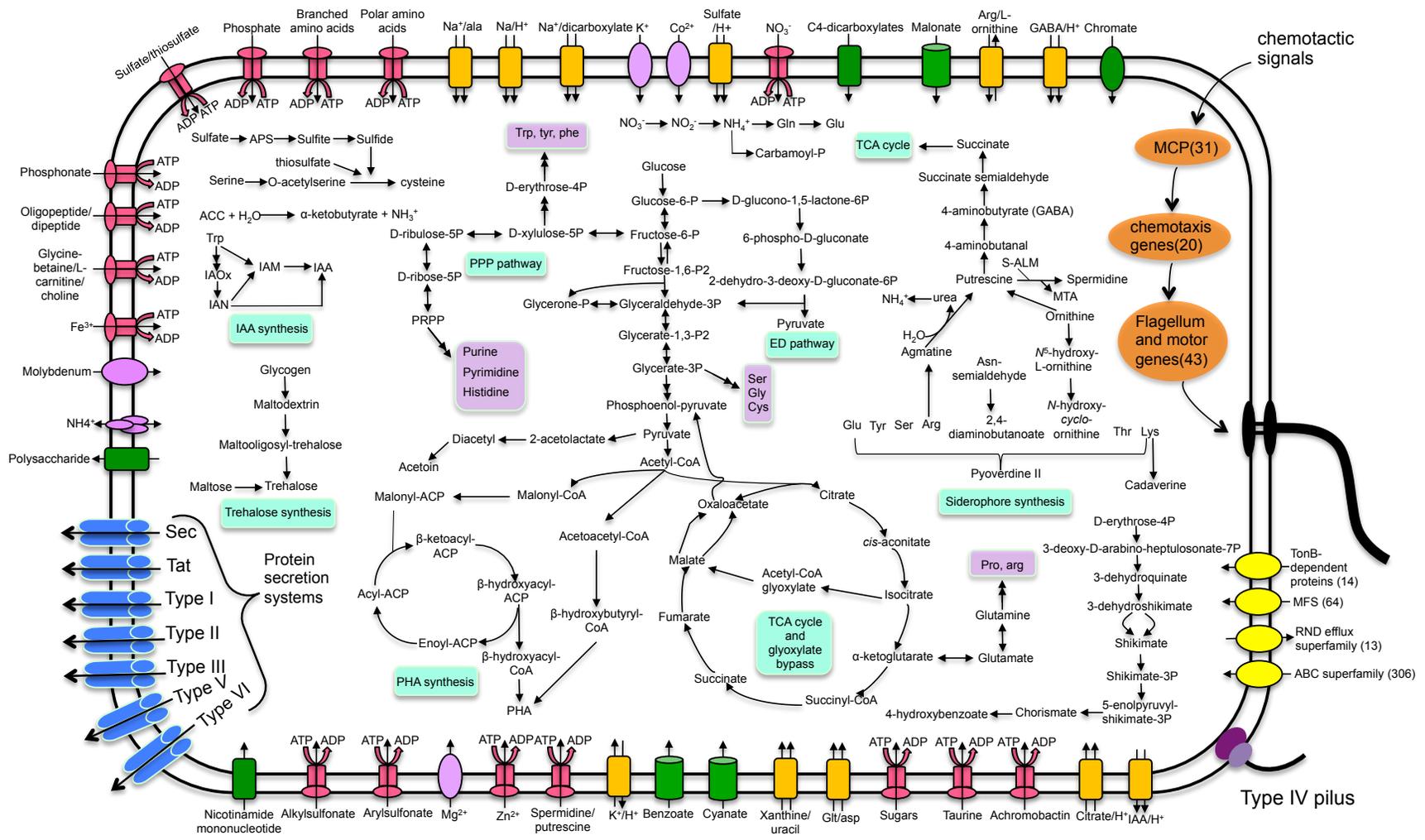
Fig. 4-26. Genetic organization of *oscA* in the genome of *P. putida* UW4. *cysP*, encoding sulphate-binding protein; *cysT*, encoding sulfate ABC transporter permease protein; *cysW*, encoding sulfate ABC transporter permease protein; *cysA*, encoding sulfate ABC transporter ATP-binding protein.



glycolysis/gluconeogenesis, a tricarboxylic acid (TCA) cycle with glyoxylate bypass, and a pentose phosphate pathway (PPP).

Metabolism of sulfur in UW4 involves assimilation of inorganic sulfate and mineralization of organic sulfonates. Inorganic sulfate or thiosulfate is transported into the cell by an ABC-type transporter including a periplasmic binding protein, Sbp (PputUW4\_3831) for sulfate or CysP (PputUW4\_0156) for thiosulfate, permease CysT and CysW, and an ATPase CysA. Sulfate and thiosulfate use the same permease components and ATPase for transport. Once in the cell, sulfate is activated to adenosine-5'-phosphosulfate (APS) by sulfate adenylyltransferase, CysDN (PputUW4\_0798, 0799), and then to sulfite by phosphoadenosine phosphosulfate reductase, CysH (PputUW4\_3670). Sulfite is further reduced to sulfide by sulfite reductase, CysI (PputUW4\_2355). This sulfide then joins *O*-acetylserine catalyzed by cysteine synthase, CysK (PputUW4\_4008) to form cysteine. In the case of thiosulfate, a gene encoding *O*-acetylserine sulfhydrylase, CysM (PputUW4\_4112), catalyzes the reaction between thiosulfate and *O*-acetylserine to generate *S*-sulfocysteine, which is then converted to cysteine (Kertesz 2004; Chambers and Trudinger 1971). In the genome of UW4, six SulP family sulfate transporters were identified (PputUW4\_0023, 0047, 0619, 2921, 3097, 4199). Although the role of these transporters in sulfate assimilation in bacteria is not clear, the homologs in several eukaryotes have been characterized and shown to be active components of sulfate transport, some of which function as sulfate:H<sup>+</sup> symporters (Piłsyk and Paszewski 2009). Organosulfur compounds are widely present in nature. For example, in aerobic soils organic sulfur can make up greater than 95% of the total sulfur in the forms of peptides/amino acids, sulfonates (C-SO<sub>3</sub>H), sulfamates (C-NH-SO<sub>3</sub>H), and sulfate esters (C-O-SO<sub>3</sub>H) (Kertesz 2004). Desulfonation of alkanesulfonates by UW4 is potentially catalyzed by alkanesulfonate monooxygenase, SsuD (PputUW4\_5216), and an NADPH-dependent FMN reductase, SsuE (PputUW4\_5218). The two

Fig. 4-27. Schematic overview of metabolic pathways and transport systems in *P. putida* UW4. Individual pathways are denoted by single-headed arrows, while reversible pathways are denoted by double-headed arrows.



- ATP-dependent transporters
  - Symporters/antiporters
  - Ion channels
  - Secondary transporters
  - Transporters families
- ABC: ATP-binding cassette  
 ACC: 1-aminocyclopropane-1-carboxylate  
 APS: adenosine-5'-phosphosulfate  
 ED: Entner Duodroff  
 GABA: 4-aminobutyrate  
 IAA: Indole-3-acetate  
 IAM: Indole-3-acetamide  
 IAN: Indole-3-acetonitrile  
 IAOx: Indole-3-acetaldoxime  
 MCP: Methyl-accepting chemotaxis protein  
 MFS: Major facilitator superfamily  
 MTA: S-methyl-5'-thioadenosine  
 PHA: Polyhydroxyalkanoate  
 PPP: Pentose phosphate pathway  
 PRPP: 5-phospho-α-D-ribose 1-diphosphate  
 RND: Resistance-nodulation-cell division  
 S-ALM: S-adenosyl-L-methioninamine  
 TCA: Tricarboxylic acid

genes are located within an operon, *ssuEADCBF* (PputUW4\_5213-5218), which also includes sulfonate transporter genes, *ssuABC*, and a molybdenum-pterin binding protein gene, *ssuF*. Similar to other *Pseudomonas putida* strains, a gene encoding the thiol-specific antioxidant, LsfA, was found upstream of *ssuE*. It has been demonstrated that expression of *lsfA* increased dramatically under sulfate starvation (Quadroni et al. 1999). Taurine is a naturally occurring aliphatic sulfonate. In the genome of UW4, two operons that each contains four genes encoding an ABC-transporter (*tauABC*) (PputUW4\_0121-0123 and PputUW4\_0200-0202) and a taurine dioxygenase (*tauD*) (PputUW4\_0120 and PputUW4\_0199) were identified. In addition, a third set of genes *tauA* (PputUW4\_5223) and *tauD* (PputUW4\_0897) are present in the genome. However, neither of them is associated with other *tau* genes. Like sulfonates, sulfate esters are components commonly present in soil. A sulfatase gene cluster that might be involved in desulfurization of aryl and alkylsulfate esters of UW4 was identified. The cluster contains seven genes, *atsACBR-sftR-atsK-sftP* (PputUW4\_0166-0172), which encode arylsulfatase, sulfate ester transporter ATP-binding component, aliphatic sulfonates ABC transporter permease, periplasmic aliphatic sulfonates-binding protein, LysR family transcriptional regulator, alkylsulfatase, and TonB-dependent receptor, correspondingly. It has been reported that in many gram-negative bacteria a LysR-type transcriptional regulator, CysB, mediated global sulfur regulation. Under the sulfur limitation conditions, CysB activates the transcription of cysteine synthesis genes in the presence of *N*-acetylserine or *O*-acetylserine, whereas sulfide and thiosulfate function as corepressors by inhibiting the binding of CysB to the promoters of the cysteine synthesis genes (Ostrowski and Kredich 1989, 1990; Hryniewicz and Kredich 1991). In UW4, a gene encoding CysB was identified (PputUW4\_1424) and it contains a typical helix-turn-helix motif at the N terminus for binding to the target DNA.

*P. putida* UW4 is unable to fix nitrogen and it also lacks the genes for denitrification. However, it contains the genes for assimilatory nitrate reduction. Two types of nitrate transporters are present on the chromosome of UW4 including an ABC-type nitrate transporter system and a NarK family transporter NasA. The locus of the ABC transporter system contains three genes that encode a nitrate transporter periplasmic protein (PputUW4\_2324), a nitrate transporter permease (PputUW4\_2325), and a nitrate transporter ATP-binding protein (PputUW4\_2326). NasA is located within a cluster of eight genes, *nasST-nasA-ppkB-nasDEC-cobA* (PputUW4\_3643-3650), which is potentially involved in nitrate/nitrite assimilation. The gene *nasS* encodes a periplasmic nitrate-binding protein and *nasT* encodes a response regulator that acts as an inducer of the *nas* operon in response to the presence of nitrate/nitrite (Gutierrez et al. 1995; Caballero et al. 2005). It has been shown that NasA is a nitrate transporter and a *nasA* mutant was unable to grow on nitrate but capable of growing on nitrite (Moir and Wood 2001). The genes *nasDEC-cobA* are located within an operon and they encode assimilatory nitrite reductase (NasDE), assimilatory nitrate reductase (NasC), and uroporphyrin III methyltransferase (CobA), respectively. Uroporphyrin III methyltransferase is an enzyme responsible for siroheme synthesis and the gene was induced strongly by nitrate (Wang et al. 2000). Furthermore, a siroheme synthetase homolog gene mutant of *Rhizobium etli* was unable to grow on nitrate as the sole nitrogen source (Tate et al. 1997). The gene *ppkB* (PputUW4\_3646), which is located immediately downstream of *nasA*, encodes a serine/threonine protein kinase. It has been demonstrated that a protein kinase carried out phosphorylation of the nitrate transporter and played an important role in nitrate deprivation response in *A. thaliana* and *Hansenula polymorpha* (Hu et al. 2009; Martín et al. 2011).

Many soil bacteria are capable of solubilizing poorly soluble mineral phosphates by synthesizing organic acids and acid phosphatases. In the genome of UW4, the genes responsible

for gluconic acid synthesis were found. The production of gluconic acid is catalyzed by glucose dehydrogenase (PputUW4\_0992) and its cofactor PQQ. The PQQ biosynthetic genes of UW4 are clustered in two separate loci on the chromosome: the *pqqABCDEFH* (PputUW4\_4969-4975) and the *pqqBCDE* (PputUW4\_2943-2946). In addition, five putative acid phosphatase-encoding genes were identified including two phosphatidic acid phosphatase (PAP2) protein genes (PputUW4\_0633, 4390), two SurE superfamily protein genes (PputUW4\_1119, 1674), and one non-specific acid phosphatase gene (PputUW4\_2829). However, no phytase gene is present in UW4. Inorganic phosphate uptake in UW4 may be facilitated by two high-affinity phosphate transport systems: PstBACS (PputUW4\_5366-5369) and PhnDCE1E2 (PputUW4\_3168-3171), and one low-affinity phosphate transport system, PitA (PputUW4\_1200). The high-affinity phosphate uptake system is composed of multi-subunit ABC transporters and is induced by phosphate-starvation, whereas the low-affinity system consists of a single membrane protein and is constitutively expressed (Gebhard et al. 2009).

*P. putida* UW4 has seven major protein secretion systems including Sec, Tat, Type I, II, III, V and VI (Table 4-14). The Sec (general secretory pathway) and Tat (twin arginine translocation) systems are the two ubiquitous systems for export across the cytoplasmic membrane. UW4 has one of each such system. MscL is a large conductance mechanosensitive channel protein and is able to export small proteins in response to osmotic pressure changes within the cell (Kloda et al. 2008). Type I secretion system (T1SS) consists of an outer membrane protein, an ABC transporter, and a membrane fusion protein. Three complete T1SS and their putative substrates were identified in UW4 (PputUW4\_0114, 0117-0119, 1722-1725, 3955-3958). In addition, one partial T1SS containing only an ABC transporter and a membrane fusion protein was found (PputUW4\_2636-2638). The putative substrate, mannuronan C-5-epimerase, is located downstream of the membrane fusion protein and is transcribed in an opposite direction.

Table 4-14. Protein secretion systems in *P. putida* UW4.

System	Gene	PputUW4
<i>(a) Sec secretory pathway</i>		
Chaperone	<i>secB</i>	0281
ATPase	<i>secA</i>	4321
TM complex	<i>secY</i>	4869
	<i>secE</i>	4902
	<i>secG</i>	0706
	<i>secD</i>	0923
Auxiliary proteins	<i>secF</i>	0924
	<i>yajC</i>	0922
	<i>ftsY</i>	5140
Membrane insertion proteins	<i>ffh</i>	4439
	<i>yidC</i>	5445
	<i>lepB</i>	0960
Spase I	<i>lspA</i>	4672
Spase II		
<i>(b) Twin arginine targeting (Tat) secretory pathway</i>		
	<i>tatA</i>	0323
	<i>tatB</i>	0322
	<i>tatC</i>	0321
<i>(c) Large conductance mechanosensitive ion channel</i>		
Channel	<i>mscL</i>	4710
<i>(d) Type I secretory system</i>		
TolC family outer membrane protein		0117
		1722
		3956
		3958
HylD family membrane fusion protein		0119
		1723
		2637
		3958
ATPase		0118
		1724
		2638
		3957
Putative substrate: RTX toxins and related Ca <sup>2+</sup> binding proteins		0114
		1725
		2636
		3955

Table 4-14. Continued.

<i>(e) Type II secretory system</i>		
Secretin	<i>gspD</i>	3290
Inner membrane protein	<i>gspF</i>	3303
	<i>gspG</i>	3287
		3288
		3289
		3302
APTase	<i>gspE</i>	3295
		5229
<i>(f) Type III secretory system</i>		
HopJ type III effector protein		0810
Apparatus protein	<i>orgAB</i>	3618
Apparatus lipoprotein	<i>prgK</i>	3619
Needle complex protein	<i>prgJ</i>	3620
Cytoplasmic protein	<i>prgI</i>	3621
Needle complex protein	<i>prgH</i>	3622
Cell invasion protein	<i>iagB</i>	3623
Invasion protein regulator	<i>hilA</i>	3624
Acyl carrier protein	<i>iacP</i>	3625
Cell invasion protein	<i>sipD</i>	3626
Invasin	<i>sipC</i>	3627
	<i>sipB</i>	3628
	<i>sicA</i>	3629
Chaperone	<i>spaS</i>	3630
Surface presentation of antigens protein	<i>spaR</i>	3631
	<i>spaQ</i>	3632
	<i>spaP</i>	3633
	<i>spaO</i>	3634
	<i>invJ</i>	3635
Invasion protein	<i>invI</i>	3636
Secretory apparatus	<i>invC</i>	3637
Secretory apparatus ATP synthase	<i>invB</i>	3638
Invasion protein	<i>invA</i>	3639
	<i>invE</i>	3640
	<i>invG</i>	3641
	<i>invF</i>	3642

Table 4-14. Continued.

<i>(g) Type V secretory system</i>		
Autotransporters	<i>estA</i>	4925
		2802
		0219
<i>(h) Type VI secretory system</i>		
OmpA/MotB domain-containing protein		3076
Lysozyme-like protein	<i>tssE</i>	3077
Lipoprotein	<i>tssJ</i>	3078
Secretion protein		3079
Secretion protein		3080
Secretion-associated protein		3081
ImcF domain-containing protein	<i>tssM</i>	3082
Hypothetical protein		3083
PAAR repeat-containing protein		3084
Hypothetical protein		3085
Hypothetical protein		3086
Hypothetical protein		3087
Vgr family protein	<i>vgrG1</i>	3088
ClpV1 family type VI secretion ATPase	<i>clpV</i>	3089
Hcp1 family effector	<i>hcp</i>	3090
OmpA/MotB domain-containing protein		3091
Hypothetical protein	<i>tssL</i>	3092
Secretion protein		3093
EvpB family secretion protein	<i>tssC</i>	3094
Secretion protein	<i>tssB</i>	3095
Vgr family protein	<i>vgrG2</i>	1753
Vgr family protein	<i>vgrG3</i>	3284

Since this system lacks the outer membrane protein, the transport mechanism of this large extracellular protein (1871 aa) is not clear. Genes involved in the type II secretion system (T2SS) of UW4 are located mainly within one cluster consisting of two separate operons (PputUW4\_3287-3295 and PputUW4\_3302-3303). The first operon contains nine genes but only five can be identified as T2SS protein genes based on sequence similarities. The other four genes encode three hypothetical proteins and a fimbrial assembly protein. The genes potentially involved in UW4 type III secretory pathway were found within one locus and showed strong similarity to those located within pathogenicity island I of *Salmonella*, suggesting that these genes were likely acquired by horizontal transfer. However, one gene *invH*, which encodes an outer membrane lipoprotein located upstream of *invF* in *Salmonella*, is absent in UW4. It has been demonstrated that in *Salmonella*, InvH is required for efficient adherence and invasion of cultured epithelial cells (Altmeyer et al. 1993). In addition, InvH is also essential for the proper localization of the secretin, InvG, in the outer membrane and the secretion of the virulence effector SipC (Daepler and Russel, 1998). Since UW4 does not have this gene, the function of the type III secretion system (T3SS) of UW4 needs to be confirmed experimentally. The type V secretion system (T5SS) of Gram-negative bacteria contains two steps: inner membrane transport via Sec pathway and outer membrane transport by a  $\beta$  barrel protein. Currently, two subtypes of T5SS have been identified including the autotransporters (ATs) and the two-partner secretion system (TPS). In UW4, three putative ATs were found. One of them, *estA*, possesses esterase activity and was shown to play an important role in twitching, swarming, and swimming motilities of *P. aeruginosa* (Wilhelm et al. 2007). The other two putative ATs in UW4 encode an outer membrane autotransporter (PputUW4\_2802) and an extracellular serine protease (PputUW4\_0219), respectively. However, none of these have been characterized experimentally.

The TPS system consists of two proteins. One protein named TpsA has a secretion motif and a catalytic domain. The other protein named TpsB contains the  $\beta$  domain involved in recruitment of the TpsA protein. Several TPS systems have been identified in *Pseudomonas* sp. such as *P. aeruginosa* PAO1, *P. fluorescens* Pf0-1 and *P. putida* KT2440 (Bleves et al. 2010; Molina et al. 2006). However, none of those systems is present in UW4. The type VI secretion system (T6SS) was first described in *Vibrio cholerae* six years ago (Pukatzke et al. 2006). Since then, the T6SS has been found in the genome of hundreds of bacteria, where it reportedly functions as a regulator of bacterial interactions and competition (Cascales and Cambillau, 2012). UW4 contains one gene cluster that is associated with T6SS. The cluster is composed of twenty genes (PputUW4\_3076\_3095) including the core components to form the minimal apparatus. Haemolysin coregulated protein (Hcp) forms hexamers and eventually assembles as nanotubes, which are responsible for transportation of other T6SS effector proteins. Another protein, valine-glycine repeat protein (VgrG), forms a trimer and serves as a puncturing device towards the targeted cells. Structures of Hcp and VgrG indicated that they are related to the needle tail and syringe components of bacteriophage T4. TssB (Type Six Secretion B) and TssC form structures similar to the bacteriophage needle sheath, and TssE resembles the needle hub. TssM, TssL, and TssJ are three proteins anchored to the bacterial cell envelope. TssJ, an outer membrane lipoprotein, interacts with the inner membrane protein TssM, which links the inner and outer membrane, and forms a stable complex with protein TssL (Felisberto-Rodrigues et al. 2011; Cascales and Cambillau, 2012; Durand et al. 2012).

Efficient plant growth stimulation requires effective root colonization that often relies on the bacterial cell surface structures, such as pili. Type IV pili are 5-7 nm fibers and the function is controlled by numerous genes. A total of twenty-four genes that are involved in type IV pili

biosynthesis were identified on the genome of UW4 (Table 4-15). These genes are arranged mainly within four clusters, *pilMNOPQ*, *pilA CD*, *pilEXWV-fimT*, *pilL/chpA-pilJIHG*, where the last cluster contains the genes involved in pili biosynthesis regulation.

#### 4.4 *Pseudomonas* Genome Comparisons and Phylogeny

A total of 1679 orthologous genes were identified between *P. putida* UW4 and other completely sequenced *Pseudomonas* genomes. Phylogenetic analysis of the 1679 conserved genes indicated that *P. putida* UW4 has a closer relationship with *P. fluorescens* than with *P. putida* (Fig. 4-28). The putative orthologous relations between UW4 and 20 completely sequenced *Pseudomonas* genomes are shown in Table 4-16. In addition, 71 CDSs were found in other *Pseudomonas* sp., whose genome sequences have not been determined (Table 4-17). In UW4 genome, 278 CDSs were considered as unique based on two criteria: 1. No hits to any CDS present in NCBI nr protein sequences database with a cutoff E-value of  $1 \text{ E}^{-20}$ ; 2. Identities are less than 30% and/or query/subject coverage is less than 80% (Table 4-18). Among the 278 CDSs, 240 have been annotated as hypothetical proteins. When comparing UW4 CDSs with those in nr database, 199 showed similarities with protein sequences in other genera only, indicating these genes probably originated from a genus outside of *Pseudomonas* (Table 4-19).

Comparisons of genome structure for UW4 vs completely sequenced *P. fluorescens* and *P. putida* genomes are illustrated in Fig. 4-29, with the red lines indicating individual TBLASTX matches and blue lines exhibiting inverted matches. The distribution of the genes among the *Pseudomonas* genomes showed that the unique genes are mostly located at the replication termini, whereas the orthologues are commonly present at the replication origin. Furthermore, the whole genome alignments showed extensive DNA rearrangement indicated by the blue lines, which is likely driven by repeat sequences within the genome. Moreover, the line plots revealed that genes in

Table 4-15. Type IV Pilus Genes in *P. putida* UW4

Gene	Gene Product	PputUW4
<i>pilM</i>	type IV pili biogenesis protein PilM	0346
<i>pilN</i>	type IV pili biogenesis protein PilN	0347
<i>pilO</i>	type IV pili biogenesis protein PilO	0348
<i>pilP</i>	type IV pili biogenesis protein PilP	0349
<i>pilQ</i>	type IV pili biogenesis protein PilQ	0350
<i>pilF</i>	type IV pili biogenesis protein PilF	0940
<i>pilZ</i>	type IV pilus assembly PilZ	3510
<i>pilZ</i>	type IV pilus assembly PilZ	4339
<i>fimD</i>	fimbrial biogenesis outer membrane usher protein	4454
<i>pilA</i>	type IV pili prepilin PilA	4645
<i>pilC</i>	type IV fimbrial assembly protein	4646
<i>pilD</i>	type IV pili prepilin peptidase PilD/XcpA	4647
<i>pilE</i>	type IV pili biogenesis protein PilE	4663
<i>pilX</i>	type IV pilus assembly protein	4665
<i>pilW</i>	type IV pilus assembly protein	4666
<i>pilV</i>	type IV fimbrial biogenesis protein PilV	4667
<i>fimT</i>	type IV pilin	4668
<i>fimT</i>	type IV pili biogenesis protein	4669
<i>traX</i>	conjugal transfer pilus acetylation protein	5018
<i>pilL/chpA</i>	type IV pili sensor histidine kinase/response regulator	5106
<i>pilJ</i>	type IV pili methyl-accepting chemotaxis transducer	5107
<i>pilI</i>	type IV pili signal transduction protein PilI	5108
<i>pilH</i>	type IV pili response regulator PilH	5109
<i>pilG</i>	type IV pili response regulator PilG	5110

Fig. 4-28. Phylogenetic tree of 21 different *Pseudomonas* species, base on 1,679 conserved genes. Numbers on nodes represent percentages of individual trees containing that relationship. The scale bar corresponds to the number of substitutions per site.

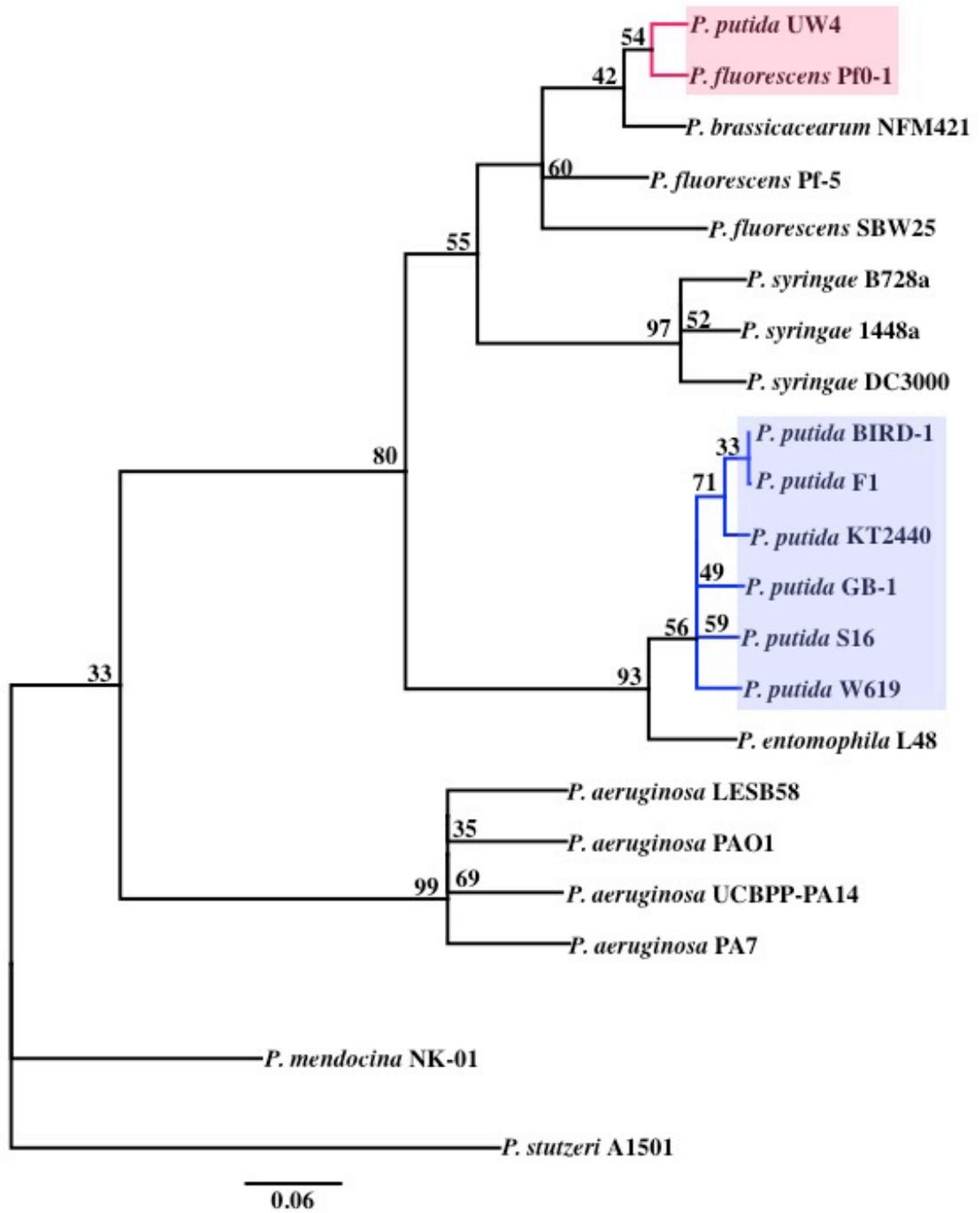


Table 4-16. Putative Orthologous Relations Between UW4 and completely sequenced *Pseudomonas* genomes

Species	Putative orthologous shared with <i>P. putida</i> UW4, no. (%)
<i>P. fluorescens</i> Pf0-1	4113 (76)
<i>P. fluorescens</i> Pf-5	3789 (70)
<i>P. fluorescens</i> SBW25	3595 (66)
<i>P. putida</i> KT2440	3356 (62)
<i>P. putida</i> GB-1	3434 (63)
<i>P. putida</i> F1	3435 (63)
<i>P. putida</i> BIRD-1	3394 (62)
<i>P. putida</i> W619	3405 (63)
<i>P. putida</i> S16	3466 (64)
<i>P. brassicacearum</i> NFM-421	3815 (69)
<i>P. entomophila</i> L48	3387 (62)
<i>P. syringae</i> DC3000	3015 (55)
<i>P. syringae</i> 1448a	2992 (55)
<i>P. syringae</i> B728a	3061 (56)
<i>P. mendocina</i> NK-01	2918 (54)
<i>P. stutzeri</i> A1501	2414 (44)
<i>P. aeruginosa</i> LESB58	3175 (58)
<i>P. aeruginosa</i> PAO1	3182 (58)
<i>P. aeruginosa</i> UCBPP-PA14	3185 (58)
<i>P. aeruginosa</i> PA7	3182 (58)

Table 4-17. Predicted CDSs that share similarities with other *Pseudomonas* sp.

locus_tag	product
PputUW4_0243	transcription repressor protein
PputUW4_0802	hypothetical protein
PputUW4_0918	type III restriction protein res subunit
PputUW4_1235	hypothetical protein
PputUW4_1655	short chain dehydrogenase
PputUW4_1656	dehydratase
PputUW4_1658	3-ketosteroid-5-isomerase
PputUW4_1662	2,6-dioxo-6-phenylhexa-3-enoate hydrolase
PputUW4_1670	short-chain dehydrogenase/reductase SDR
PputUW4_1671	lipase
PputUW4_1676	hypothetical protein
PputUW4_1677	3-oxosteroid-delta-1-dehydrogenase
PputUW4_1678	short chain dehydrogenase
PputUW4_1690	2,3-dihydroxybiphenyl-1,2-dioxygenase
PputUW4_1691	CoA-transferase subunit alpha
PputUW4_1692	CoA-transferase, beta subunit
PputUW4_1693	enoyl-CoA hydratase
PputUW4_1694	2-nitropropane dioxygenase
PputUW4_1695	enoyl-CoA hydratase
PputUW4_1696	acyl-CoA dehydrogenase
PputUW4_1697	acyl-CoA dehydrogenase
PputUW4_1699	short chain dehydrogenase
PputUW4_1740	LysR family transcriptional regulator
PputUW4_1815	leucyl-tRNA synthetase
PputUW4_1893	glutamine synthetase
PputUW4_1993	AraC family transcriptional regulator
PputUW4_2075	AMP-dependent synthetase and ligase
PputUW4_2077	AMP-dependent synthetase and ligase
PputUW4_2080	acyl-CoA dehydrogenase
PputUW4_2082	enoyl-CoA hydratase
PputUW4_2197	diguanylate cyclase
PputUW4_2257	ribosomal-protein-alanine acetyltransferase
PputUW4_2269	exported protein
PputUW4_2270	alpha/beta fold family hydrolase
PputUW4_2280	phenylserine aldolase
PputUW4_2308	long-chain-fatty-acid--CoA ligase
PputUW4_2568	plasmid stabilization system protein
PputUW4_2610	phage integrase

PputUW4_2622	AraC family transcriptional regulator
Table 4-17. Continued	
PputUW4_2624	alpha/beta hydrolase fold family protein
PputUW4_2646	hypothetical protein
PputUW4_3047	betaine aldehyde dehydrogenase
PputUW4_3144	aldehyde dehydrogenase
PputUW4_3257	hypothetical protein
PputUW4_3279	MutT/nudix-family hydrolase
PputUW4_3554	dTDP-4-dehydrorhamnose 3,5-epimerase
PputUW4_3593	hypothetical protein
PputUW4_3617	hypothetical protein
PputUW4_3618	oxygen-regulated gene required for bacterial internalization
PputUW4_3619	type III secretion apparatus lipoprotein
PputUW4_3620	type III secretion system needle complex protein
PputUW4_3621	type III secretion cytoplasmic protein
PputUW4_3622	type III secretion system needle complex protein
PputUW4_3624	invasion protein regulator
PputUW4_3625	acyl carrier protein
PputUW4_3626	type III effector protein cell invasion protein
PputUW4_3627	invasin
PputUW4_3628	invasin
PputUW4_3629	chaperone protein SicA
PputUW4_3630	surface presentation of antigens protein
PputUW4_3631	surface presentation of antigens protein
PputUW4_3632	surface presentation of antigens protein
PputUW4_3634	hypothetical protein
PputUW4_3635	hypothetical protein
PputUW4_3636	hypothetical protein
PputUW4_3640	invasion protein
PputUW4_3642	invasion protein
PputUW4_3921	hypothetical protein
PputUW4_4962	transcriptional regulator
PputUW4_4963	hypothetical protein
PputUW4_5044	hypothetical protein

Table 4-18. Putative unique CDSs in *P. putida* UW4

locus_tag	product
PputUW4_0044	hypothetical protein
PputUW4_0101	hypothetical protein
PputUW4_0115	conserved repeat domain protein
PputUW4_0143	hypothetical protein
PputUW4_0144	hypothetical protein
PputUW4_0145	hypothetical protein
PputUW4_0146	peptidase
PputUW4_0151	IS3 family transposase
PputUW4_0204	hypothetical protein
PputUW4_0219	extracellular serine protease
PputUW4_0231	hypothetical protein
PputUW4_0239	hypothetical protein
PputUW4_0359	hypothetical protein
PputUW4_0488	hypothetical protein
PputUW4_0499	hypothetical protein
PputUW4_0500	hypothetical protein
PputUW4_0539	hypothetical protein
PputUW4_0544	hypothetical protein
PputUW4_0545	hypothetical protein
PputUW4_0573	hypothetical protein
PputUW4_0594	hypothetical protein
PputUW4_0635	hypothetical protein
PputUW4_0639	hypothetical protein
PputUW4_0640	hypothetical protein
PputUW4_0645	hypothetical protein
PputUW4_0646	hypothetical protein
PputUW4_0647	hypothetical protein
PputUW4_0651	hypothetical protein
PputUW4_0658	hypothetical protein
PputUW4_0677	hypothetical protein
PputUW4_0726	hypothetical protein
PputUW4_0727	hypothetical protein
PputUW4_0737	hypothetical protein
PputUW4_0740	hypothetical protein
PputUW4_0741	hypothetical protein
PputUW4_0861	hypothetical protein
PputUW4_0866	hypothetical protein
PputUW4_0894	hypothetical protein
PputUW4_0895	hypothetical protein

Table 4-18. Continued

PputUW4_0903	hypothetical protein
PputUW4_0912	ion channel family protein
PputUW4_0926	hypothetical protein
PputUW4_0968	hypothetical protein
PputUW4_1014	hypothetical protein
PputUW4_1028	hypothetical protein
PputUW4_1030	hypothetical protein
PputUW4_1071	hypothetical protein
PputUW4_1183	hypothetical protein
PputUW4_1199	hypothetical protein
PputUW4_1213	hypothetical protein
PputUW4_1218	hypothetical protein
PputUW4_1234	LuxR family transcriptional regulator
PputUW4_1236	hypothetical protein
PputUW4_1238	hypothetical protein
PputUW4_1266	hypothetical protein
PputUW4_1267	hypothetical protein
PputUW4_1314	hypothetical protein
PputUW4_1375	hypothetical protein
PputUW4_1393	K <sup>+</sup> -transporting ATPase subunit F
PputUW4_1414	hypothetical protein
PputUW4_1487	transcriptional regulator
PputUW4_1488	hypothetical protein
PputUW4_1489	hypothetical protein
PputUW4_1490	hypothetical protein
PputUW4_1499	hypothetical protein
PputUW4_1547	hypothetical protein
PputUW4_1556	hypothetical protein
PputUW4_1557	hypothetical protein
PputUW4_1559	hypothetical protein
PputUW4_1567	hypothetical protein
PputUW4_1619	metallothionein
PputUW4_1666	hypothetical protein
PputUW4_1687	hypothetical protein
PputUW4_1706	hypothetical protein
PputUW4_1719	hypothetical protein
PputUW4_1726	hypothetical protein
PputUW4_1730	hypothetical protein
PputUW4_1743	hypothetical protein
PputUW4_1761	hypothetical protein
PputUW4_1812	Class II Aldolase/Adducin Family Protein

Table 4-18. Continued

PputUW4_1877	hypothetical protein
PputUW4_1921	hypothetical protein
PputUW4_1932	hypothetical protein
PputUW4_1946	hypothetical protein
PputUW4_1959	hypothetical protein
PputUW4_1996	hypothetical protein
PputUW4_2009	hypothetical protein
PputUW4_2015	IS3 family transposase
PputUW4_2019	hypothetical protein
PputUW4_2021	hypothetical protein
PputUW4_2046	aldehyde dehydrogenase family protein
PputUW4_2047	hypothetical protein
PputUW4_2097	hypothetical protein
PputUW4_2098	hypothetical protein
PputUW4_2100	hypothetical protein
PputUW4_2106	hypothetical protein
PputUW4_2124	IS1182 family transposase
PputUW4_2125	ThiF family protein
PputUW4_2126	hypothetical protein
PputUW4_2128	hypothetical protein
PputUW4_2129	hypothetical protein
PputUW4_2130	hypothetical protein
PputUW4_2131	hypothetical protein
PputUW4_2133	IS1182 family transposase, truncated
PputUW4_2135	hypothetical protein
PputUW4_2136	IS1182 family transposase
PputUW4_2147	hypothetical protein
PputUW4_2164	IS1182 family transposase
PputUW4_2167	hypothetical protein
PputUW4_2169	hypothetical protein
PputUW4_2171	hypothetical protein
PputUW4_2179	hypothetical protein
PputUW4_2183	hypothetical protein
PputUW4_2187	heme peroxidase
PputUW4_2188	hypothetical protein
PputUW4_2192	hypothetical protein
PputUW4_2202	hypothetical protein
PputUW4_2203	hypothetical protein
PputUW4_2205	hypothetical protein
PputUW4_2216	LysR family transcriptional regulator
PputUW4_2217	hypothetical protein

Table 4-18. Continued

PputUW4_2223	IS1182 family transposase
PputUW4_2284	hemolysin-type calcium-binding region
PputUW4_2285	hypothetical protein
PputUW4_2288	hypothetical protein
PputUW4_2307	LuxR family ATP-dependent transcriptional regulator
PputUW4_2353	hypothetical protein
PputUW4_2356	hypothetical protein
PputUW4_2371	hypothetical protein
PputUW4_2395	hypothetical protein
PputUW4_2397	hypothetical protein
PputUW4_2415	hypothetical protein
PputUW4_2447	hypothetical protein
PputUW4_2457	hypothetical protein
PputUW4_2481	hypothetical protein
PputUW4_2520	hypothetical protein
PputUW4_2534	hypothetical protein
PputUW4_2535	hypothetical protein
PputUW4_2537	hypothetical protein
PputUW4_2570	hypothetical protein
PputUW4_2574	hypothetical protein
PputUW4_2588	hypothetical protein
PputUW4_2589	AraC family transcriptional regulator
PputUW4_2601	hypothetical protein
PputUW4_2607	AraC family transcriptional regulator
PputUW4_2608	hypothetical protein
PputUW4_2609	hypothetical protein
PputUW4_2611	hypothetical protein
PputUW4_2620	hypothetical protein
PputUW4_2621	hypothetical protein
PputUW4_2632	hypothetical protein
PputUW4_2633	hypothetical protein
PputUW4_2657	hypothetical protein
PputUW4_2658	hypothetical protein
PputUW4_2668	hypothetical protein
PputUW4_2674	ion transport 2 domain-containing protein
PputUW4_2688	hypothetical protein
PputUW4_2705	hypothetical protein
PputUW4_2725	hypothetical protein
PputUW4_2727	hypothetical protein
PputUW4_2728	CadC family transcriptional regulator

Table 4-18. Continued

PputUW4_2729	hypothetical protein
PputUW4_2730	hypothetical protein
PputUW4_2731	AraC family transcriptional regulator
PputUW4_2732	hypothetical protein
PputUW4_2733	OmpA-like transmembrane domain protein
PputUW4_2734	hypothetical protein
PputUW4_2750	hypothetical protein
PputUW4_2761	DNA-binding response regulator
PputUW4_2762	hypothetical protein
PputUW4_2777	hypothetical protein
PputUW4_2807	hypothetical protein
PputUW4_2845	hypothetical protein
PputUW4_2870	hypothetical protein
PputUW4_2928	hypothetical protein
PputUW4_2993	hypothetical protein
PputUW4_2994	hypothetical protein
PputUW4_3001	IS110 family transposase
PputUW4_3027	hypothetical protein
PputUW4_3042	hypothetical protein
PputUW4_3048	hypothetical protein
PputUW4_3064	hypothetical protein
PputUW4_3085	hypothetical protein
PputUW4_3088	type VI secretion system Vgr family protein
PputUW4_3098	hypothetical protein
PputUW4_3103	hypothetical protein
PputUW4_3130	hypothetical protein
PputUW4_3133	hypothetical protein
PputUW4_3141	hypothetical protein
PputUW4_3142	hypothetical protein
PputUW4_3145	hypothetical protein
PputUW4_3146	hypothetical protein
PputUW4_3159	hypothetical protein
PputUW4_3189	hypothetical protein
PputUW4_3191	hypothetical protein
PputUW4_3214	glutamine amidotransferase class-I
PputUW4_3226	hypothetical protein
PputUW4_3246	hypothetical protein
PputUW4_3255	hypothetical protein
PputUW4_3270	hypothetical protein
PputUW4_3280	hypothetical protein
PputUW4_3286	hypothetical protein

Table 4-18. Continued

PputUW4_3297	hypothetical protein
PputUW4_3304	hypothetical protein
PputUW4_3311	hypothetical protein
PputUW4_3312	hypothetical protein
PputUW4_3324	hypothetical protein
PputUW4_3329	hypothetical protein
PputUW4_3349	hypothetical protein
PputUW4_3359	hypothetical protein
PputUW4_3416	hypothetical protein
PputUW4_3417	hypothetical protein
PputUW4_3418	hypothetical protein
PputUW4_3421	hypothetical protein
PputUW4_3478	hypothetical protein
PputUW4_3482	hypothetical protein
PputUW4_3486	hypothetical protein
PputUW4_3547	hypothetical protein
PputUW4_3548	hypothetical protein
PputUW4_3551	IS630 family transposase, truncated
PputUW4_3560	hypothetical protein
PputUW4_3587	hypothetical protein
PputUW4_3588	hypothetical protein
PputUW4_3590	hypothetical protein
PputUW4_3591	hypothetical protein
PputUW4_3594	hypothetical protein
PputUW4_3595	hypothetical protein
PputUW4_3597	hypothetical protein
PputUW4_3616	hypothetical protein
PputUW4_3677	hypothetical protein
PputUW4_3678	hypothetical protein
PputUW4_3679	surface antigen protein
PputUW4_3680	surface antigen protein
PputUW4_3681	hypothetical protein
PputUW4_3683	hypothetical protein
PputUW4_3763	hypothetical protein
PputUW4_3768	hypothetical protein
PputUW4_3812	hypothetical protein
PputUW4_3954	hypothetical protein
PputUW4_3967	hypothetical protein
PputUW4_3968	hypothetical protein
PputUW4_3996	aldehyde dehydrogenase family protein
PputUW4_4022	hypothetical protein

Table 4-18. Continued

PputUW4_4069	hypothetical protein
PputUW4_4083	hypothetical protein
PputUW4_4113	hypothetical protein
PputUW4_4114	hypothetical protein
PputUW4_4154	hypothetical protein
PputUW4_4157	hypothetical protein
PputUW4_4163	hypothetical protein
PputUW4_4164	hypothetical protein
PputUW4_4166	hypothetical protein
PputUW4_4172	hypothetical protein
PputUW4_4194	hypothetical protein
PputUW4_4232	hypothetical protein
PputUW4_4234	hypothetical protein
PputUW4_4311	hypothetical protein
PputUW4_4336	hypothetical protein
PputUW4_4362	hypothetical protein
PputUW4_4363	hypothetical protein
PputUW4_4500	hypothetical protein
PputUW4_4518	hypothetical protein
PputUW4_4519	hypothetical protein
PputUW4_4625	hypothetical protein
PputUW4_4664	type IV pilus-associated protein
PputUW4_4696	hypothetical protein
PputUW4_4698	hypothetical protein
PputUW4_4828	hypothetical protein
PputUW4_4911	hypothetical protein
PputUW4_4964	hypothetical protein
PputUW4_5019	hypothetical protein
PputUW4_5020	hypothetical protein
PputUW4_5024	hypothetical protein
PputUW4_5088	hypothetical protein
PputUW4_5206	glutamine amidotransferase
PputUW4_5210	hypothetical protein
PputUW4_5218	hypothetical protein
PputUW4_5350	hypothetical protein

Table 4-19. Predicted UW4 CDSs that share sequence similarities to those in other genera only

locus_tag	product
PputUW4_0444	hypothetical protein
PputUW4_0525	hypothetical protein
PputUW4_0644	hypothetical protein
PputUW4_0657	ADP-Ribosylglycohydrolase
PputUW4_0666	hypothetical protein
PputUW4_0739	hypothetical protein
PputUW4_0863	hypothetical protein
PputUW4_0864	hypothetical protein
PputUW4_0868	radical SAM family protein
PputUW4_0869	branched-chain amino acid aminotransferase
PputUW4_0870	hypothetical protein
PputUW4_0871	hypothetical protein
PputUW4_0872	hypothetical protein
PputUW4_0873	GntR family transcriptional regulator with aminotransferase domain
PputUW4_0884	histidine kinase
PputUW4_0885	two component transcriptional regulator
PputUW4_0886	methionine-R-sulfoxide reductase
PputUW4_0887	methionine sulfoxide reductase A
PputUW4_0892	hypothetical protein
PputUW4_1333	Glucose/sorbose dehydrogenases
PputUW4_1334	tat (twin-arginine translocation) pathway signal sequence domain protein
PputUW4_1373	hypothetical protein
PputUW4_1374	hypothetical protein
PputUW4_1376	hypothetical protein
PputUW4_1561	hypothetical protein
PputUW4_1593	hypothetical protein
PputUW4_1653	acyl-CoA dehydrogenase
PputUW4_1654	acyl-CoA dehydrogenase
PputUW4_1659	hypothetical protein
PputUW4_1660	hypothetical protein
PputUW4_1661	oxidoreductase, short chain dehydrogenase/reductase family
PputUW4_1668	LuxR family transcriptional regulator
PputUW4_1669	rieske [2Fe-2S] domain protein
PputUW4_1674	5'-nucleotidase SurE
PputUW4_1679	acyl-CoA synthetase
PputUW4_1680	hypothetical protein
PputUW4_1681	FAD-binding oxidoreductase
PputUW4_1683	fumarate reductase/succinate dehydrogenase

Table 4-19. Continued

PputUW4_1685	acyl-CoA dehydrogenase
PputUW4_1686	flavin-dependent oxidoreductase
PputUW4_1729	17 kDa surface antigen
PputUW4_1735	hypothetical protein
PputUW4_1736	arylsulfatase
PputUW4_1737	multiple antibiotic resistance (MarC)-like protein
PputUW4_1741	AMP-dependent synthetase and ligase
PputUW4_1744	enoyl-CoA hydratase
PputUW4_1808	Citrate lyase subunit beta
PputUW4_1809	MaoC Domain Protein Dehydratase
PputUW4_1810	ornithine cyclodeaminase
PputUW4_1811	D-3-phosphoglycerate dehydrogenase
PputUW4_1813	histidinol-phosphate aminotransferase
PputUW4_1858	LuxR family transcriptional regulator
PputUW4_1859	amidohydrolase family protein
PputUW4_1863	hypothetical protein
PputUW4_1864	D-alanyl-D-alanine dipeptidase
PputUW4_1883	hypothetical protein
PputUW4_1890	amino acid permease family protein
PputUW4_2012	AraC family transcriptional regulator
PputUW4_2022	hypothetical protein
PputUW4_2030	hypothetical protein
PputUW4_2033	Arginine/lysine/ornithine decarboxylases
PputUW4_2034	putrescine transporter
PputUW4_2053	hypothetical protein
PputUW4_2054	hypothetical protein
PputUW4_2061	dihydroxy-acid dehydratase
PputUW4_2066	TetR family transcriptional regulator
PputUW4_2067	short-chain dehydrogenase/reductase SDR
PputUW4_2068	MaoC-like dehydratase
PputUW4_2069	hypothetical protein
PputUW4_2070	thiolase
PputUW4_2072	LuxR family MalT-like ATP-dependent transcriptional regulator
PputUW4_2074	enoyl-CoA hydratase/carnithine racemase
PputUW4_2076	thioesterase superfamily protein
PputUW4_2079	thiolase
PputUW4_2081	acyl-CoA dehydrogenase
PputUW4_2087	alcohol dehydrogenase class III
PputUW4_2090	hypothetical protein
PputUW4_2091	hypothetical protein

Table 4-19. Continued

PputUW4_2092	AraC family transcriptional regulator
PputUW4_2096	hypothetical protein
PputUW4_2120	RND family efflux transporter, MFP subunit
PputUW4_2122	peptide ABC transporter permease
PputUW4_2123	peptide ABC transporter permease
PputUW4_2134	hypothetical protein
PputUW4_2140	HxlR family transcriptional regulator
PputUW4_2142	glutathione-dependent formaldehyde-activating GFA
PputUW4_2143	hypothetical protein
PputUW4_2145	RND family efflux transporter MFP subunit
PputUW4_2146	RND efflux system outer membrane lipoprotein
PputUW4_2148	transcriptional regulator ATPase, winged helix family
PputUW4_2150	TetR family transcriptional regulator
PputUW4_2152	TetR family transcriptional regulator
PputUW4_2154	fumarylacetoacetate (FAA) hydrolase
PputUW4_2157	hypothetical protein
PputUW4_2174	hypothetical protein
PputUW4_2185	LuxR family transcriptional regulator
PputUW4_2186	histidine kinase family protein
PputUW4_2189	sensor histidine kinase
PputUW4_2190	sensor histidine kinase
PputUW4_2191	two component LuxR family transcriptional regulator
PputUW4_2195	hypothetical protein
PputUW4_2206	hypothetical protein
PputUW4_2239	hypothetical protein
PputUW4_2260	hypothetical protein
PputUW4_2264	hypothetical protein
PputUW4_2283	twin-arginine translocation pathway signal
PputUW4_2286	patatin-like phospholipase family protein
PputUW4_2298	thiolase
PputUW4_2299	MaoC-like dehydratase
PputUW4_2300	MaoC-like dehydratase
PputUW4_2301	3-oxoacyl-[acyl-carrier-protein] reductase
PputUW4_2331	major facilitator superfamily MFS_1
PputUW4_2554	hypothetical protein
PputUW4_2571	hypothetical protein
PputUW4_2573	hypothetical protein
PputUW4_2585	hypothetical protein
PputUW4_2586	hypothetical protein
PputUW4_2587	hypothetical protein

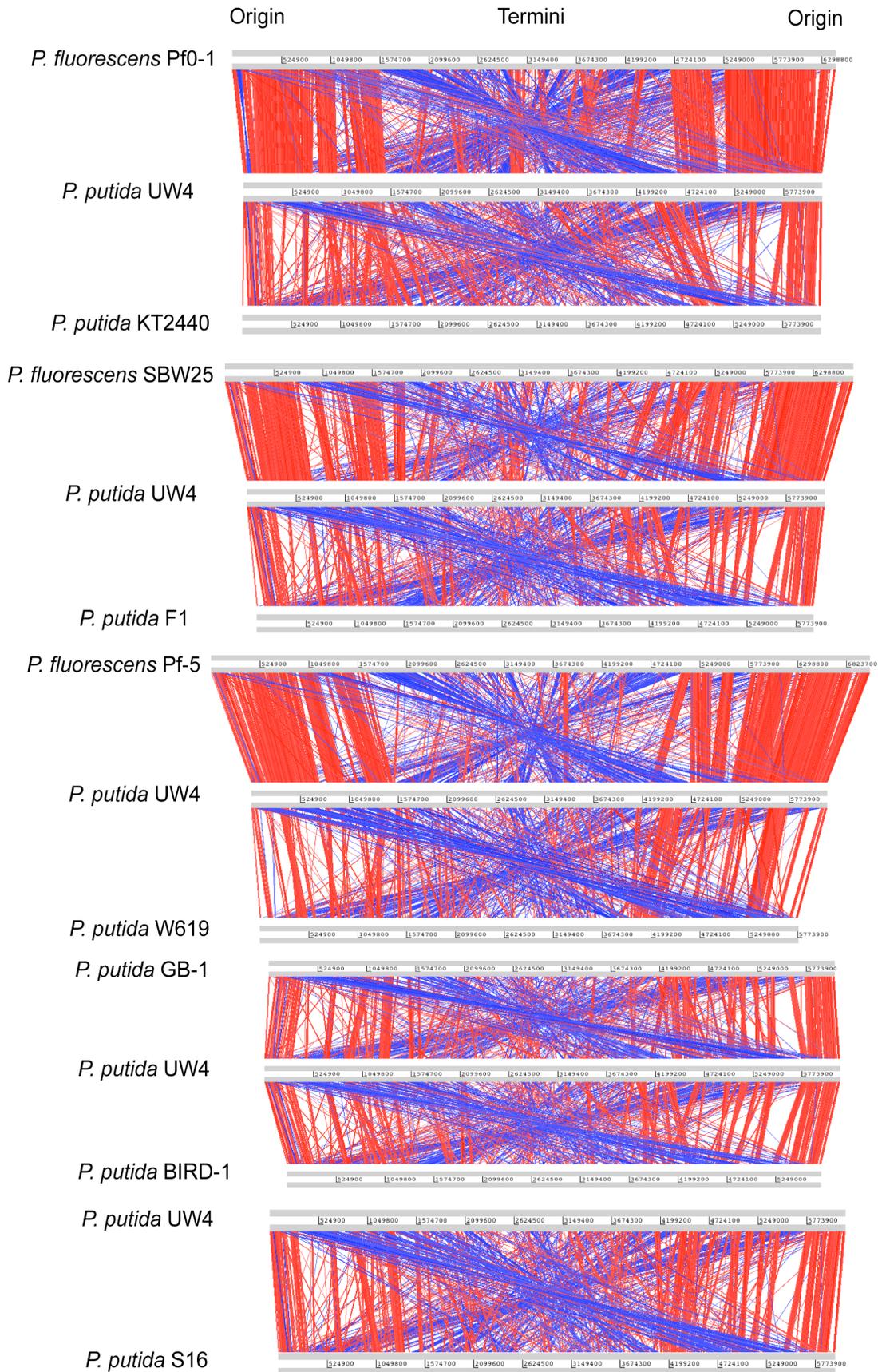
Table 4-19. Continued

PputUW4_2591	LuxR family ATP-dependent transcriptional regulator
PputUW4_2592	acetyl-CoA acetyltransferase
PputUW4_2597	Radical SAM domain protein
PputUW4_2598	oxidoreductase
PputUW4_2599	arylsulfatase
PputUW4_2600	2,4-dihydroxyacetophenone dioxygenase
PputUW4_2612	hypothetical protein
PputUW4_2625	major facilitator superfamily protein
PputUW4_2626	enoyl-CoA hydratase/isomerase
PputUW4_2627	AMP-dependent synthetase and ligase
PputUW4_2628	carbon monoxide dehydrogenase small subunit
PputUW4_2630	carbon-monoxide dehydrogenase, large subunit
PputUW4_2676	TetR family transcriptional regulator
PputUW4_2677	hypothetical protein
PputUW4_2681	glutathione-dependent formaldehyde-activating GFA
PputUW4_2692	hypothetical protein
PputUW4_2693	major facilitator family protein
PputUW4_2701	hypothetical protein
PputUW4_2702	MarR family transcriptional regulator
PputUW4_2708	isochorismatase family protein
PputUW4_2713	cytochrome b561 family protein
PputUW4_2714	hypothetical protein
PputUW4_2716	hypothetical protein
PputUW4_2717	cytochrome c oxidase subunit III
PputUW4_2718	cytochrome c oxidase subunit III
PputUW4_2719	cytochrome c oxidase, subunit I
PputUW4_2720	cytochrome c oxidase subunit II
PputUW4_2723	hypothetical protein
PputUW4_2724	hypothetical protein
PputUW4_2745	ABC transporter permease protein
PputUW4_2746	ABC transporter permease protein
PputUW4_2747	ABC transporter family protein
PputUW4_2748	ABC transporter family protein
PputUW4_2749	ABC transporter substrate-binding protein
PputUW4_2847	hypothetical protein
PputUW4_2995	LuxR family two component transcriptional regulator
PputUW4_3012	IclR family transcriptional regulator
PputUW4_3024	nitrilase
PputUW4_3026	glutathione S-transferase
PputUW4_3043	spermidine/putrescine ABC transporter ATPase subunit

Table 4-19. Continued

PputUW4_3044	spermidine/putrescine-binding periplasmic protein
PputUW4_3058	prolyl-tRNA synthetase
PputUW4_3086	hypothetical protein
PputUW4_3097	sulfate transporter
PputUW4_3104	hypothetical protein
PputUW4_3127	peroxidase
PputUW4_3148	phosphotransferase family protein
PputUW4_3149	hypothetical protein
PputUW4_3150	hypothetical protein
PputUW4_3151	macrolide glycosyltransferase
PputUW4_3152	spermidine/putrescine ABC transporter, ATP-binding protein
PputUW4_3155	spermidine/putrescine-binding periplasmic protein
PputUW4_3156	GntR family transcriptional regulator
PputUW4_3157	glycosyl transferase family protein
PputUW4_3158	FAD dependent oxidoreductase
PputUW4_3172	hypothetical protein
PputUW4_3173	ATP-dependent DNA helicase
PputUW4_3176	succinylglutamate desuccinylase/aspartoacylase
PputUW4_3177	hypothetical protein
PputUW4_3242	glutathione S-transferase
PputUW4_3325	hypothetical protein
PputUW4_3334	alpha/beta hydrolase fold protein
PputUW4_3335	LysR family transcriptional regulator
PputUW4_3336	demethylmenaquinone methyltransferase
PputUW4_3338	dimethylmenaquinone methyltransferase
PputUW4_3350	hypothetical protein
PputUW4_3383	LuxR family transcriptional regulator
PputUW4_3545	glycerol-3-phosphate cytidyltransferase
PputUW4_3546	polysaccharide biosynthesis protein
PputUW4_3549	glucosyltransferase
PputUW4_3550	glycosyltransferase
PputUW4_3589	hypothetical protein
PputUW4_3598	hypothetical protein
PputUW4_3638	surface presentation of antigens protein
PputUW4_4107	hypothetical protein
PputUW4_4153	two component LuxR family transcriptional regulator
PputUW4_4155	dihydrolipoamide acetyltransferase
PputUW4_4192	hypothetical protein
PputUW4_4195	histidine kinase
PputUW4_4695	GntR family transcriptional regulator
PputUW4_4697	hypothetical protein

Fig. 4-29. Comparative synteny line plots of the complete six-frame translations of the whole genome sequences of *P. putida* UW4 with other *P. fluorescens* and *P. putida* genomes. The analysis was carried out using Artemis Comparison Tool and computed using TBLASTX with a cutoff E value of  $1 \text{ E}^{-5}$ . The red bars between the DNA lines indicate individual TBLASTX matches, and the blue lines exhibit inverted matches. The cutoff identities and alignments length are 75% and 30 amino acids, respectively.



UW4 are more closely related to those in *P. fluorescens* than in *P. putida*, illustrated by the number of matches. This result is consistent with the results obtained from whole genome phylogenetic analysis.

#### 4.5 16S rRNA Genes Phylogenetic Analysis

16S rRNA gene sequences are highly conserved among the same bacterial species and are frequently used to identify and classify microorganisms. It has been observed that the number of rRNA genes in prokaryotic genomes can vary from one to as many as 15 copies and the intragenic diversity ranges from 0.06% to 20.38% (Pei et al. 2010). On the chromosome of UW4, seven ribosomal RNA (*rrn*) operons were identified. Among the seven 16S rRNA genes, three were found to have identical sequences (i.e., RNAs 3, 4 and 6). A ML phylogenetic tree was constructed for the unique 16S rRNA genes of *Pseudomonas* genomes (Fig. 4-30). Although the 16S rRNA genes of UW4 are grouped with those of *P. putida*, the node support is only 0.45, indicating low confidence for the classification.

Additional analysis was conducted using the unique 16S rRNA genes of UW4, *P. fluorescens* SBW25, *P. putida* KT2440, as well as the type strain of *P. fluorescens* and *P. putida*, IAM12022 (D84013) and IAM1236<sup>T</sup> (D84020), respectively (Anzai et al. 1997; Peix et al. 2009). Sequence alignment of the twelve 16S rRNA genes is shown in Fig. 4-31 and the pairwise identities are shown in Table 4-20. Results of the identities table shows that the intragenomic diversity of 16S rRNA genes of UW4 is below the common threshold (1 to 1.3%), and overall UW4 is phylogenetically closer to *putida* than to *fluorescens*. However, the diversities between UW4 and *putida* 16S rRNA gene sequences are all above the threshold, ranging from 1.7%-2.9%. In the case of *fluorescens*, the diversities range from 2.1%-3.5%. Also, it is important to note that

Fig. 4-30. ML phylogenetic tree of 16S rRNA sequences from completely sequenced *Pseudomonas* genomes. Nodal support was evaluated by aLRT. Different species are shown in different colors. Only unique sequences from each genome were included for this analysis.

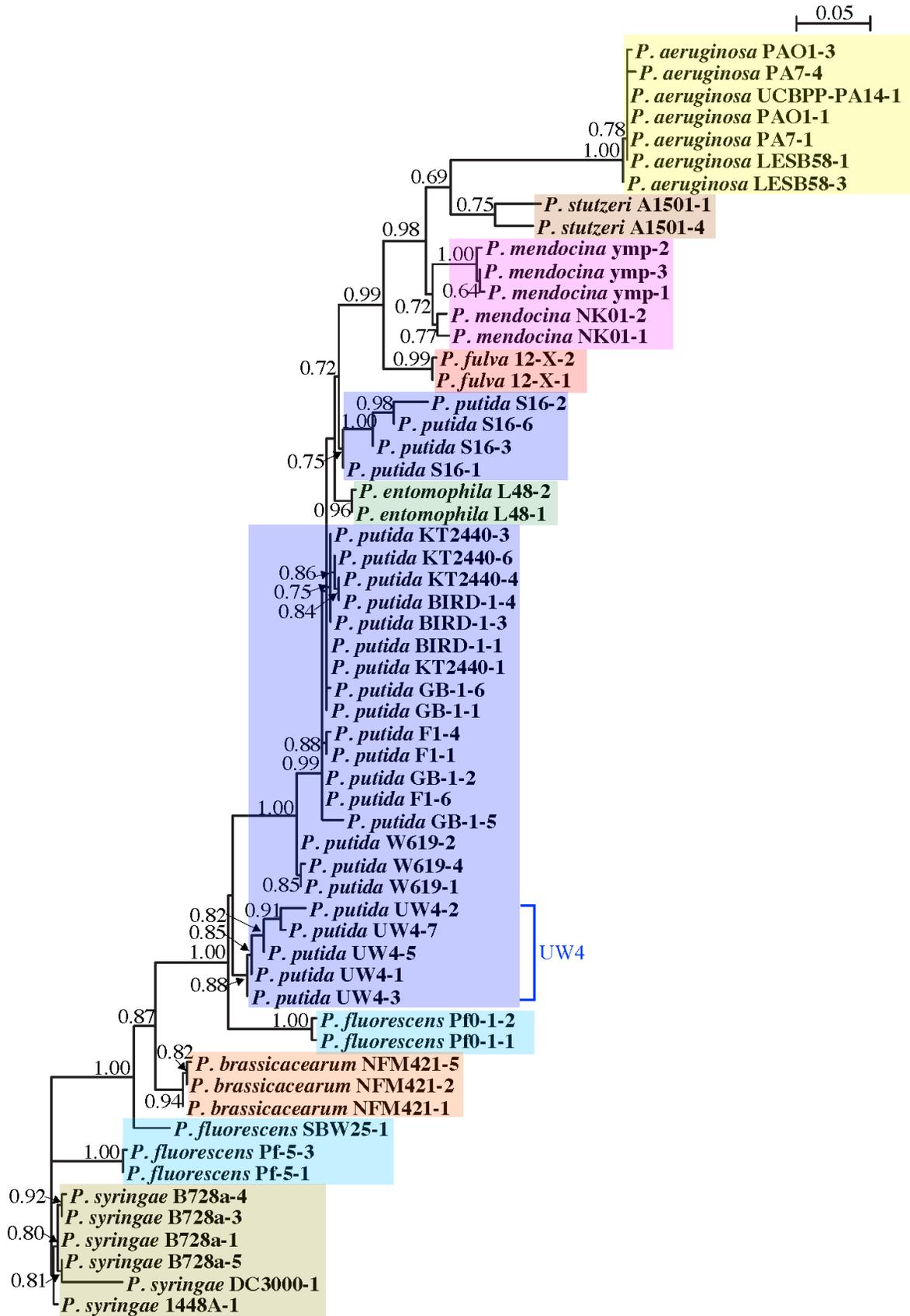


Fig. 4-31. 16S rRNA gene sequences alignment of UW4, *P. putida* KT2440, *P. fluorescens* SBW25, and the type strain of *P. putida* IAM1236<sup>T</sup> (D84020) and *P. fluorescens* IAM12022 (D84013). Only the unique genes were included for the analysis. The consensus sequence is showing above the alignment. The dots represent the nucleotides that are identical to the consensus sequences.



Consensus  
Identity

520 530 540 550 560 570 580 590

CCGCGGTAATACAGAGGGTCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCTRGGTGGTTTGGTTAAGTTGGATGTGAAA

1. SBW25\_16S-1  
2. IAM 12022  
3. UW4\_16S-2  
4. UW4\_16S-7  
5. UW4\_16S-5  
6. UW4\_16S-1  
7. UW4\_16S-3  
8. IAM 1236T  
9. KT2440\_16S-6  
10. KT2440\_16S-4  
11. KT2440\_16S-3  
12. KT2440\_16S-1

Consensus  
Identity

600 610 620 630 640 650 660 670 680

KCCCCGGGCTCAACCTGGGAACCTGCATYCAAAACTGRCWRRCTAGAGTAYGGTAGAGGGTGGTGGAAATTCCTGTGTAGCGGTGA

1. SBW25\_16S-1  
2. IAM 12022  
3. UW4\_16S-2  
4. UW4\_16S-7  
5. UW4\_16S-5  
6. UW4\_16S-1  
7. UW4\_16S-3  
8. IAM 1236T  
9. KT2440\_16S-6  
10. KT2440\_16S-4  
11. KT2440\_16S-3  
12. KT2440\_16S-1

Consensus  
Identity

690 700 710 720 730 740 750 760

AATCGGYAGATATAGGAAGGAACACCAGTGGCGAARGCCAGCCACTGGACTRATACAGACTGAGGTGCGAAASSGTGGGGAGC

1. SBW25\_16S-1  
2. IAM 12022  
3. UW4\_16S-2  
4. UW4\_16S-7  
5. UW4\_16S-5  
6. UW4\_16S-1  
7. UW4\_16S-3  
8. IAM 1236T  
9. KT2440\_16S-6  
10. KT2440\_16S-4  
11. KT2440\_16S-3  
12. KT2440\_16S-1

Consensus  
Identity

770 780 790 800 810 820 830 840 850

AAACAGGATTAGATACCCCTGGTAGTCCACGCCGTAAACGATGTC AAC TAGCCGTGGRAKCCCTTGACMTYTTAGTGGCGMRSTA

1. SBW25\_16S-1  
2. IAM 12022  
3. UW4\_16S-2  
4. UW4\_16S-7  
5. UW4\_16S-5  
6. UW4\_16S-1  
7. UW4\_16S-3  
8. IAM 1236T  
9. KT2440\_16S-6  
10. KT2440\_16S-4  
11. KT2440\_16S-3  
12. KT2440\_16S-1

Consensus  
Identity

860 870 880 890 900 910 920 930

ACGCATTAAGTTGACCGCTGGGGAGTACGGCCCAANGTTAAACTCAAATGAATTGACGGGGCCCGCACAAACNGTGGAGCA

1. SBW25\_16S-1  
2. IAM 12022  
3. UW4\_16S-2  
4. UW4\_16S-7  
5. UW4\_16S-5  
6. UW4\_16S-1  
7. UW4\_16S-3  
8. IAM 1236T  
9. KT2440\_16S-6  
10. KT2440\_16S-4  
11. KT2440\_16S-3  
12. KT2440\_16S-1

Consensus  
Identity

940 950 960 970 980 990 1,000 1,010 1,020

TGTGGTTTAAATTCGAAGCAACCGGAAGARCCCTTACCAGGCCCTTGACATSCARWGAACCTTTCYAGAGATRGATTGGTGCCTTCGGG

1. SBW25\_16S-1  
2. IAM 12022  
3. UW4\_16S-2  
4. UW4\_16S-7  
5. UW4\_16S-5  
6. UW4\_16S-1  
7. UW4\_16S-3  
8. IAM 1236T  
9. KT2440\_16S-6  
10. KT2440\_16S-4  
11. KT2440\_16S-3  
12. KT2440\_16S-1



Table 4-20. 16S rRNA Genes Sequences Pairwise Identities.

	UW4_16S-1	UW4_16S-5	UW4_16S-3	UW4_16S-7	UW4_16S-2
UW4_16S-1	ID	0.998	0.998	0.992	0.99
UW4_16S-5	0.998	ID	0.996	0.994	0.991
UW4_16S-3	0.998	0.996	ID	0.994	0.99
UW4_16S-7	0.992	0.994	0.994	ID	0.993
UW4_16S-2	0.99	0.991	0.99	0.993	ID
KT2440_16S-1	0.982	0.98	0.983	0.977	0.973
IAM_1236T	0.981	0.98	0.982	0.978	0.973
KT2440_16S-3	0.981	0.979	0.982	0.977	0.973
KT2440_16S-6	0.98	0.98	0.982	0.977	0.972
KT2440_16S-4	0.98	0.979	0.981	0.977	0.971
SBW25_16S-1	0.978	0.977	0.979	0.974	0.969
IAM_12022	0.973	0.973	0.975	0.97	0.965

ID: Identical

only one copy of 16S rRNA gene from each of the type strain, IAM12022 and IAM1236<sup>T</sup>, was available for the analysis.

#### 4.6 Heat Shock Protein Genes Phylogenetic and Promoter Analyses

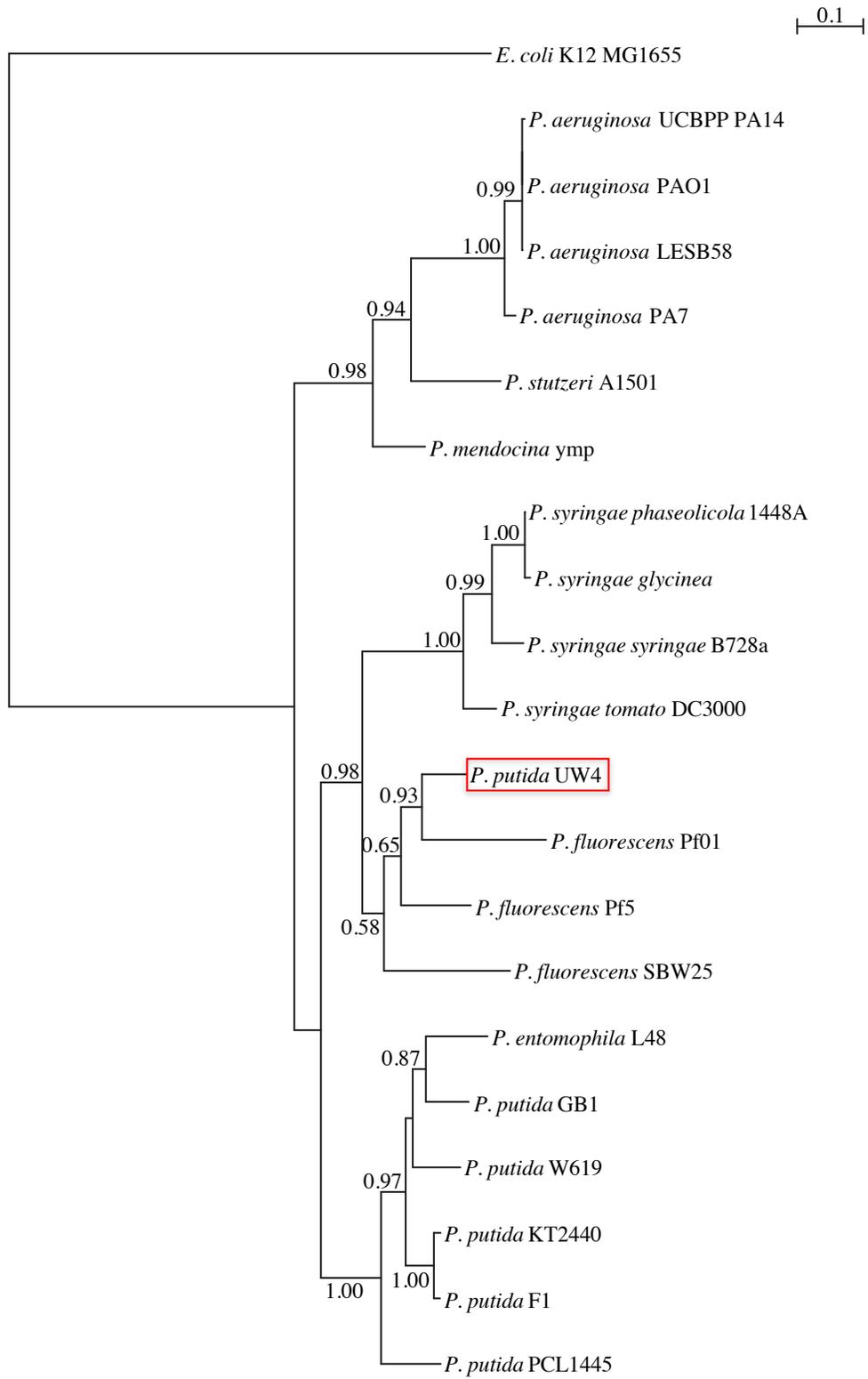
Heat shock protein (HSP) genes can be used in bacterial classification along with 16S rRNA genes because of their conservation and ubiquity (Ahmad et al. 2000). Phylogenetic analysis of seven HSPs in UW4 based on nucleotide sequences is shown in Figure 4-33A-G. ML trees of larger HSPs including DnaK, DnaJ, GroEL and ClpB are in close agreement with the *Pseudomonas* whole genome phylogenetic tree based on 1679 conserved genes, which illustrates the *aeruginosa* clade branching first, followed by *putida* clade, and finally the *syringae* and *fluorescens* clades. In addition, the phylogenetic relationships of *stutzeri* and *mendocina* vary among the four trees. For example, the DnaK tree shows *stutzeri* and *mendocina* as part of *aeruginosa*, whereas the GroEL and ClpB trees show *stutzeri* and *mendocina* after the *aeruginosa* clade. In the DnaJ tree, *mendocina* clade is before *aeruginosa* while *stutzeri* is after.

For smaller HSPs such as GrpE, GroES and IbpA, the trees are more dissimilar. In the GrpE tree, the *fluorescens* group is shown first, followed by *syringae*, and then *putida* and *aeruginosa* clades. Similar to the GrpE tree, the GroES tree also depicts the *fluorescens* and *syringae* groups earlier than the *putida* and *aeruginosa* groups. However, the IbpA tree is entirely different from the other trees, where the *fluorescens* group is shown first, followed by *aeruginosa*, and then by *syringae* and *putida* groups.

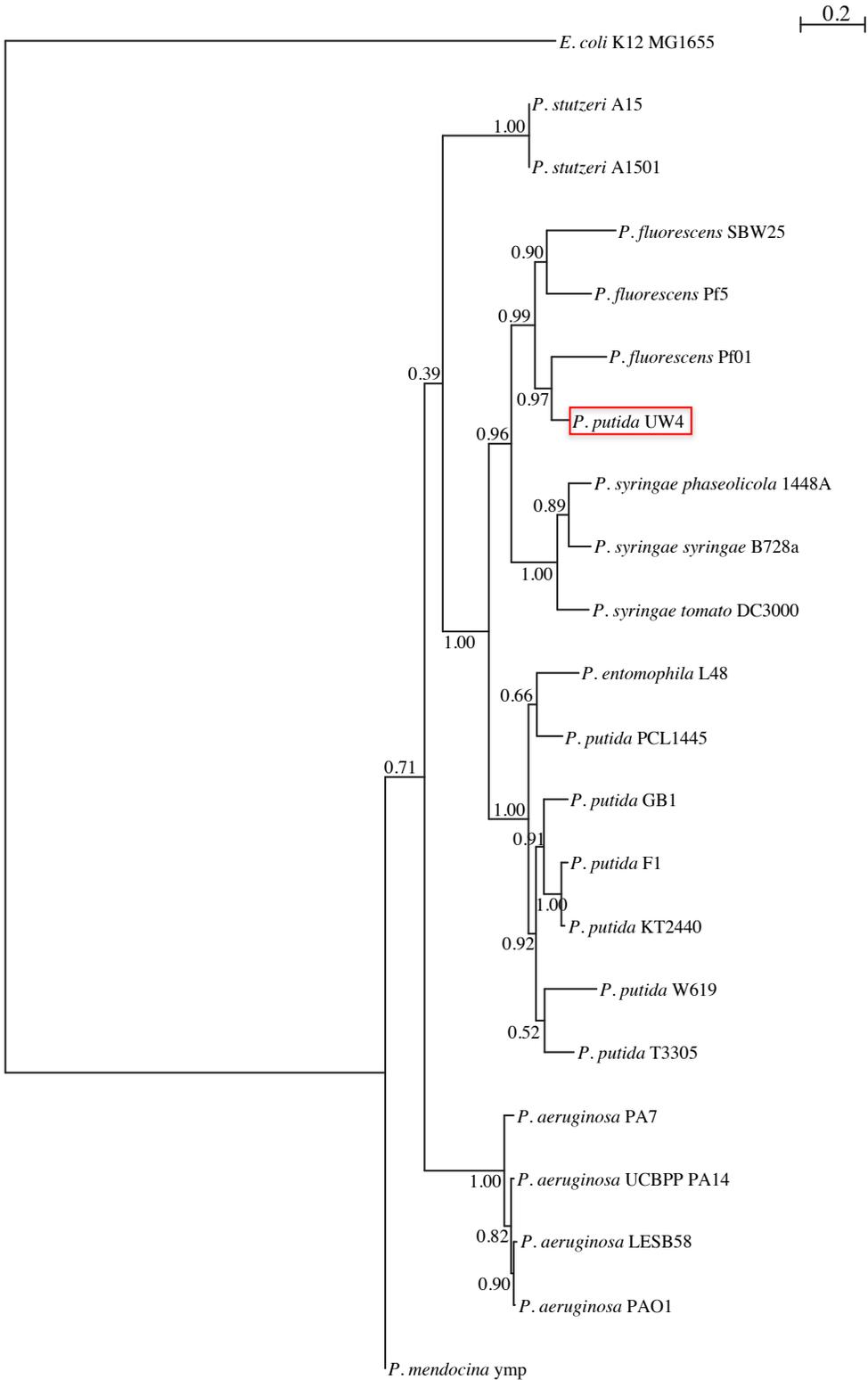
In every HSP tree constructed, *P. putida* UW4 was grouped with the *fluorescens* clade rather than the *putida* clade, which agrees with the whole genome phylogenetic tree.

Fig. 4-32. HSPs ML trees. The phylogenetic analysis of the *Pseudomonas* species was based on a comparison of the nucleotide coding sequences of the A) DnaK, B) DnaJ, C) GrpE, D) GroEL, E) GroES, F) ClpB and G) IbpA. *Escherichai coli* K12 MG1655 orthologous were used as outgroups. Nodal support was evaluated by aLRT.

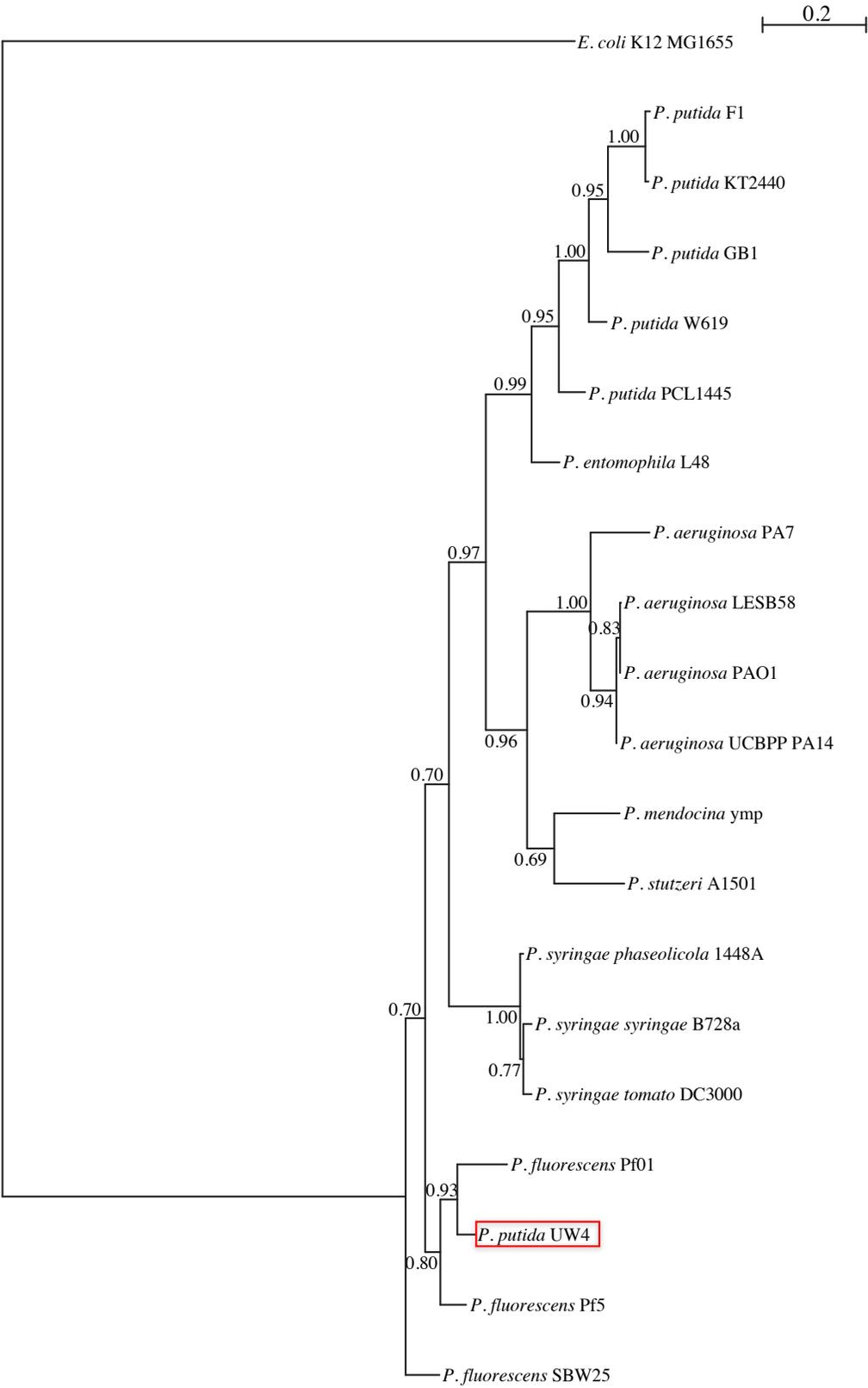
A. DnaK

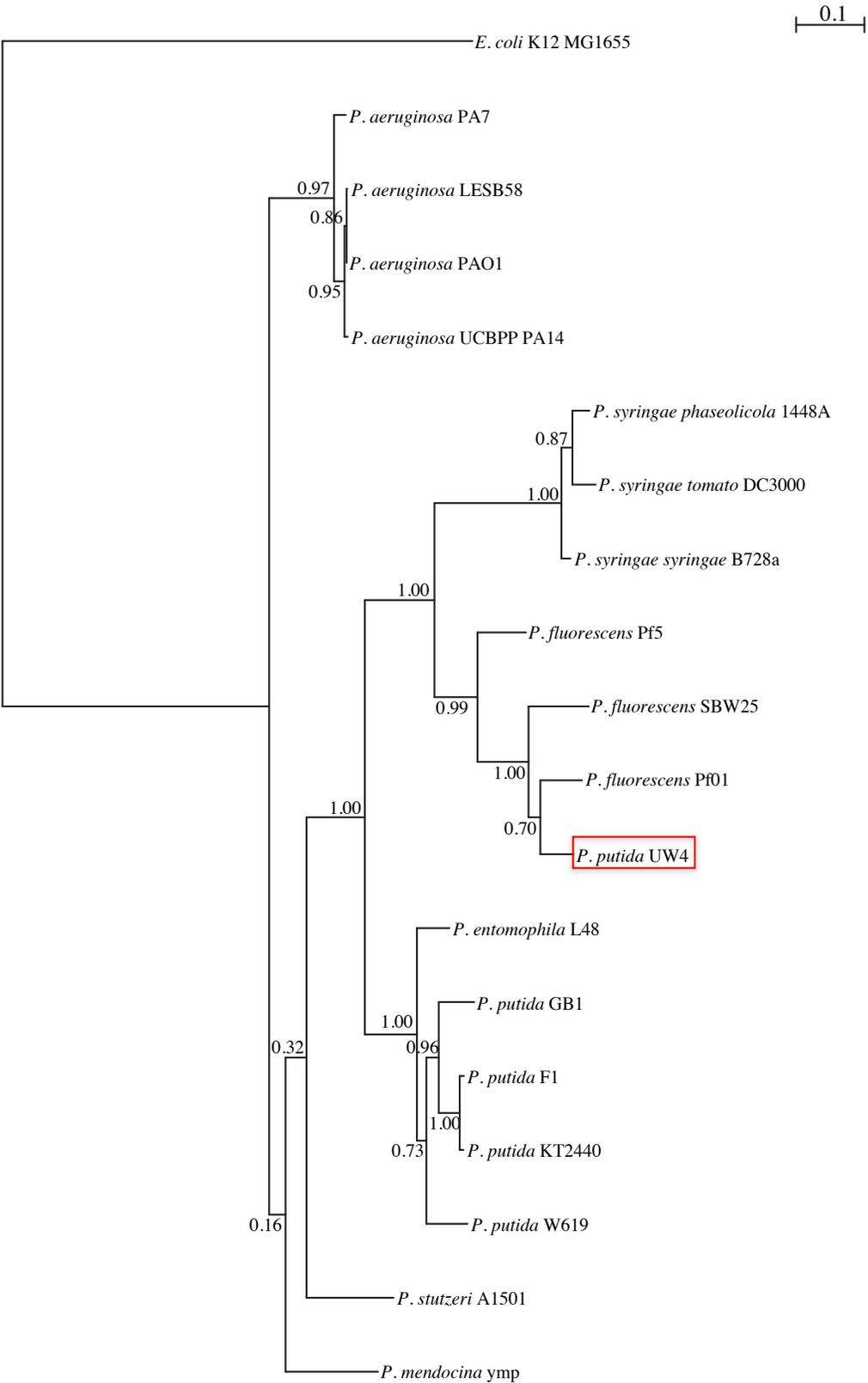


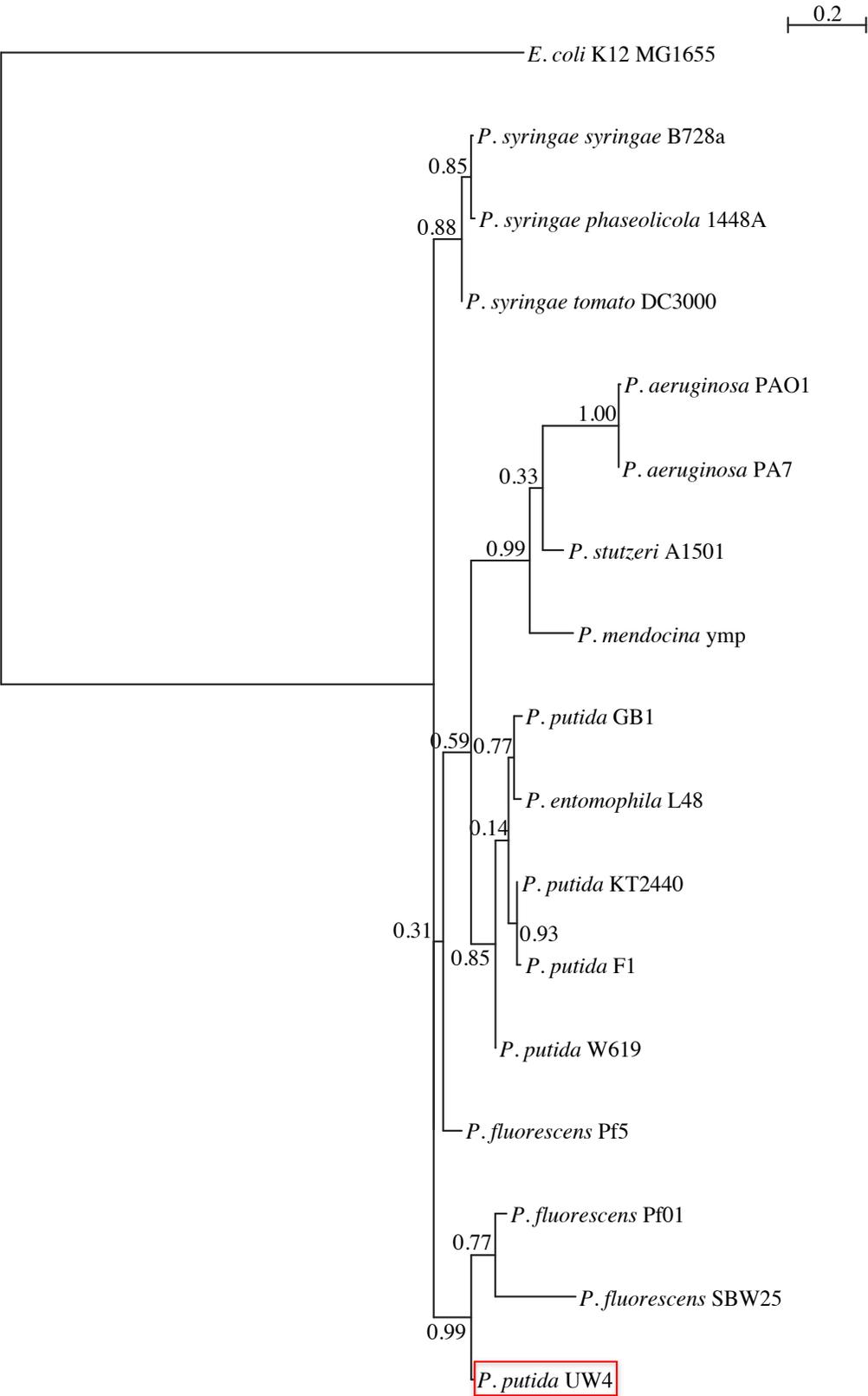
B. DnaJ



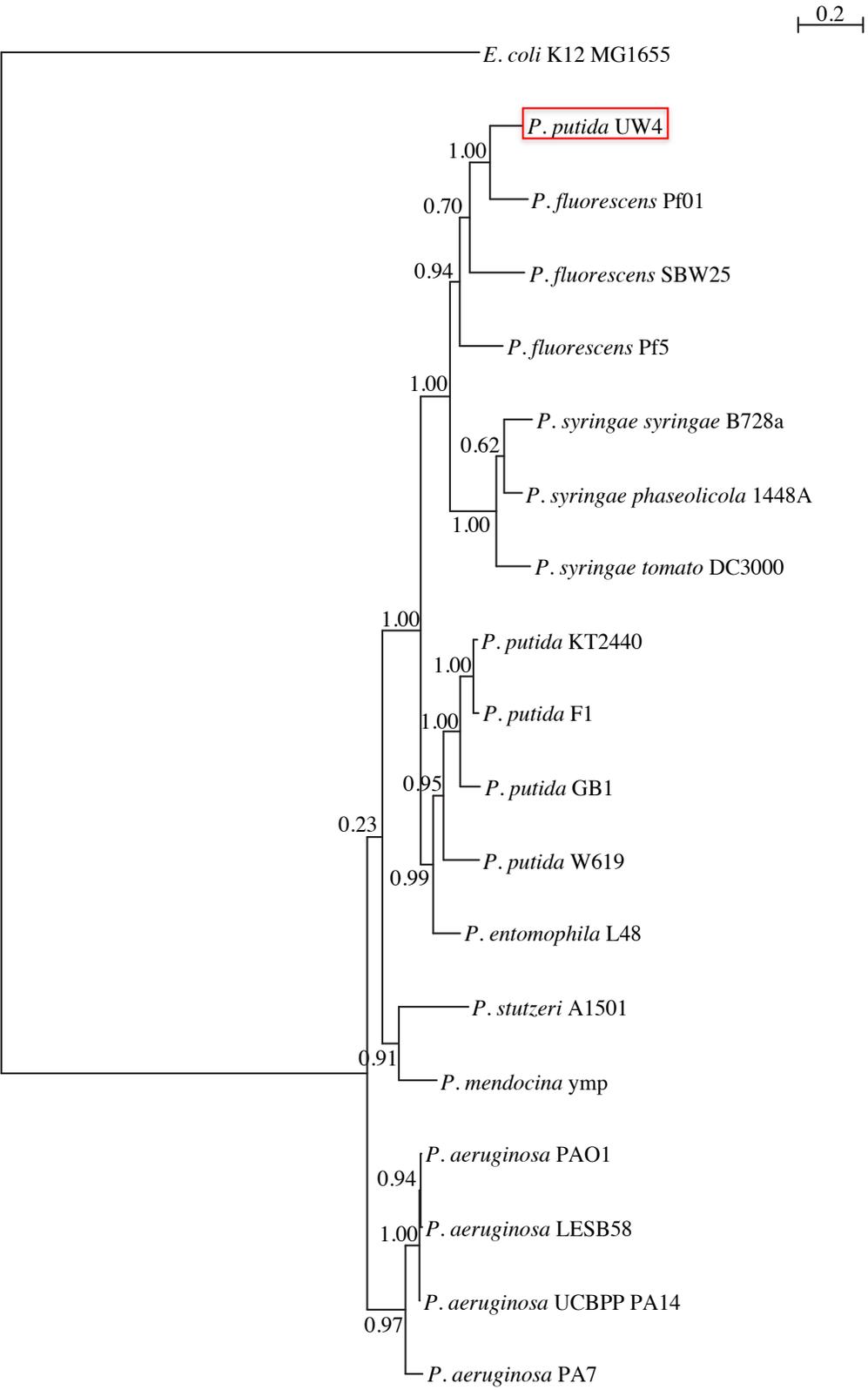
C. GrpE



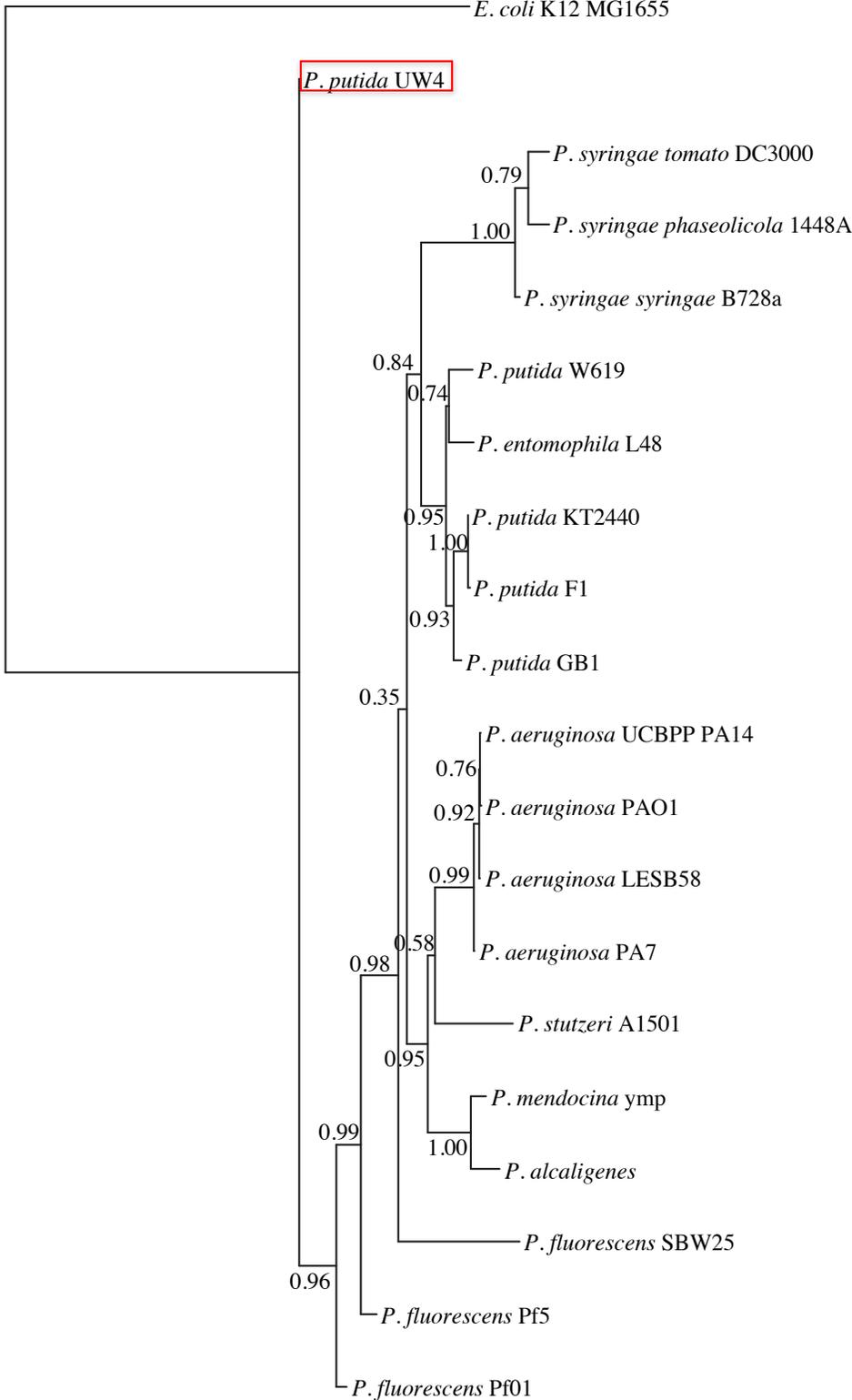




F. ClpB



0.2



The protein,  $\sigma^{32}$  is bacterial heat shock sigma factor that can turn on the transcription of many genes in response to heat shock stress. The alignment of the  $\sigma^{32}$  promoter sequences of HSP genes including *dnaK* operon (*dnaK*, *dnaJ* and *grpE*), the *groE* operon (*groEL* and *groES*), *clpB* and *ibpA* are shown in Fig. 4-33. The promoter regions were determined based on the sequence similarity with the consensus sequence from *E. coli* (Nonaka et al. 2006). In all cases, the -35 region is more conserved than the -10 region.

Fig. 4-33.  $\sigma^{32}$  promoter sequence alignment of HSPs. A. *dnaK* operon, B. *groE* operon, C. *clpB*, D. *ibpA*. Dots indicate the nucleotides that are identical to the promoter sequences of *P. putida* UW4. Gaps are indicated by dashes. The consensus sequences of -35 and -10 regions are shown above the alignment. PpUW4: *Pseudomonas putida* UW4; PaLESB58: *Pseudomonas aeruginosa* LESB58; PaPAO1: *Pseudomonas aeruginosa* PAO1; PaUCBPPPA14: *Pseudomonas aeruginosa* UCBPP-PA14; PaPA7: *Pseudomonas aeruginosa* PA7; PsA1501: *Pseudomonas stutzeri* A1501; Pmymp: *Pseudomonas mendocina* ymp; PpKT2440: *Pseudomonas putida* KT2440; PpF1: *Pseudomonas putida* F1; PpGB1: *Pseudomonas putida* GB-1; PpW619: *Pseudomonas putida* W619; PpPCL1445: *Pseudomonas putida* PCL1445; PeL48: *Pseudomonas entomophila* L48; PfPf01: *Pseudomonas fluorescens* Pf0-1; PfPf5: *Pseudomonas fluorescens* Pf-5; PfSBW25: *Pseudomonas fluorescens* SBW25; Ps1448A: *Pseudomonas syringae* pv. *phaseolicola* 1448A; PsB728a: *Pseudomonas syringae* pv. *syringae* B728a; PsDC3000: *Pseudomonas syringae* pv. *tomato* DC3000; EcMG1655: *Escherichia coli* str. K-12 substr. MG1655

### A. DnaK Operon

	CTTGAA	RNCCCCATNT
PpUW4	CTTGAA	GCCGGATTTTGC
PaLESB58	C . GT . GAAAGC .	T . . . . .
PaPAO1	C . GT . GAAAGC .	T . . . . .
PaUCBPPPA14	C . GT . GAAAGC .	T . . . . .
PaPA7	. GT . CGAACC .	A . . . . .
PsaA1501	. . . . .	T . C . C . . . . .
Pmymp	AG . T . TGGAATT .	. . . . . T . . . . .
PpKT2440	C . T . T . . . . .	C . . . . .
PpF1	C . T . T . . . . .	C . . . . .
PpGB1	. . . . .	T . T . . . . CA . . . . .
PpW619	. . . . .	T . AT . . . . . C . . . . .
PpPCL1445	. . . . .	T . T . . . . . C . . . . .
PeL48	. . . . .	C . T . T . . . . . C . . . . .
PfPf01	. . . . .	T . . . . .
PfPf5	C . TT . AAA . . . . .	A . . . . .
PfSBW25	C . T . . . . .	AA . . . . . T . . . . .
Ps1448A	. . . . .	AA . C . . . . . AT . . . . .
PsB728a	. . . . .	AAA . C . . . . . AT . . . . .
PsDC3000	. . . . .	G . AA . C . . . . . AT . . . . .
EcMG1655	. . . . . T . A . TGG . A .	. . . . . A . . . . . T . . . . .

### B. GroE Operon

	CTTGAA	RNCCCCATNT
PpUW4	CTTGAA	ATAAG - CTTATGC
PaLESB58	. . . . . A . . . . .	GGC - T . GTG . . . . . A . . . . .
PaPAO1	. . . . . A . . . . .	GGC - T . GTG . . . . . A . . . . .
PaUCBPPPA14	. . . . . A . . . . .	GGC - T . GTG . . . . . A . . . . .
PaPA7	. . . . . A . . . . .	GGC - T . GTGA . . . . . A . . . . .
PsaA1501	. . . . . A . . . . .	GCT . . . . . GAA . . . . . GC . . . . .
Pmymp	. . . . . A . . . . .	GCC . . . . . GCC . . . . . AC . . . . .
PpKT2440	. . . . .	CG . - A . GTG . . . . . G . . . . . C . . . . .
PpF1	. . . . .	TG . - A . GTG . . . . . G . . . . . C . . . . .
PpGB1	. . . . .	TGA - T . ATG . . . . . G . . . . . C . . . . .
PpW619	. . . . .	TGA - TCA TGC . . . . . G . . . . . C . . . . .
PeL48	. . . . .	TGA - T . CACT . . . . . G . . . . . C . . . . .
PfPf01	. . . . .	. . . . .
PfPf5	. . . . .	. . . . . C . . . . .
PfSBW25	. . . . .	. . . . .
Ps1448A	. . . . .	. . . . . C . C - T . . . . . GCT . . . . .
PsB728a	. . . . .	. . . . . C . C - T . . . . . GCT . . . . .
PsDC3000	. . . . .	. . . . . C . A - A . . . . . GCC . . . . .
EcMG1655	. . . . . A . . . . .	GGGGCGAAGCCT . . . . . AT . . . . . CC . . . . . T . . . . .

### C. ClpB

	CTTGAA	RNCCCCATNT
PpUW4	CTTGAA	ACTCCCAGAATC
PaLESB58	C . G . GG . A . . . . .	C . . . . . TT . . . . .
PaPAO1	C . G . GG . A . . . . .	C . . . . . TT . . . . .
PaUCBPPPA14	C . G . GG . A . . . . .	C . . . . . TT . . . . .
PaPA7	C . G . GG . A . . . . .	C . . . . . TT . . . . .
PsaA1501	. . . . .	T . . . . . A . C . . . . . AT . . . . . T . . . . .
Pmymp	. . . . .	TACGG . GT . . . . .
PpKT2440	. . . . .	T . . . . .
PpF1	. . . . .	T . . . . .
PpGB1	. . . . .	T . . . . .
PpW619	. . . . .	. . . . . CT . . . . .
PeL48	. . . . .	T . . . . .
PfPf01	. . . . .	T . . . . .
PfPf5	. . . . .	T . . . . .
PfSBW25	. . . . .	T . . . . . TG . . . . .
Ps1448A	. . . . .	. . . . .
PsB728a	. . . . .	. . . . .
PsDC3000	. . . . .	. . . . .
EcMG1655	. . . . .	. . . . . TAATTG . . . . . GGAT . . . . .

### D. IbpA

	CTTGAA	RNCCCCATNT
PpUW4	CTTGAA	ATGCAATCAGCG
PaLESB58	. . . . .	A . TTC . GA . . . . . A . . . . .
PaPAO1	. . . . .	A . TTC . GA . . . . . A . . . . .
PaUCBPPPA14	. . . . .	A . TTC . GAT . . . . . A . . . . .
PaPA7	. . . . .	ATTC . C . . . . . TAC . . . . .
PsaA1501	. . . . .	C . GT . TTT . . . . . T . . . . .
Pmymp	. . . . .	ATTCTC . C . T . A . . . . . TC . GT . . . . .
PaIcali	. . . . .	TGCG . GCGTGCGA . . . . . CC . T . AAC . . . . .
PpKT2440	. . . . .	GCATT . CCTC . . . . . G . . . . . A . . . . .
PpF1	. . . . .	CCGT . CACCGTT . C . CTG . C . C . G . . . . .
PpGB1	. . . . .	GCATT . CCTC . . . . . G . . . . . A . . . . .
PpW619	. . . . .	GCT . T . TCCT . . . . . AG . . . . . T . . . . .
PeL48	. . . . .	GC . CG . GGC . C . . . . . G . . . . . A . . . . .
PfPf01	. . . . .	ACA . . . . . A . GCA . . . . .
PfPf5	. . . . .	C . . . . . TGGAGC . CA . . . . .
PfSBW25	. . . . .	TTTTTC . GC . T . AA . . . . . CT . . . . .
Ps1448A	. . . . .	A . TG . ATG . GTA . . . . . T . . . . .
PsB728a	. . . . .	A . G . AAG . T . . . . . T . . . . .
PsDC3000	. . . . .	A . . . . . A . G . C . . . . . T . . . . .
EcMG1655	. . . . .	A . TTCATTT . CAGA . . . . . C . . . . . T . . . . .

## 5 Discussion

### 5.1 Taxonomy of *P. putida* UW4: *putida* or *fluorescens*

Whole genome phylogenetic analysis, HSPs phylogeny and the HSPs  $\sigma^{32}$  promoter analysis suggested that UW4 is closer to *fluorescens* than to *putida*. However, 16S rRNA gene phylogeny of completely sequenced *Pseudomonas* genomes showed that UW4 is grouped with the *putida* clade, albeit with low confidence. Additional analysis of the 16S rRNA gene phylogeny including the type strains of *P. fluorescens* and *P. putida* as well as the unique 16S sequences of KT2440 and SBW25 showed that UW4 has a closer relationship with *putida* than *fluorescens*. Nevertheless, the diversities of UW4 and *putida* 16S rRNA gene sequences are above the common threshold, 1.3%. These results raise the question whether UW4 belongs to *putida* or *fluorescens*.

*Pseudomonas* sp. is one of the most diverse and prevalent genera that are present in all natural environments. Since its first discovery by Migula in 1894, the taxonomy of *Pseudomonas* has always been controversial (Peix et al. 2009). The initial classification of *Pseudomonas* in the 1920s contained only very limited phenotypic characteristics including Gram-negative, aerobic non-sporulated rods that are motile through polar flagella, and did not show a clear differentiation from other Gram-negative bacteria (Peix et al. 2009). In 1974, genetic information such as G+C content was first added in Bergey's Manual to assist bacterial classification. Meanwhile, another genotypic criterion based on RNA-DNA relatedness was used to classify *Pseudomonas* into five rRNA subgroups, and only the strains in group I were kept in genus *Pseudomonas* (Palleroni et al. 1973; Peix et al. 2009). In 1984, a new bacterial identification scheme based on 16S ribosomal RNA was proposed by Woese and collaborators (Woese et al. 1984). Since then, sequencing of 16S rRNA gene has become a routine method to identify bacteria, mainly because its

evolutionary rate is high enough to differentiate different species, and also there is sufficient sequence conservation within the same species. Furthermore, with the advancement of mathematical models for construction of trees, phylogenetic classification of prokaryotes can be readily achieved. As bacterial taxonomy has progressed, many *Pseudomonas* sp. have been reclassified as other species and/or genera through the years (Johnson and Palleroni, 1989; Willems et al., 1990, 1992; Yabuuchi et al. 1992, 1995; Palleroni and Bradbury, 1993; Segers et al., 1994; Grimes et al. 1997; Denner et al. 1999; Anzai et al. 2000; Brown et al. 2001; Coenye et al. 2001; Satomi et al. 2002; Peçonek et al. 2006; Peix et al., 2007; Kämpfer et al. 2008).

Although 16S rRNA genes are the basis of the current bacterial classification, it is known that very closely related species of bacteria cannot be differentiated based on this gene (Fox et al. 1992; Lechner et al. 1998; Wink et al., 2003; Valverde et al., 2006b; Dutta and Gachhui, 2007; Rivas et al., 2007; Zurdo-Piñeiro et al., 2007). Therefore, many studies have shown that other genes, such as “housekeeping” genes *recA*, *atpD*, *carA*, *gyrB*, *rpoB*, *trpB*, should be used to assist bacterial species classification (Hilario et al., 2004; Maiden 2006; Peix et al. 2007; Guo et al. 2008). Furthermore, the fact that most bacteria have multiple copies of 16S rRNA genes and their intragenomic diversities within individual genomes indicate that it is necessary to include all unique 16S rRNA genes of one bacterium for its identification. However, without knowing the complete genome sequence of the bacterium, one can hardly obtain all the sequences of its 16S rRNA genes. Thanks to the continually improved sequencing technologies, more and more complete bacterial genome sequences will become available, which will greatly facilitate 16S rRNA gene based bacterial taxonomy. Another asset of genome sequencing is that it allows whole genome phylogenetic analysis among the species of the same genus, which will help determine the core and pan-genome and provide valuable information to aid bacterial classification.

Since the resolution of 16S rRNA tree was not sufficient to differentiate UW4 from other closely related *Pseudomonas* species, the classification of this bacterium should follow the whole genome phylogeny based on the conserved genes among all sequenced *Pseudomonas* genomes, which indicated that it belongs to *P. fluorescens*. Furthermore, according to the Bergey's Manual of Determinative Bacteriology (Holt, 1994), *P. fluorescens* is positive for nitrate reduction, whereas *P. putida* is negative. In the genome of UW4, the presence of a putative nitrate reductase (PputUW4\_3649) supports the reclassification of UW4 into *fluorescens*. However, experimental evidence is necessary to confirm the validity of this classification.

## 5.2 Comparative Genomics of UW4 with Other *Pseudomonas*

At the time of this writing, 20 complete sequenced *Pseudomonas* genomes were available in the NCBI genome database and their general features are shown in Table 5-1.

Among the 20 genomes, four belong to *aeruginosa*, which is one of the major opportunistic human pathogens. *P. aeruginosa* PAO1 was the first sequenced *Pseudomonas* genome, and was originally isolated from a wound (Stover et al. 2000). PA14 is a human clinical isolate from a burn patient and has been the cause of disease in various hosts (Lee et al. 2006). LESB58 represents Liverpool Epidemic Strain B58 and was obtained from a cystic fibrosis patient in the United Kingdom (Winstanley et al. 2009). The last one, PA7 was isolated from a non-respiratory patient in Argentina (Roy et al. 2010). *P. brassicacearum* NFM421 is a PGPB that is associated with the roots of *A. thaliana* and *Brassica napus* (Ortet et al. 2011). It can be used as a biocontrol strain because of the production of antifungal compounds. *P. entomophila* L48 is an entomopathogenic bacterium that was isolated from a fruit fly of Guadeloupe and it kills insects upon ingestion (Vodovar et al. 2006). *P. fluorescens* strains are well known for their physiological diversities that they can colonize various environments such as soil, water and plant

Table 5-1. General Features of the *Pseudomonas* Genomes.

<i>Pseudomonas</i> sp.	Genome Size, bp	CDS no.	Pseudogenes	G+C content, %	Coding density, %	tRNA no.	rRNA genes (operon no.)	Plasmid no.	Ref.
UW4	6,183,388	5,430	18	60.92	87.4	72	22 (7)	—	This study
<i>aeruginosa</i> LESB58	6,601,757	5,925	34	66.3	88.4	67	13 (4)	—	Winstanley et al. 2009
<i>aeruginosa</i> PA7	6,588,339	6,286	8	66.5	89.5	63	12 (4)	—	Roy et al. 2010
<i>aeruginosa</i> PAO1	6,264,404	5,566	5	66.6	89.3	63	13 (4)	—	Stover et al. 2000
<i>aeruginosa</i> UCBPP-PA14	6,537,648	5,892	none	66.3	89.4	59	13 (4)	—	Lee et al. 2006
<i>brassicacearum</i> subsp. <i>brassicacearum</i> NFM421	6,843,248	6,097	N/A	60.8	88.2	65	16 (5)	—	Ortet et al. 2011
<i>entomophila</i> L48	5,888,780	5,169	N/A	64.2	89.1	78	22 (7)	—	Vodovar et al. 2006
<i>fluorescens</i> Pf-5	7,074,893	6,144	N/A	63.3	88.8	71	16 (5)	—	Paulsen et al. 2005
<i>fluorescens</i> Pf0-1	6,438,405	5,741	9	60.6	89.8	73	19 (6)	—	Silby et al. 2009
<i>fluorescens</i> SBW25	6,722,539	6,009	88	60.5	88.3	66	16 (5)	—	Silby et al. 2009
<i>mendocina</i> NK-01	5,434,353	4,958	N/A	62.5	88.7	65	12 (4)	—	Guo et al. 2011
<i>putida</i> BIRD-1	5,731,541	5,124	N/A	61.7	86.9	64	22 (7)	—	Matilla et al. 2011
<i>putida</i> F1	5,959,964	5,300	49	61.9	88.7	76	19 (6)	—	Wu et al. 2010
<i>putida</i> GB-1	6,078,430	5,417	8	61.9	89.4	74	22 (7)	—	Wu et al. 2010
<i>putida</i> KT2440	6,181,863	5,420	N/A	61.6	86.7	73	22 (7)	—	Nelson et al. 2002
<i>putida</i> S16	5,984,790	5,218	N/A	62.3	84.9	70	19 (6)	—	Yu et al. 2011
<i>putida</i> W619	5,774,330	5,471	26	61.4	88.9	75	22 (7)	—	Wu et al. 2010
<i>stutzeri</i> A1501	4,567,418	4,146	N/A	63.8	89.8	61	12 (4)	—	Yan et al. 2008
<i>syringae</i> pv. <i>phaseolicola</i> 1448A	5,928,787	5,144	N/A	58	87	64	16 (5)	2	Joardar et al. 2005
<i>syringae</i> pv. <i>syringae</i> B728a	6,093,698	5,137	47	59.2	88.5	64	16 (5)	—	Feil et al. 2005
<i>syringae</i> pv. <i>tomato</i> str. DC3000	6,397,126	5,615	N/A	58.4	86.8	63	15 (5)	2	Buell et al. 2003

surface. Strain Pf-5 is a soil bacterium that possesses biocontrol capabilities such as the production of broad spectrum of antibiotics and secondary metabolite (Paulsen et al. 2005). SBW25 is a plant beneficial bacterium that was obtained from the leaf surface of a sugar beet plant in the United Kingdom, and Pf0-1 was isolated from loam soil in Sherborn, Massachusetts, USA in 1987 (Silby et al. 2009). *P. mendocina* NK-01 was isolated from farmland soil in Tianjin, China. It produces PHA and alginate oligosaccharides under nitrogen starvation, which made it a candidate for genome sequencing (Guo et al. 2011). Complete genome sequences of six *P. putida* strains are publically available in the genome database. Strain KT2440 is a rhizospheric bacterium isolated from garden soil in Japan. It is certified as a safety strain for cloning and expression of foreign genes for Gram-negative soil bacteria, and is recognized as the best characterized *putida* strain that serves as the workhorse for *Pseudomonas* research (Nelson et al. 2002; Wu et al. 2011). *P. putida* F1 was obtained from a polluted creek in Urbana, IL, USA and can be used in bioremediation owing to its ability to degrade aromatic hydrocarbon compounds such as benzene, toluene, ethylbenzene and p-cymene (Wu et al. 2011). Strain GB-1 is a manganese oxidizer that was isolated from fresh water of Green Bay, WI, USA. It serves as a model organism for molecular genetic studies of Mn<sup>2+</sup> oxidation (Wu et al. 2011). W619 is an endophyte that was isolated from *Populus trichocarpa* × *deltoides* cv. “Hoogvorst” (Wu et al. 2011). It is a PGPB that can improve plant growth by decreasing the activities of antioxidative defence related enzymes such as glutathione reductase and superoxide dismutase, resulting in lowered oxidative stress level. It can also reduce stomatal resistance, leading to increased plant fitness (Weyens et al. 2011). BIRD-1 is a rhizospheric PGPB that is highly tolerant to desiccation, is able to solubilize inorganic phosphate, synthesize siderophores and phytohormones such as IAA (Matilla et al. 2011). S16 is the first completely sequenced nicotine-degrading

microorganism, which can serve as a reference strain in research on biodegradation of N-heterocyclic compound (Yu et al. 2011). *P. stutzeri* A1501 was isolated from rice paddy soils and has been applied in the field as a crop inoculant in China. It has the ability to fix nitrogen, which is not considered to be a common characteristic of *Pseudomonas* (Yan et al. 2008). *P. syringae* are well known plant pathogens that can grow epiphytically on various plants. *P. syringae* pv. *tomato* Strain DC3000 was isolated from tomato grown in the Channel Islands, Guernsey, UK. It is pathogenic on tomato and the model plant *A. thaliana* (Buell et al. 2003). *P. syringae* pv. *syringae* Strain B728a was isolated from a snap bean leaflet in Wisconsin, USA. It differs from DC3000 in terms of host range and it has higher abiotic stress tolerance (Feil et al. 2005). *P. syringae* pv. *phaseolicola* Strain 1448a was isolated from common bean, *Phaseolus vulgaris*, in Ethiopia in 1985, which causes halo blight of bean (Joardar et al. 2005).

Currently, the largest *Pseudomonas* genome is *P. fluorescens* Pf-5 (7 MB), whereas the smallest one is *P. stutzeri* A1501 (4.5 MB), indicating a high degree of physiological and genetic versatility of *Pseudomonas* sp. (Table 5-1). UW4 has a similar genome size (6.18 MB) and number of predicted protein coding genes (5,430) compared with *P. putida* KT2440 (6.18 MB and 5,420, respectively). The number of CDSs usually reflects the size of the genome, with one exception represented by PA7, whose genome size is smaller than Pf-5 but has 142 more predicted CDSs. The number of pseudogenes among the 21 genomes ranges from zero (*P. aeruginosa* UCBPP-PA14) to 88 (*P. fluorescens* SBW25). However, pseudogenes were not analyzed in most sequenced *Pseudomonas* genomes, and only 11 out of 21 has the data available for comparison. The three *P. syringae* strains have the lowest G+C content (58-59.2%), whereas *P. aeruginosa* strains have the highest (66.3-66.6%). In the case of UW4, its G+C content is in between that of the *putida* strains and two *fluorescens* strains, SBW25 and Pf0-1. The coding densities of *Pseudomonas* genomes range from 84.9% (*putida* S16) to 89.8% (*stutzeri* A1501),

and UW4 has slightly lower value (87.4%) than that of *brassicacearum* NFM421 (88.2%) and higher value than that of *syringae* 1448A (87%). The number of *rrn* operons in *Pseudomonas* genomes range from 4 to 7, with most of them comprising an additional 5S rRNA. Similar to most *putida* strains, UW4 has 7 *rrn* operons and an additional 5S rRNA. The presence of plasmids has been documented in only two sequenced *Pseudomonas* genomes, *P.syringae* strain DC3000 and 1448A. The plasmids of the two pathovars are significantly different in size and gene content. For example, genes involved in virulence located on p1448A-A and -B are present on the chromosome of DC3000, whereas three plasmid borne virulence factors of DC3000 are found on the 1448A chromosome (Feil et al. 2005).

Whole genome alignment among the 21 *Pseudomonas* sp. identified 1679 conserved genes, which makes the core genome of *Pseudomonas*. The putative orthologous shared between UW4 and other *Pseudomonas* are shown in Table 4-16, with the two *P. fluorescens* strains, Pf0-1 and Pf-5 being the top two. The rest of the CDSs including those shared with (1) one or several but not all completely sequenced *Pseudomonas* genomes; (2) other *Pseudomonas* sp. whose genomes are not available; (3) CDSs present only in other genera, and (4) the unique genes in UW4, will contribute to the pan genome of *Pseudomonas*.

Since UW4 is a well-studied PGPB, the following discussion will focus on genomic comparisons between UW4 and other *Pseudomonas* sp. with respect to plant growth-promotion. The discussion of the two features, antibiotic resistance and heavy metal resistance, have been combined with the results, thus they won't be included here.

### **5.2.1 ACC deaminase**

ACC deaminase gene is present in five *Pseudomonas* genomes including UW4, *P. brassicacearum* NFM421, *P. syringae* DC3000, *P. syringae* B728a, and *P. syringae* 1448A.

Pairwise amino acid sequence identities between UW4 *acdS* and the other four genomes range from 89% to 99%, and they all contain the important active sites (Glick et al. 2007b), suggesting that the putative *acdS* gene in those genomes is likely functional. Furthermore, the common *acdS* regulatory gene, *acdR*, was found immediately upstream of *acdS* in all five genomes, and the amino acid sequence identities between UW4 *acdR* and the other four genomes range from 80% to 93%. This type of *acdS* regulation scheme has been observed in many bacteria and was proposed as a main feature of the functioning of bacterial ACC deaminase (Glick et al. 2007a).

### 5.2.2 Siderophores

Siderophore production is a typical characteristic possessed by fluorescent pseudomonads. Among the 21 *Pseudomonas* genomes, only *P. stutzeri* A1501 does not have the genes for siderophore biosynthesis, and it also has the smallest genome compared to the other 20 species, suggesting loss of functions in A1501 (Yan et al. 2008). Compared with UW4, 11 genomes contain a gene encoding PvdYII including *P. putida* KT2440 (locus\_tag: PP\_4245), *P. putida* BIRD-1 (PPUBIRD1\_1611), *P. putida* F1 (Pput\_1682), *P. putida* GB-1 (PputGB1\_3811), *P. putida* W619 (PputW619\_3564), *P. putida* S16 (PPS\_3636), *P. brassicacearum* NFM421 (PSEBR\_a1665), *P. entomophila* L48 (PSEEN\_1813), *P. fluorescens* Pf0-1 (Pfl01\_3942), *P. mendocina* NK-01 (MDS\_1799), *P. aeruginosa* PA7 (PSPA7\_2826), indicating these species likely produce type II pyoverdine since this gene was only observed in the strains of *P. aeruginosa* that make type II pyoverdine (Smith et al. 2005; Lamont et al. 2006). However, the precise structure of the siderophore needs to be confirmed experimentally.

In *P. fluorescens* Pf-5, genes responsible for pyoverdine as well as pyochelin were identified. The genes required for pyoverdine synthesis are located in three clusters; whereas genes necessary for pyochelin synthesis are present in a single cluster (Paulsen et al. 2005). In *P.*

*entomophila* L48, two gene clusters of pyoverdine synthesis are present, which exhibits similar organization compared with that found in other fluorescent pseudomonads. In addition, one gene cluster related to acinetobactin was observed on the chromosome, and it contains a salicylamide moiety (Vodovar et al. 2006). *P. syringae* DC3000 produces two types of siderophores, pyoverdine and yersiniabactin, and in both cases the required genes are present in a single cluster (Buell et al. 2003). *P. syringae* B728a also secretes two types of siderophores. The first type is pyoverdine, and as in DC3000, the determinants are located in one gene cluster. The second type is achromobactin, which is a citrate siderophore produced by *Pectobacterium chrysanthemi* and *Escherichia carotovora* pv. *atroceptica* (Feil et al. 2005). The ability of bacteria to produce multiple siderophores surely benefits these organisms, as they may function in different environments, making them more competitive against other organisms in the same niche.

### 5.2.3 IAA production

Although many bacteria are able to synthesize IAA, the amounts produced vary significantly between strains. Depending on the concentration, bacterially produced IAA can either stimulate or inhibit plant growth. In UW4, two potential tryptophan-dependent IAA biosynthesis pathways, indole-3-acetamide (IAM) and indole-3-acetonitrile (IAN), were identified, and 7 genes might be involved. When searching those genes against the other 20 *Pseudomonas* genomes, 7 orthologous genes were found in one PGPB, *P. fluorescens* SBW25, suggesting similar IAA synthesis pathways compared to UW4. *P. putida* BIRD-1 has 6 homologs that complete the IAM and IAN pathways, but it lacks the gene encoding the nitrilase (PputUW4\_2466). *P. putida* F1 lacks the homolog for one of the tryptophan 2-monooxygenases (PputUW4\_4535), but it contains all of the other 6 genes. It is possible that one of the tryptophan

2-monooxygenases is dominant over the other, but currently it is not known which one plays a more important role in IAA biosynthesis in these strains.

Many studies have shown that numerous bacterial strains possess multiple IAA synthesis pathways. Besides the above mentioned strains, it has been observed that putative TAM and IAM pathways are present in *P. putida* W619, GB-1, and F1 (Wu et al. 2011). Therefore, to study the role of each gene in IAA biosynthesis of a particular bacterium it is necessary to construct a large number of mutants, single or multiple, and test the functioning of each one. Currently, mutational analyses are ongoing for UW4 IAA biosynthesis genes.

#### **5.2.4 Trehalose**

It has been reported that trehalose can protect bacterial cells from environmental stresses such as desiccation, high salinity, freezing, and heat (Freeman et al. 2010). In bacteria, five trehalose biosynthetic pathways are known including OtsA/OtsB, TreS, TreY/TreZ, TreP, and TreT (Paul et al. 2008). In UW4, two trehalose synthesis pathways, TreS and TreY/TreZ, were identified. When searching the orthologous in the other *Pseudomonas* genomes, all 20 species contain the genes involved in those two trehalose synthesis pathways, and they are organized in a similar way, indicating the ubiquity and importance of this sugar. In addition, *P. stutzeri* A1501 has a third trehalose synthesis pathway, OtsA/OtsB, which is the most widespread pathway present in both eukaryotes and prokaryotes, and this may further contribute to its survival under different environmental stresses.

#### **5.2.5 Acetoin**

Bacterial volatile compounds such as acetoin can stimulate the growth of plants such as *A. thaliana* by increasing the total leaf surface area and reducing the disease symptoms triggered

by pathogenic bacterium, *Erwinia carotovora* subsp. *carotovora* (Ryu et al. 2003, 2004). In the UW4 genome, a potential acetoin biosynthesis pathway was identified, and three enzymes are involved. First, pyruvate is converted to 2-acetolactate by acetolactate synthase (PputUW4\_4617, 4618). Because the intermediate, 2-acetolactate, is unstable, it undergoes a spontaneous decarboxylation in the presence of O<sub>2</sub>, resulting in the production of diacetyl. Next, diacetyl is converted to acetoin by diacetyl reductase/acetoin dehydrogenase (PputUW4\_3051).

When the genomes of the other *Pseudomonas* were examined for acetoin synthesis, the same pathway was observed in all 20 species, although the enzyme that catalyzes the last step couldn't be determined definitely due to ambiguous annotations. Since the genes identified in acetoin biosynthesis pathway were predicted based on sequence similarity only, experiments need to be conducted to verify the production of acetoin by UW4.

### 5.2.6 Antimicrobial compounds

It has been demonstrated that fungal elicitors such as *Pythium aphanidermatum* and *Fusarium oxysporum* can induce the production of 4-hydroxybenzoate in carrot cell cultures and alfalfa plants, respectively (Schnitzler et al. 1992; Cvrikova et al. 1993). In bacteria, 4-hydroxybenzoate is formed from chorismate directly by chorismate lyase encoded by *ubiC*. When searching the gene *ubiC* in *Pseudomonas* genomes, it was found in all 21 species including UW4, suggesting 4-hydroxybenzoate synthesis is a common pathway in the genus of *Pseudomonas*.

HCN is another antimicrobial agent that confers biocontrol ability onto some PGPB (Blumer and Haas 2000; Haas and Défago 2005). In the genome of UW4, the gene cluster *hcnABC*, which is responsible for HCN synthesis, is absent. When the three genes were searched against the genomes of the other *Pseudomonas*, it was found that they are present in 8 of the 21 strains including *P. fluorescens* Pf-5, *P. fluorescens* Pf0-1, *P. brassicacearum* NFM421, *P.*

*entomophila* L48, and the four *P. aeruginosa* strains LESB58, PAO1, PA7 and UCBPP-PA14. Since HCN can be produced by both PGPB and pathogens, other plant growth-promoting traits should also be considered when trying to isolate PGPB from the environments.

### 5.2.7 PHAs biosynthesis

The genes involved in PHA synthesis are found in many *Pseudomonas* sp. such as *P. putida* KT2440, *P. putida* GPo1, *P. aeruginosa* PAO1, *P. fluorescens* Pf0-1, *P. fluorescens* Pf-5, *P. syringae* pv *phaseolicola*, and *P. syringae* DC3000 (Prieto et al. 2007). The gene cluster typically contains six genes including *phaC1*, *phaZ*, *phaC2*, *phaD*, *phaF*, and *phaI*. The order of the six genes is highly conserved (*phaC1ZC2DFI*) in the above mentioned strains, and was also observed in UW4 (PputUW4\_0335-0330). The gene *phaC* encodes the key enzyme, PHA synthase or PHA polymerase, for the biosynthesis of PHA. The PhaC1 and PhaC2 belong to the class II PHA synthases that preferentially use 3-hydroxyalkanoates consisting of 6-14 carbons as substrates, and the class II PHA synthases are primarily found in *Pseudomonas* sp. The *phaZ* gene encodes a depolymerase that is responsible for PHA degradation. The gene product of *phaD* is a transcriptional regulator that positively regulates the expression of the downstream genes, *phaI* and *phaF*, which code for phasins (Prieto et al. 2007). When this *pha* gene cluster was searched against the other *Pseudomonas* genomes, orthologs were found to be absent in *P. syringae* pv. *syringae* B728a. In addition, the genome of *P. stutzeri* A1501 contains a gene cluster different from *phaC1ZC2DFI*, designated *phaCABR* that is responsible for poly-hydroxybutyrate (PHB) synthesis (Yan et al. 2008). In the genome of UW4, a second *phaC1* gene was identified (PputUW4\_2305). Compared with the *phaC1* in the *pha* gene cluster, the second *phaC1* showed 69% identities and 83% similarities. It is likely that the redundant *phaC1* gene also contributes to the production of PHA in UW4, however this has to be confirmed experimentally.

### 5.2.8 Degradation of aromatic compounds

In polluted environments, *P. putida* strains are often isolated as predominant microorganisms and are therefore commonly used in bioremediation. Aromatic compounds are among the most abundant and recalcitrant pollutants in the soil and their degradation by bacteria usually involves ring-cleavage in the presence of O<sub>2</sub> by oxygenase (Fuchs et al. 2011). For example, the toluene degradation pathway in *P. putida* F1 is composed of the toluene dioxygenase operon *todABC1C2DE* (Zylstra and Gibson, 1989). However, this toluene degradation pathway is absent in all the other 20 *Pseudomonas* sp., including UW4. In the genome of *P. putida* W619, the genes involved in 3-HPP were identified previously (Wu et al. 2011). Nevertheless, this pathway seems unique in this strain because in the other 20 *Pseudomonas* genomes, it is either absent or incomplete, such as in *P. putida* F1 (Wu et al. 2011) and UW4. In the genome of UW4, catechol and protocatechuate branches of the  $\beta$ -ketoadipate pathway are present. Since this pathway is considered to be one of the central pathways for the catabolism of aromatic compounds in *Pseudomonas* sp., its presence is ubiquitous in the completely sequenced *Pseudomonas* genomes.

### 5.3 Type III secretion system in non-pathogenic *Pseudomonas*

T3SS have been known for some time to be expressed by Gram-negative pathogens to deliver virulence effectors into host cells. Members of these bacteria include *Salmonella* sp., enteropathogenic *E. coli*, *Yersinia* sp., *Shigella* sp., *Erwinia carotovora*, *P. aeruginosa*, and *P. syringae* (Blocker et al. 2003; Galán and Wolf-Watz 2006; Cornelis 2006).

Surprisingly, a putative T3SS was found in UW, which has been recognized as a PGPB since it was first isolated in 1995. The T3SS of UW4 consists of 26 genes, with 25 genes located in one cluster (PputUW4\_3618-3642), and one gene encoding a HopJ type III effector located elsewhere (PputUW4\_0810). Sequence analyses showed that this system is likely acquired from *Salmonella* spp. since the genes are highly similar to those of the typical T3SS of *Salmonella*. However, one gene, *invH*, required for the functioning of the T3SS in *Salmonella* is absent in UW4, therefore it is not clear whether or not this T3SS is functional.

Actually, UW4 is not the only PGPB that has been found to have T3SS. It has been observed that *P. fluorescens* SBW25 has a 20-kb cluster containing 22 CDSs of T3SS-related genes (Preston et al. 2001). This system resembles the T3SS of *P. syringae* at the level of amino acid sequence and with respect to genomic organization. Although the wild-type SBW25 is a PGPB and does not induce a hypersensitive response (HR) in host plants, a modified strain that over-expressed the sigma factor RspL specific to T3SS did elicit HR in *A. thaliana* and *Nicotiana clevelandii* (Preston et al. 2001). T3SS have been found in three other *P. fluorescens* strains including WH6, KD, and Q8r1-96 (Kimbrel et al. 2010; Rezzonico et al. 2004, 2005; Mavrodi et al. 2011). WH6 seems to have a complete and functional T3SS (PFWH6\_0718-0737) consisting of 20 genes, and it is highly homologous to the T3SS region of *P. syringae* (Kimbrel et al. 2010). The T3SS of the biocontrol strain KD is also thought to originate from *P. syringae*. It has been demonstrated that this T3SS is functional in KD, and the T3SS mutant of KD had low biocontrol activity against *Pythium ultimum* on cucumber while maintaining its root-colonization ability (Rezzonico et al. 2005). Similar to SBW25 and KD, the strain Q8r1-96 has a functional T3SS with a *P. syringae* origin. However, the genomic organization of the gene cluster is divergent from SBW25 and KD (Mavrodi et al. 2011).

Although T3SS has been found in many non-pathogenic *Pseudomonas* sp., the ecological significance of this observation is still unclear. Several studies have shown that the T3SS enhanced the biocontrol ability of the wild-type strains instead of eliciting HR in the host plants. However, the T3SS in strain UW4 apparently has a different origin compared to all of the above mentioned *P. fluorescens* strains, and it doesn't seem to have a complete pathway compared to the T3SS of its putative origin, *Salmonella* sp. Therefore, experimental approaches are necessary to verify the functionality (or lack thereof) of the T3SS in UW4. If it is functional, further studies need to be performed to investigate the effect of these genes on the plant growth-promotion activity of UW4.

#### **5.4 Heat shock proteins and $\sigma^{32}$ promoter analysis**

In bacterial cells, HSPs are a group of highly conserved proteins that are rapidly induced when the cells are exposed to environmental stresses such as higher temperature, addition of ethanol, hydrogen peroxide or heavy metals, and extreme pH values in order to deal with the increased denaturation of cellular proteins. The increased transcription of HSPs is the result of an increased level of heat shock sigma factor,  $\sigma^{32}$ , which has a short half-life under normal physiological conditions. Upon a shift to abnormal conditions that trigger cellular protein denaturation, enhanced translation of *ropH* (encoding  $\sigma^{32}$ ) and increased stability of the  $\sigma^{32}$  occur, leads to preferential expression of HSP genes (Craig 1985; Bukau 1993; Yura et al. 1993).

Bacteria are frequently exposed to various environmental stresses in their natural habitats. Therefore, the heat shock response likely plays an important role in the behavior of bacteria when they are exposed to abnormal conditions. In this study, seven HSPs including DnaK, DnaJ, GroEL, ClpB, GrpE, GroES, and IbpA of UW4 were analyzed by comparing the nucleotide

sequences of this bacterium with the corresponding sequences found in other *Pseudomonas* species and *E. coli*. Phylogenetic trees were subsequently constructed in order to illustrate the evolutionary relationships among those strains. Furthermore, the promoter sequences of the transcriptional units *grpE-dnaKJ*, *groES-groEL*, *clpB*, and *ibpA* of UW4 were compared visually with corresponding sequences of other *Pseudomonas* spp. and *E. coli*.

The results obtained from the phylogeny of seven HSP genes are consistent with the whole genome phylogenetic analysis, which suggest that UW4 is a *fluorescens* and not a *putida* (Fig. 4-32). Also, the -35 and -10 regions of the  $\sigma^{32}$  promoter of the *fluorescens* strains are the closest to UW4 (Fig. 4-33), especially in the cases of the GroE operon and ClpB, supporting the reclassification of UW4. In the phylogenetic trees of the three small HSPs including GrpE, GroES, and IbpA, members of the *fluorescens* form a paraphyletic group, and in the phylogenetic trees of the large HSPs such as DnaK, DnaJ, GroEL, and ClpB, the *fluorescens* clade varies considerably in topology. These results indicate that the *fluorescens* species are more genetically diverse than the other *Pseudomonas* species.

## 6 Conclusions

In this work, the genome of *P. putida* UW4, a well-studied PGPB, was sequenced by one of the next-generation sequencing methods, pyrosequencing. Genome assembly was done by PCR and subsequent Sanger sequencing, and genome annotation was performed first by automated pipelines available online, followed by manual corrections. Strain UW4 genome analyses included general genome features, genes involved in plant growth promotion and UW4 lifestyle, central metabolic pathways, protein secretion systems, *Pseudomonas* genome comparisons, 16S rRNA gene phylogeny, and heat shock protein and promoter analyses.

As expected, many plant growth-promoting determinants were observed in the UW4 genome, including genes encoding ACC deaminase, which were previously characterized, and potential genes encoding siderophore, IAA, trehalose, and acetoin biosynthesis. Furthermore, pathways that are thought to contribute to the fitness of UW4 were also identified including production of antimicrobial compounds, aromatic compound degradation, and heavy metal resistance. The central metabolic pathways and protein secretion systems provide an overview of the physiology of strain UW4 and the strategies that it uses to interact with the environment. Comparisons among the completely sequenced *Pseudomonas* genomes provided valuable information on determining the pan and core-*Pseudomonas* genome, and offered insights into evolutionary changes between *Pseudomonas* sp.

From the results of genome analyses, two important questions with regard to the taxonomy of UW4 and the significance of the presence of a type III secretion system were discussed. It was concluded that UW4 has a better fit within the *fluorescens* group rather than the *putida* group, and this should be verified by more detailed phenotypic characteristics. Although it was a surprise to observe a T3SS in UW4, it appears that T3SS is not uncommon in *P.*

*fluorescens* strains, and more importantly the presence of this system didn't induce a hypersensitive response under the experimental conditions shown by Preston et al. (2001). Therefore, it will be very interesting to investigate the function of the T3SS in UW4.

Genome sequencing of UW4 has opened up a number of opportunities to study this PGPB from different aspects in the future, and it will absolutely benefit the development of a more complete understanding of the mechanisms used by this bacterium to promote plant growth. Knowing the complete genome sequence of UW4 allows us to see this bacterium from a whole new point of view. Because biological functions rely on interactions between different biomolecules, rather than a single gene product, the availability of the whole genetic contents of this organism will surely help to provide more additional insight in unraveling the complex biological mechanisms that UW4 and other similar organisms use to promote plant growth. This work aims to initiate a more comprehensive study of the strain UW4. The analyses that have been done will provide a fundamental basis for future studies towards fully understanding the functioning of this organism.

## References

- Abascal F, Zardoya R, Posada D. 2005. ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics* 21:2104-2105
- Abd-Alla MH. 1994. Use of organic phosphorus by *Rhizobium leguminosarum* biovar *vicia* phosphatases. *Biology and Fertility of Soils* 18:216-218
- Abeles FB, Morgan PW, Saltveit ME Jr. 1992. Ethylene in Plant Biology. 2nd ed. pp: 147-154  
Academic Press, New York
- Adaikkalam V and Swarup S. 2002. Molecular characterization of an operon, *cueAR*, encoding a putative P1-type ATPase and a MerR-type regulatory protein involved in copper homeostasis in *Pseudomonas putida*. *Microbiology* 148:2857-2867
- Adaikkalam V and Swarup S. 2005. Characterization of *copABCD* operon from a copper-sensitive *Pseudomonas putida* strain. *Canadian Journal of Microbiology* 51:209-216
- Akiyoshi DE, Regier DA, Gordon MP. 1987. Cytokinin production by *Agrobacterium* and *Pseudomonas* spp. *Journal of Bacteriology* 169:4242-4248
- Ahmad S, Selvapandiyan A, Bhatnagar RK. 2000. Phylogenetic analysis of gram-positive bacteria based on *grpE*, encoded by the *dnaK* operon. *International Journal of Systematic and Evolutionary Microbiology* 50:1761-1766
- Altmeyer RM, McNern JK, Bossio JC, Rosenshine I, Finlay BB, Galán JE. 1993. Cloning and molecular characterization of a gene involved in Salmonella adherence and invasion of cultured epithelial cells. *Molecular Microbiology* 7:89-98
- Alvarez AH, Moreno-Sánchez R, Cervantes C. 1999. Chromate efflux by means of the ChrA chromate resistance protein from *Pseudomonas aeruginosa*. *Journal of Bacteriology* 181:7398-7400

- Andersson RA, Eriksson ARB, Heikinheimo R, Mäe A, Pirhonen M, Kõiv V, Hyytiäinen, Tuikkala A, Palva ET. 2000. Quorum sensing in the plant pathogen *Erwinia carotovora* subsp. *carotovora*: The role of *expR<sub>Ecc</sub>*. *Molecular Plant-Microbe Interactions* 13:384-393
- Andrews SC, Robinson AK, Rodríguez-Quñones F. 2003. Bacterial iron homeostasis. *FEMS Microbiology Reviews* 27:215-237
- Anfoka G and Buchenauer H. 1997. Systemic acquired resistance in tomato against *Phytophthora infestans* by pre-inoculation with tobacco necrosis virus. *Physiological and Molecular Plant Pathology* 50:85-90
- Anisimova M and Gascuel O. 2006. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic Biology* 55:539-552
- Ankenbauer R, Hanne LF, Cox CD. 1986. Mapping of mutations in *Pseudomonas aeruginosa* defective in pyoverdinin production. *Journal of Bacteriology* 167:7-11
- Anzai Y, Kudo Y, Oyaizu H. 1997. The phylogeny of the genera *Chryseomonas*, *Flavimonas*, and *Pseudomonas* supports synonymy of these three genera. *International Journal of Systematic Bacteriology* 47:249-251
- Anzai Y, Kim H, Park JY, Wakabayashi H, Oyaizu H. 2000. Phylogenetic affiliation of the pseudomonads based on 16S rRNA sequence. *International Journal of Systematic and Evolutionary Microbiology* 50:1563-1589
- Atzorn R, Crozier A, Wheeler CT, Sandberg G. 1988. Production of gibberellins and indole-3-acetic acid by *Rhizobium phaseoli* in relation to nodulation of *Phaseolus vulgaris* roots. *Planta* 175:532-538
- Audenaert K, Pattery T, Cornelis P, Höfte M. 2002. Induction of systemic resistance to *Botrytis cinerea* in tomato by *Pseudomonas aeruginosa* 7NSK2: Role of salicylic acid, pyochelin, and pyocyanin. *Molecular Plant-Microbe Interactions* 15:1147-1156

- Aury JM, Cruaud C, Barbe V, Rogier O, Mangenot S, Samson G, Poulain J, Anthouard V, Scarpelli C, Artiguenave F, Wincker P. 2008. High quality draft sequences for prokaryotic genomes using a mix of new sequencing technologies. *BMC Genomics* 9:603
- Ayub ND, Tribelli PM, López NI. 2009. Polyhydroxyalkanoates are essential for maintenance of redox state in the Antarctic bacterium *Pseudomonas* sp. 14-3 during low temperature adaptation. *Extremophiles* 13:59-66
- Azco Biotech. Polonator Brochure. Retrieved Dec 29<sup>th</sup>, 2011, from <http://www.azcobiotech.com/literature.php>
- Azcón-Aguilar C, Gianinazzi-Pearson V, Fardeau JC, Gianinazzi S. 1986. Effect of vesicular-arbuscular mycorrhizal fungi and phosphate-solubilizing bacteria on growth and nutrition of soybean in a neutral-calcareous soil amended with <sup>32</sup>P-<sup>45</sup>Ca-tricalcium phosphate. *Plant and Soil* 96:3-15
- Babalola OO, Osir EO, Sanni AI, Odhaimbo GD, Bulimo WD. 2003. Amplification of 1-aminocyclopropane-1-carboxylic (ACC) deaminase from plant growth promoting rhizobacteria in *Striga*-infested soils. *African Journal of Biotechnology* 2:157-160
- Babana AH and Antoun H. 2006. Effect of Tilemsi phosphate rock-solubilizing microorganisms on phosphorus uptake and yield of field-grown wheat (*Triticum aestivum* L.) in Mali. *Plant and Soil* 287:51-58
- Badenochjones J, Rolfe BG, Letham DS. 1983. Phytohormones, *Rhizobium* mutants, and nodulation in legumes. III. Auxin metabolism in effective and ineffective pea root nodules. *Plant Physiology* 73:347-352
- Bailey TL and Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, pp: 28-36 AAAI Press, Menlo Park, California

- Bairoch A. 2000. The ENZYME database in 2000. *Nucleic Acids Research* 28:304-305
- Bais HP, Park SW, Weir TL, Callaway RM, Vivanco JM. 2004. How plants communicate using the underground information superhighway. *Trends in Plant Science* 9:26-32
- Bakker PAHM, Pieterse CMJ, Van Loon LC. 2007. Induced systemic resistance by fluorescent *Pseudomonas* spp. *Phytopathology* 97:239-243
- Barbieri P, Zanelli T, Galli E, Zanetti G. 1986. Wheat inoculation with *Azospirillum brasilense* Sp6 and some mutants altered in nitrogen fixation and indole-3-acetic acid production. *FEMS Microbiology Letters* 36:87-90
- Báscones E, Imperial J, Rulz-Argüeso T, Palacios JM. 2000. Generation of new hydrogen-recycling *Rhizobiaceae* strains by introduction of a novel *hup* minitransposon. *Applied and Environmental Microbiology* 66:4292-4299
- Bashan Y and Holguin G. 1997. *Azospirillum*-plant relationships: environmental and physiological advances (1990-1996). *Canadian Journal of Microbiology* 43:103-121
- Bashan Y and Holguin G. 1998. Proposal for the division of plant growth-promoting rhizobacteria into two classifications: biocontrol-PGPB (Plant Growth-Promoting Bacteria) and PGPB. *Soil Biology & Biochemistry* 30:1225-1228
- Bayliss C, Bent E, Culham DE, MacLellan S, Clarke AJ, Brown GL, Wood JM. 1997. Bacterial genetic loci implicated in the *Pseudomonas putida* GR12-2R3-canola mutualism: identification of an exudate-inducible sugar transporter. *Canadian Journal of Microbiology* 43:809-818
- Beattie GA. 2006. Plant-associated bacteria: survey, molecular phylogeny, genomics and recent advances. In: Gnanamanickam SS (ed.) *Plant-Associated Bacteria*, pp: 1-56 Springer, Dordrecht, The Netherlands
- Belimov AA, Safronova VI, Sergeyeva TA, Egorova TN, Matveyeva VA, Tsyganov VE, Borisov

- AY, Tikhonovich IA, Kluge C, Preisfeld A, Dietz KJ and Stepanok VV. 2001. Characterization of plant growth promoting rhizobacteria isolated from polluted soils and containing 1-aminocyclopropane-1-carboxylate deaminase. *Canadian Journal of Microbiology* 47:642-652
- Belimov AA, Hontzeas N, Safronova VI, Demchinskaya SV, Piluzza G, Bullitta S, Glick BR. 2005. Cadmium-tolerant plant growth-promoting bacteria associated with the roots of Indian mustard (*Brassica juncea* L. Czern.). *Soil Biology and Biochemistry* 37:241-250
- Belimov AA, Dodd IC, Safronova VI, Hontzeas N, Davies WJ. 2007. *Pseudomonas brassicacearum* strain Am3 containing 1-aminocyclopropane-1-carboxylate deaminase can show both pathogenic and growth-promoting properties in its interaction with tomato. *Journal of Experimental Botany* 24:1-11
- Benson DR. 1988. The genus *Frankia*: actinomycete symbionts of plants. *Microbiological Sciences* 5:9-12
- Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* 27:573-580
- Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Cheetham RK, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IMJ, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, Aniebo IC,
- Bailey DMD, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA,

Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Catenazzi MCE, Chang S, Neil Cooley R, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Echin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Furey WS, George D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, Hoschler K, Hurwitz S, Ivanov DV, Johnson MQ, James T, Huw Jones TA, Kang GD, Kerelska TH, Kersey AD, Khrebtukova I, Kindwall AP, Kingsbury Z, Kokko-Gonzales PI, Kumar A, Laurent MA, Lawley CT, Lee SE, Lee X, Liao AK, Loch JA, Lok M, Luo S, Mammen RM, Martin JW, McCauley PG, McNitt P, Mehta P, Moon KW, Mullens JW, Newington T, Ning Z, Ng BL, Novo SM, O'Neill MJ, Osborne MA, Osnowski A, Ostadan O, Paraschos LL, Pickering L, Pike AC, Pike AC, Pinkard DC, Pliskin DP, Podhasky J, Quijano VJ, Raczy C, Rae VH, Rawlings SR, Rodriguez AC, Roe PM, Rogers J, Bacigalupo MCR, Romanov N, Romieu A, Roth RK, Rourke NJ, Ruediger ST, Rusman E, Sanches-Kuiper RM, Schenker MR, Seoane JM, Shaw RJ, Shiver MK, Short SW, Sizto NL, Sluis JP, Smith MA, Sohna JES, Spence EJ, Stevens K, Sutton N, Szajkowski L, Tregidgo CL, Turcatti G, vandeVondele S, Verhovsky Y, Virk SM, Wakelin S, Walcott GC, Wang J, Worsley GJ, Yan J, Yau L, Zuerlein M, Rogers J, Mullikin JC, Hurles ME, McCooke NJ, West JS, Oaks FL, Lundberg PL, Klenerman D, Durbin R, Smith AJ. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456:53-59

Bertagnolli BL, Soglio FKD and Sinclair JB. 1996. Extracellular enzyme profiles of the fungal pathogen *Rhizoctonia solani* isolate 2B-12 and of two antagonists, *Bacillus megaterium* strain B153-2-2 and *Trichoderma harzianum* isolate Th008. I. Possible correlation with inhibition of growth and biocontrol. *Physiological and Molecular Plant Pathology* 48:145-

- Bhattacharjee RB, Singh A, Mukhopadhyay SN. 2008. Use of nitrogen-fixing bacteria as biofertiliser for non-legumes: prospects and challenges. *Applied Microbiology and Biotechnology* 80:199-209
- Bianco C and Defez R. 2010. Improvement of phosphate solubilization and *Medicago* plant yield by an indole-3-acetic acid-overproducing strain of *Sinorhizobium meliloti*. *Applied and Environmental Microbiology* 76:4626-4632
- Blaha D, Prigent-Combaret C, Mirza MS, Moenne-Loccoz Y. 2006. Phylogeny of the 1-aminocyclopropane-1-carboxylic acid deaminase-encoding gene *acdS* in phytobeneficial and pathogenic Proteobacteria and relation with strain biogeography. *FEMS Microbiology Ecology* 56:455-470
- Bleves S, Viarre V, Salacha R, Michel GPF, Filloux A, Voulhoux R. 2010. Protein secretion systems in *Pseudomonas aeruginosa*: A wealth of pathogenic weapons. *International Journal of Medical Microbiology* 300: 534-543
- Blindauer CA, Harrison MD, Robinson AK, Parkinson JA, Bowness PW, Sadler PJ, Robinson NJ. 2002. Multiple bacteria encode metallothioneins and SmtA-like zinc fingers. *Molecular Microbiology* 45:1421-1432
- Blocker A, Komoriya K, Aizawa S. 2003. Type III secretion systems and bacterial flagella: insights into their function from structural similarities. *Proceedings of the National Academy of Sciences of the United States of America* 100:3027-3030
- Blumer C and Haas D. 2000. Iron regulation of the *hcnABC* genes encoding hydrogen cyanide synthase depends on the anaerobic regulator ANR rather than on the global activator GacA in *Pseudomonas fluorescens* CHAo. *Microbiology* 146:2417-2424
- Bnayahu BY. 1991. Root excretions and their environmental effects: influence on availability of

- phosphorus. In: Waisel Y, Eshel A, Kafkafi U (eds.) Plant roots: the hidden half, pp: 529-557 Dekker, New York
- Boiero L, Perrig D, Masciarelli O, Penna C, Cassan F, Luna V. 2007. Phytohormone production by three strains of *Bradyrhizobium japonicum* and possible physiological and technological implications. *Applied Microbiology and Biotechnology* 74:874-880
- Bolam DN, Roberts S, Proctor MR, Turkenburg JP, Dodson EJ, Martinez-Fleites C, Yang M, Davis BG, Davies GJ, Gilbert HJ. 2007. The crystal structure of two macrolide glycosyltransferases provides a blueprint for host cell antibiotic immunity. *Proceedings of the National Academy of Sciences of the United States of America* 104:5336-5341
- Bottiglieri M and Keel C. 2006. Characterization of PhlG, a hydrolase that specifically degrades the antifungal compound 2,4-diacetylphloroglucinol in the biocontrol agent *Pseudomonas fluorescens* CHA0. *Applied and Environmental Microbiology* 72:418-427
- Bottini R, Fulchieri M, Pearce D, Pharis RP. 1989. Identification of Gibberellins A<sub>1</sub>, A<sub>3</sub>, and isoA<sub>3</sub> in cultures of *Azospirillum lipoferum*. *Plant Physiology* 90:45-47
- Bottini R, Cassán F, Piccoli P. 2004. Gibberellin production by bacteria and its involvement in plant growth promotion and yield increase. *Applied Microbiology and Biotechnology* 65:497-503
- Bowers J, Mitchell J, Beer E, Buzby PR, Causey M, Efcavitch JW, Jarosz M, Krzymanska-Olejniak E, Kung L, Lipson D, Lowman GM, Marappan S, McInerney P, Platt A, Roy A, Siddiqi SM, Steinmann K, Thompson JF. 2009. Virtual terminator nucleotides for next-generation DNA sequencing. *Nature Methods* 6:593-595
- Braslavsky I, Hebert B, Kartalov E, Quake SR. 2003. Sequence information can be obtained from single DNA molecules. *Proceedings of the National Academy of Sciences of the United States of America* 100:3960-3964

- Brown GR, Sutcliffe IC, Cummings SP. 2001. Reclassification of [*Pseudomonas*] *doudoroffii* (Baumann et al. 1983) into the genus *Oceanomonas* gen. nov. as *Oceanomonas doudoroffii* comb. nov., and description of a phenol-degrading bacterium from estuarine water as *Oceanomonas baumannii* sp. nov. *International Journal of Systematic and Evolutionary Microbiology* 51:67-72
- Brown JS and Holden DW. 2002. Iron acquisition by Gram-positive bacterial pathogens. *Microbes and Infection* 4:1149-1156
- Buch A, Archana G, Naresh Kumar, G. 2008. Metabolic channeling of glucose towards gluconate in phosphate-solubilizing *Pseudomonas aeruginosa* P4 under phosphorus deficiency. *Research in Microbiology* 159:635–642
- Buell CR, Joardar V, Lindeberg M, Selengut J, Paulsen IT, Gwinn ML, Dodson RJ, Deboy RT, Durkin AS, Kolonay JF, Madupu R, Daugherty S, Brinkac L, Beanan MJ, Haft DH, Nelson WC, Davidsen T, Zafar N, Zhou L, Liu J, Yuan Q, Khouri H, Fedorova N, Tran B, Russell D, Berry K, Utterback T, Van Aken SE, Feldblyum TV, D'Ascenzo M, Deng WL, Ramos AR, Alfano JR, Cartinhour S, Chatterjee AK, Delaney TP, Lazarowitz SG, Martin GB, Schneider DJ, Tang X, Bender CL, White O, Fraser CM, Collmer A. 2003. The complete genome sequence of the *Arabidopsis* and tomato pathogen *Pseudomonas syringae* pv. *tomato* DC3000. *Proceedings of the National Academy of Sciences of the United States of America* 100:10181-10186
- Bukau B. 1993. Regulation of the *Escherichia coli* heat-shock response. *Molecular Microbiology* 9:671-680
- Burd GI, Dixon DG, Glick BR. 1998. A plant growth-promoting bacterium that decreases nickel toxicity in seedlings. *Applied and Environmental Microbiology* 64:3663-3668
- Burd GI, Dixon DG, Glick BR. 2000. Plant growth promoting bacteria that decrease heavy metal

- toxicity in plants. *Canadian Journal of Microbiology* 46:237-245
- Burg SP and Burg EA. 1966. The interaction between auxin and ethylene and its role in plant growth. *American Journal of Botany* 55:262–269
- Buurman ET, Ten Voorde GJ, Teixeira de Mattos MJ. 1994. The physiological function of periplasmic glucose oxidation in phosphate-limited chemostat cultures of *Klebsiella pneumoniae* NCTC 418. *Microbiology* 140:2451–2458
- Caballero-Mellado J, Martinez-Aguilar L, Paredes-Valdez G, Estrada-de Los Santos P. 2004. *Burkholderia unamae* sp. nov., and N<sub>2</sub>-fixing rhizospheric and endophytic species. *International Journal of Systematic and Evolutionary Microbiology* 54:1165-1172
- Caba J, Centeno ML, Fernandez B, Gresshoff PM, Ligeró F. 2000. Inoculation and nitrate alter phytohormone levels in soybean roots: differences between a supernodulating mutant and the wild type. *Planta* 211:98-104
- Caballero A, Esteve-Núñez A, Zylstra GJ, Ramos JL. 2005. Assimilation of nitrogen from nitrite and trinitrotoluene in *Pseudomonas putida* JLR11. *Journal of Bacteriology* 187:396-399
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421
- Campbell BG and Thomson JA. 1996. 1-aminocyclopropane-1-carboxylate deaminase genes from *Pseudomonas* strains. *FEMS Microbiology Letters* 138:207-210
- Cánovas D, Cases I, de Lorenzo V. 2003. Heavy metal tolerance and metal homeostasis in *Pseudomonas putida* as revealed by complete genome analysis. *Environmental Microbiology* 5:1242-1256
- Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J. 2005. ACT: the Artemis Comparison Tool. *Bioinformatics* 21:3422-3423

- Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. 2009. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 25:119-120
- Cascales E and Cambillau C. 2012. Structural biology of type VI secretion systems. *Philosophical Transactions of the Royal Society* 367:1102-1111
- Castagno LN, Estrella MJ, Sannazzaro AI, Grassano AE, Ruiz OA. 2011. Phosphate-solubilization mechanism and *in vitro* plant growth promotion activity mediated by *Pantoea eucalypti* isolated from *Lotus tenuis* rhizosphere in the Salado River Basin (Argentina). *Journal of Applied Microbiology* 110:1151-1165
- Cervantes C, Ohtake H, Chu L, Misra TK, Silver S. 1990. Cloning, nucleotide sequence, and expression of the chromate resistance determinant of *Pseudomonas aeruginosa* plasmid pUM505. *Journal of Bacteriology* 172:287-291
- Chabot R, Antoun H, Cescas MP. 1996. Growth promotion of maize and lettuce by phosphate-solubilizing *Rhizobium leguminosarum* biovar *phaseoli*. *Plant and Soil* 184:311-321
- Chaiet L and Wolf FJ. 1964. The properties of streptavidin, a biotin-binding protein produced by *Streptomyces*. *Archives of Biochemistry and Biophysics* 106:1-5
- Chaiharn M, Chunnaleuchanon S, Lumyong S. 2009. Screening siderophore producing bacteria as potential biological control agent for fungal rice pathogens in Thailand. *World Journal of Microbiology and Biotechnology* 25:1919-1928
- Chambers LA and Trudinger PA. 1971. Cysteine and S-sulfocysteine biosynthesis in bacteria. *Archives of Microbiology* 77:165-184
- Chen F, Gaucher EA, Leal NA, Hutter D, Havemann SA, Govindarajan S, Ortlund EA, Benner SA. 2010. Reconstructed evolutionary adaptive paths give polymerases accepting reversible terminators for sequencing and SNP detection. *Proceedings of the National Academy of Sciences of the United States of America* 107:1948-1953

- Cheng Z, Park E, Glick BR. 2007. 1-aminocyclopropane-1-carboxylate deaminase from *Pseudomonas putida* UW4 facilitates the growth of canola in the presence of salt. *Canadian Journal of Microbiology* 53:912-918
- Cheng Z, Duncker BP, McConkey BJ, Glick BR. 2008. Transcription regulation of ACC deaminase gene expression in *Pseudomonas putida* UW4. *Canadian Journal of Microbiology* 54:128-136
- Cheng Z, Duan J, Hao Y, McConkey BJ, Glick BR. 2009a. Identification of bacterial proteins mediating the interactions between *Pseudomonas putida* UW4 and *Brassica napus* (Canola). *Molecular Plant-Microbe Interactions* 22:686-694
- Cheng Z, Wei YC, Sung WWL, Glick BR, McConkey BJ. 2009b. Proteomic analysis of the response of the plant growth-promoting bacterium *Pseudomonas putida* UW4 to nickel stress. *Proteome Science* 7:18
- Cheng Z, Woody OZ, Song J, Glick BR, McConkey BJ. 2009c. Proteome reference map for the plant growth-promoting bacterium *Pseudomonas putida* UW4. *Proteomics* 9:4271-4274
- Cheng Z, Woody OZ, McConkey BJ, Glick BR. 2011. Combined effects of the plant growth-promoting bacterium *Pseudomonas putida* UW4 and salinity stress on the *Brassica napus* proteome. *Applied Soil Ecology*. In Press
- Chin-A-Woeng TFC, Bloemberg GV, van der Bij AJ, van der Drift KMGM, Schripsema J, Kroon B, Scheffer RJ, Keel C, Bakker PAHM, Tichy HV, De Bruijn FJ, Thomas-Oates JE, Lugtenberg BJJ. 1998. Biocontrol by phenazine-1-carboxamide-producing *Pseudomonas chlororaphis* PCL1391 of tomato root rot caused by *Fusarium oxysporum* f.sp. *radicis-lycopersici*. *Molecular Plant-Microbe Interactions* 11:1069-1077
- Chin-A-Woeng TFC, Bloemberg GV, Lugtenberg BJJ. 2003. Mechanisms of biological control of phytopathogenic fungi by *Pseudomonas* spp. In: Stacey G and Keen NT (eds.) *Plant-*

Microbe Interactions 6:173-224 American Phytopathological Society, St. Paul, MN

Clark TA, Murray IA, Morgan RD, Kislyuk AO, Spittle KE, Boitano M, Fomenkov A, Roberts

RJ, Korlach J. 2011. Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Research* Dec 7 Epub ahead of print

Clark TA, Spittle KE, Turner SW, Korlach J. 2011. Direct detection and sequencing of damaged DNA bases. *Genome Integrity* 2:10

Claudel-Renard C, Chevalet C, Faraut T, Kahn D. 2003. Enzyme-specific profiles for genome annotation: PRIAM. *Nucleic Acids Research* 31:6633-6639

Coenye T, Laevens S, Gillis M, Vandamme P. 2001. Genotypic and chemotaxonomic evidence for the reclassification of *Pseudomonas woodsii* as *Burkholderia andropogonis*. *International Journal of Systematic and Evolutionary Microbiology* 51:183-185

Compant S, Duffy B, Nowak J, Clément C, Barka EA. 2005. Use of plant growth-promoting bacteria for biocontrol of plant diseases: principles, mechanisms of action, and future prospects. *Applied and Environmental Microbiology* 71:4951-4959

Cornelis GR. 2006. The type III secretion injectisome. *Nature Reviews Microbiology* 4:811-825

Costacura A and Vanderleyden J. 1995. Synthesis of phytohormones by plant-associated bacteria. *Critical Reviews in Microbiology* 21:1-18

Craig EA. 1985. The heat shock response. *CRC Critical Reviews in Biochemistry*. 18:239-280

Crosa JH, Mey AR, Payne SM (eds.). 2004. Iron transport in bacteria. ASM Press, Washington, DC, USA

Cunliffe HE, Merriman TR, Lamont IL. 1995. Cloning and characterization of *pvdS*, a gene required for pyoverdine synthesis in *Pseudomonas aeruginosa*: PvdS is probably an alternative sigma factor. *Journal of Bacteriology* 177:2744-2750

- Cvrikova M, Nedelnik J, Eder J, Binarova P. 1993. Changes in pattern of phenolic acids induced by culture filtrate of *Fusarium oxysporum* in alfalfa plants differing in susceptibility to the pathogen. *Journal of Plant Physiology* 142:1-5
- Daefler S and Russel M. 1998. The *Salmonella typhimurium* InvH protein is an outer membrane lipoprotein required for the proper localization of InvG. *Molecular Microbiology* 28:1367-1380
- De Freitas JR, Banerjee MR, Germida JJ. 1997. Phosphate-solubilizing rhizobacteria enhance the growth and yield but not phosphorus uptake of canola (*Brassica napus* L). *Biology and Fertility of Soils* 24:358-364
- De Weert S, Vermeiren H, Mulders IHM, Kuiper I, Hendrickx N, Bloemberg GV, Vanderleyden J, De Mot R, Lugtenberg BJJ. 2002. Flagella-driven chemotaxis toward exudate components is an important trait for tomato root colonization by *Pseudomonas fluorescens*. *Molecular Plant-Microbe Interactions* 15:1173-1180
- De Weger LA, Van der Vlugt CIM, Wijfjes AHM, Bakker PAHM, Schippers B, Lugtenberg B. 1987. Flagella of a plant-growth-stimulating *Pseudomonas fluorescens* strain are required for colonization of potato roots. *Journal of Bacteriology* 169:2769-2773
- De Weger LA, Bakker PAHM, Schippers B, van Loosdrecht MCM, Lugtenberg B. 1989. *Pseudomonas* spp. with mutational changes in the O-antigenic side chain of their lipopolysaccharides are affected in their ability to colonize potato roots. In: Lugtenberg BJJ (ed.) *Signal molecules in plant-microbe interactions*, pp: 197-202 Springer-Verlag, Berlin, Germany.
- De Werra P, Pechy-Tarr M, Keel C, Maurhofer M. 2009. Role of gluconic acid production in the regulation of biocontrol traits of *Pseudomonas fluorescens* CHA0. *Applied and Environmental Microbiology* 75:4162-4174

- Dekkers LC, Phoelich CC, van der Fits L, Lugtenberg BJJ. 1998. A site-specific recombinase is required for competitive root colonization by *Pseudomonas fluorescens* WCS365. Proceedings of the National Academy of Sciences of the United States of America 95:7051-7056
- Dekkers LC, van der Bij AJ, Mulders IHM, Phoelich CC, Wentwoord RAR, Glandorf DCM, Wijffelman CA, Lugtenberg BJJ. 1998. Role of the O-antigen of lipopolysaccharide, and possible roles of growth rate and of NADH:ubiquinone oxidoreductase (*nuo*) in competitive tomato root-tip colonization by *Pseudomonas fluorescens* WCS365. Molecular Plant-Microbe Interactions 11:763-771
- Dekkers LC, Mulders IH, Phoelich CC, Chin-A-Woeng TFC, Wijfjes AH, Lugtenberg BJJ. 2000. The *sss* colonization gene of the tomato-*Fusarium oxysporum* f. sp. *radicis-lycopersici* biocontrol strain *Pseudomonas fluorescens* WCS365 can improve root colonization of other wild-type *Pseudomonas* spp. bacteria. Molecular Plant-Microbe Interactions 13:1177-1183
- Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics 23:673-679
- DeLuca TF, Wu IH, Pu J, Monaghan T, Peshkin L, Singh S, Wall DP. 2006. Roundup: a multi-genome repository of orthologs and evolutionary distances. Bioinformatics 22:2044-2046
- Denner EBM, Kämpfer P, Busse HJ, Moore ERB. 1999. Reclassification of *Pseudomonas echinoides* Heumann 1962, 343<sup>AL</sup>, in the genus *Sphingomonas* as *Sphingomonas echinoides* comb. nov. International Journal of Systematic Bacteriology 49:1103-1109
- Dey R, Pal KK, Bhatt DM, Chauhan SM. 2004. Growth promotion and yield enhancement of peanut (*Arachis hypogaea* L.) by application of plant growth-promoting rhizobacteria. Microbiological Research 159:371-394
- Di Simine CD, Sayer JA, Gadd GM. 1998. Solubilization of zinc phosphate by a strain of

*Pseudomonas fluorescens* isolated from a forest soil. *Biology and Fertility of Soils* 28:87-94

Dixon R. 2004. Historical perspective – development of *nif* genetics and regulation in *Klebsiella pneumoniae*. In: Klipp W, Masepohl B, Gallon JR, Newton WE (eds.) *Genetics and Regulation of Nitrogen Fixation in Free-living Bacteria*, pp: 1-25 Kluwer Academic Publishers, Dordrecht, The Netherlands

Dobbelaere S, Croonenborghs A, Thys A, Vande Broek A, Vanderleyden J. 1999. Phytostimulatory effect of *Azospirillum brasilense* wild type and mutant strains altered in IAA production on wheat. *Plant and Soil* 212:155-164

Dohm JC, Lottaz C, Borodina T, Himmelbauer H. 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Research* 36:e105

Dörr J, Hurek T, Reinhold-Hurek B. 1998. Type IV pili are involved in plant-microbe and fungus-microbe interactions. *Molecular Microbiology* 30:7–17

Dos Santos PC and Dean DR. 2011. Co-ordination and fine-tuning of nitrogen fixation in *Azotobacter vinelandii*. *Molecular Microbiology* 79:1132-1135

Doty SL. 2011. Nitrogen-fixing endophytic bacteria for improved plant growth. In: Maheshwari DK (ed.) *Bacteria in Agrobiolgy: Plant Growth Responses*, pp: 183-199 Springer-Verlag, Berlin, Heidelberg

Drechsel H and Jung G. 1998. Peptide siderophores. *Journal of Peptide Science* 4:147-181

Drmanac R, Burns NL, Callow MJ, Carnevali P, Dahl F, Fernandez A, Halpern AL, Kermani BG, Nazarenko I, Nilsen GB, Pant KP, Sparks AB, Staker B, Yeung G, Baccash J, Borcherding AP, Brownley A, Cedeno R, Chen L, Chernikoff D, Cheung A, Chirita R, Curson B, Ebert J, Hacker CR, Hartlage R, Hauser B, Huang S, Jiang Y, Karpinchyk V, Koenig M, Kong C, Landers T, Le C, Liu J, McBride CE, Morenzoni M, Morey RE, Mutch

- K, Perazich H, Perry K, Peters BA, Peterson J, Pethiyagoda CL, Pothuraju K, Richter C, Rosenbaum AM, Roy S, Shafto J, Shananhovich V, Shannon KW, Sheppy CG, Sun M, Thakuria JV, Tran A, Vu D, Wu X, Zaranek AW, Drmanac S, Oliphant AR, Banyai WC, Martin B, Ballinger DG, Church GM, Reid CA. 2009. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 327:78-81
- Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A. 2011. Geneious v5.4, Available from <http://www.geneious.com>
- Drummond M, Walmsley J, Kennedy C. 1996. Expression from the *nifB* promoter of *Azotobacter vinelandii* can be activated by *nifA*, *vnfA*, or *anfA* transcriptional activators. *Journal of Bacteriology* 178:788–792
- Duan J, Müller KM, Charles TC, Vesely S, Glick BR. 2009. 1-aminocyclopropane-1-carboxylate (ACC) deaminase genes in Rhizobia from southern Saskatchewan. *Microbial Ecology* 57:423-436
- Duan J, Heikkila JJ, Glick BR. 2010. Sequencing a bacterial genome: an overview. In: Méndez-Vilas A (ed.) *Current Research, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology Vol. 2*, pp: 1443-1451 Formatex, Badajoz, Spain
- Dubeikovskiy AN, Mordukhova EA, Kochetkov VV, Polikarpova FY, Boronin AM. 1993. Growth promotion of black currant softwood cuttings by recombinant strain *Pseudomonas fluorescens* Bsp53A synthesizing an increased amount of indole-3-acetic acid. *Soil Biology & Biochemistry* 25:1277-1281
- Duijff BJ, Gianinazzi-Pearson V, Lemanceau P. 1997. Involvement of the outer membrane lipopolysaccharides in the endophytic colonization of tomato roots by biocontrol

*Pseudomonas fluorescens* strain WCS417r. *New Phytologist* 135:325–334

Durand E, Zoued A, Spinelli S, Watson PJH, Aschtgen MS, Journet L, Cambillau C, Cascales E.

2012. Structural characterization and oligomerization of the TssL protein, a component shared by the bacterial Type VI and Type IVb secretion systems. *The Journal of Biological Chemistry* Feb 27 [Epub ahead of print]

Dutta D and Gachhui R. 2007. Nitrogen-fixing and cellulose-producing *Gluconacetobacter kombuchae* sp. nov., isolated from Kombucha tea. *International Journal of Systematic and Evolutionary Microbiology* 57:353-357

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32:1792-1797

Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, deWinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Viecelli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* 323:133-138

Emmert EA, Klimowicz AK, Thomas MG, Handelsman J. 2004. Genetics of zwittermicin A production by *Bacillus cereus*. *Applied and Environmental Microbiology* 70:104-113

Evans ML, Ishikawa H, Estelle MA. 1994. Responses of *Arabidopsis* roots to auxin studied with high temporal resolution: comparison of wild type and auxin-response mutants. *Planta* 194:215-222

Farwell AJ, Vesely S, Nero V, Rodriguez H, Shah S, Dixon DG, Glick BR. 2006. The use of

transgenic canola (*Brassica napus*) and plant growth-promoting bacteria to enhance plant biomass at a nickel-contaminated field site. *Plant and Soil* 288:309–318

Farwell AJ, Vesely S, Nero V, Rodriguez H, McCormack K, Shah S, Dixon DG, Glick BR. 2007. Tolerance of transgenic canola plants (*Brassica napus*) amended with plant growth-promoting bacteria to flooding stress at a metal-contaminated field site. *Environmental Pollution* 147:540-545

Feil H, Feil WS, Chain P, Larimer F, DiBartolo G, Copeland A, Lykidis A, Trong S, Nolan M, Glotsman E, Thiel J, Malfatti S, Loper JE, Lapidus A, Detter JC, Land M, Richardson PM, Kyrpides NC, Ivanova N, Lindow SE. 2005. Comparison of the complete genome sequences of *Pseudomonas syringae* pv. *syringae* B728a and pv. *tomato* DC3000. *Proceedings of the National Academy of Sciences of the United States of America* 102:11064-11069

Felisberto-Rodrigues C, Durand E, Aschtgen MS, Blangy S, Ortiz-Lombardia M, Douzi B, Cambillau C, Cascales E. 2011. Towards a structural comprehension of bacterial Type VI secretion systems: characterization of the TssJ-TssM complex of an *Escherichia coli* Pathovar. *PLoS Pathogens* 7:e1002386

Feng K, Lu HM, Sheng HJ, Wang XL, Mao J. 2004. Effect of organic ligands on biological availability of inorganic phosphorus in soils. *Pedosphere* 14:85-92

Finking R and Marahiel MA. 2004. Biosynthesis of nonribosomal peptides. *Annual Review of Microbiology* 58:453-488

Fischer HM. 1994. Genetic regulation of nitrogen fixation in rhizobia. *Microbiological Reviews* 58:352-386

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM, McKenney K, Sutton G, FitzHugh W, Fields C,

- Gocayne JD, Scott J, Shirley R, Liu LI, Glodek A, Kelley JM, Weidman JF, Phillips CA, Spriggs T, Hedblom E, Cotton MD, Utterback TR, Hanna MC, Nguyen DT, Saudek DM, Brandon RC, Fine LD, Fritchman JL, Fuhrmann JL, Geoghagen NSM, Gnehm CL, McDonald LA, Small KV, Fraser CM, Smith HO, Venter JC. 1995. Whole-Genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269:496-498+507-512
- Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW. 2010. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature Methods* 7:461-465
- Foquet M, Samiee KT, Kong X, Chauduri BP, Lundquist PM, Turner SW, Freudenthal J, Roitman DB. 2008. Improved fabrication of zero-mode waveguides for single-molecule detection. *Journal of Applied Physics* 103:034301
- Fox G, Wisotzkey JD, Jurtshuk P Jr. 1992. How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *International Journal of Systematic Bacteriology* 42:166-170
- Freeman BC, Chen C, Beattie GA. 2010. Identification of the trehalose biosynthetic loci of *Pseudomonas syringae* and their contribution to fitness in the phyllosphere. *Environmental Microbiology* 12:1486-1497
- Fuchs G, Boll M, Heider J. 2011. Microbial degradation of aromatic compounds – from one strategy to four. *Nature Reviews Microbiology* 9:803-816
- Fuentes-Ramirez LE, Jimenez-Salgado T, Abarca-Ocampo IR, Cabellero-Mellado J. 1993. *Acetobacter diazotrophicus*, an indoleacetic acid producing bacterium isolated from sugar cane cultivars of Mexico. *Plant and Soil* 154:145-150
- Fujita MQ, Yoshikawa H, Ogasawara N. 1989. Structure of the *dnaA* region of *Pseudomonas*

*putida*: Conservation among three bacteria, *Bacillus subtilis*, *Escherichia coli* and *P. putida*.  
Molecular and General Genetics 215:381-387

Fulchieri M, Lucangeli C, Bottini R. 1993. Inoculation with *Azospirillum lipoferum* affects growth and gibberellin status of corn seedling roots. Plant and Cell Physiology 34:1305-1309

Gaffney T, Friedrich L, Vernooij B, Negrotto D, Nye G, Uknes S, Ward E, Kessmann H, Ryals J. 1993. Requirement of salicylic acid for the induction of systemic acquired resistance. Science 261:754-756

Galán JE and Wolf-Watz H. 2006. Protein delivery into eukaryotic cells by type III secretion machines. Nature 444:567-573

Gamalero E, Berta G, Massa N, Glick BR, Lingua G. 2009. Interactions between *Pseudomonas putida* UW4 and *Gigaspora rosea* BEG9 and their consequences for the growth of cucumber under salt stress conditions. Journal of Applied Microbiology 108:236-245

Gamalero E and Glick BR. 2011. Mechanisms used by plant growth-promoting bacteria. In: Maheshwari DK (ed.) Bacteria in Agrobiolgy: Plant Nutrient Management, pp: 17-46 Springer-Verlag, Berlin, Heiderlberg

García de Salamone IE. 2000. Direct beneficial effect of cytokinin-producing rhizobacteria on plant growth. PhD thesis. University of Saskatchewan, Saskatoon, Canada

García de Salamone IE, Hynes RK, Nelson LM. 2006. Role of cytokinins in plant growth promotion by rhizosphere bacteria. In: Siddiqui ZA (ed.) PGPR: Biocontrol and Biofertilization, pp: 173-195 Springer, Dordrecht, The Netherlands

Garg AK, Kim JK, Owens TG, Ranwala AP, Choi YD, Kochian LV, Wu RJ. 2002. Trehalose accumulation in rice plants confers high tolerance levels to different abiotic stresses. Proceedings of the National Academy of Sciences of the United States of America

99:15893-15903

- Ge L and Seah SY. 2006. Heterologous expression, purification, and characterization of an L-ornithine N<sup>5</sup>-hydroxylase involved in pyoverdine siderophore biosynthesis in *Pseudomonas aeruginosa*. *Journal of Bacteriology* 188:7205-7210
- Gebhard S, Ekanayaka N, Cook GM. 2009. The low-affinity phosphate transporter PitA is dispensable for in vitro growth of *Mycobacterium smegmatis*. *BMC Microbiology* 9:254
- Gellert M, O'Dea MH, Itoh T, Tomizawa J. 1976. Novobiocin and coumermycin inhibit DNA supercoiling catalyzed by DNA gyrase. *Proceedings of the National Academy of Sciences of the United States of America* 73:4474-4478
- Ghosh S, Penterman JN, Little RD, Chavez R, Glick BR. 2003. Three newly isolated plant growth-promoting bacilli facilitate the seedling growth of canola seedlings. *Plant Physiology and Biochemistry* 41:277-281
- Glenn TC. 2011. Field guide to next-generation DNA sequencers. *Molecular Ecology Resources* 11:759-769
- Glick BR. 1995. The enhancement of plant growth by free-living bacteria. *Canadian Journal of Microbiology* 41:109-117
- Glick BR, Karaturovic DM, Newell PC. 1995. A novel procedure for rapid isolation of plant growth promoting pseudomonads. *Canadian Journal of Microbiology* 41:533-536
- Glick BR, Penrose DM, Li J. 1998. A model for the lowering of plant ethylene concentrations by plant growth-promoting bacteria. *Journal of Theoretical Biology* 190:63-68
- Glick BR, Patten CL, Holguin G, Penrose DM. 1999. Biochemical and genetic mechanisms used by plant growth-promoting bacteria. Imperial College Press, London.
- Glick BR. 2003. Phytoremediation: synergistic use of plants and bacteria to clean up the environment. *Biotechnology Advances* 21:383-393

- Glick BR. 2004. Bacterial ACC deaminase and the alleviation of plant stress. *Advances in Applied Microbiology* 56:291-312
- Glick BR, Cheng ZY, Czarny J, Duan J. 2007a. Promotion of plant growth by ACC deaminase-producing soil bacteria. *European Journal of Plant Pathology* 119:329-339
- Glick BR, Todorovic B, Czarny J, Cheng Z, Duan J, McConkey B. 2007b. Promotion of plant growth by bacterial ACC deaminase. *Critical Reviews in Plant Sciences* 26:227-242
- Glick BR, Pasternak JJ, Patten CL. 2010. *Molecular Biotechnology: Principles and Applications of Recombinant DNA*. ASM Press, Washington DC, USA, pp: 608-612, 621-639
- Gohel V, Singh A, Vimal M, Ashwini P, Chhatpar HS. 2006. Bioprospecting and antifungal potential of chitinolytic microorganisms. *African Journal of Biotechnology* 5:54-72
- Goldberg SMD, Johnson J, Busam D, Feldblyum T, Ferriera S, Friedman R, Halpern A, Khouri H, Kravitz SA, Lauro FM, Li K, Rogers YH, Strausberg R, Sutton G, Tallon L, Thoman T, Venter E, Frazier M, Venter JC. 2006. A Sanger/pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes. *Proceedings of the National Academy of Sciences of the United States of America* 103:11240-11245
- Goldstein AH. 1986. Bacterial solubilization of mineral phosphates: historical perspectives and future prospects. *American Journal of Alternate Agriculture* 1:51-57
- Goldstein AH, Rogers RD, Mead G. 1993. Mining by microbe. *Biotechnology* 11:1250-1254
- Goldstein AH. 1994. Involvement of the quinoprotein glucose dehydrogenase in the solubilization of exogenous phosphates by gram-negative bacteria. In: Torriani-Gorini A, Yagil E, Silver S (eds.) *Phosphate in microorganisms: cellular and molecular biology*, pp: 197-203 ASM, Washington, DC
- Goldstein A. 1995. Recent progress in understanding the molecular genetics and biochemistry of calcium phosphate solubilization by Gram negative bacteria. *Biological Agriculture and*

Horticulture 12:185–193

- Gonzalez-Lopez J, Salmeron V, Martinez-Toledo MV, Ballesteros F, Ramos-Cormenzana A. 1986. Production of auxins, gibberellins and cytokinins by *Azotobacter vinelandii* ATCC 12837 in chemically defined media and dialysed soil media. *Soil Biology & Biochemistry* 18:119-120
- Göttfert M. 1993. Regulation and function of rhizobial nodulation genes. *FEMS Microbiology Letters* 104:39-63
- Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution* 27:221-224
- Govindarajan M, Balandreau J, Muthukumarasamy R, Revathi G, Lakshminarasimhan C. 2006. Improved yield of micropropagated sugarcane following inoculation by endophytic *Burkholderia vietnamiensis*. *Plant and Soil* 280:239-252
- Govindarajan M, Balandreau J, Kwon SW, Weon HY, Lakshminarasimhan C. 2008. Effects of the inoculation of *Burkholderia vietnamiensis* and related endophytic diazotrophic bacteria on grain yield of rice. *Microbial Ecology* 55:21-37
- Grichko VP and Glick BR. 2000. Identification of DNA sequences that regulate the expression of the *Enterobacter cloacae* UW4 1-aminocyclopropane-1-carboxylate deaminase gene. *Canadian Journal of Microbiology* 46:1159-1165
- Grichko VP, Filby B, Glick BR. 2000. Increased ability of transgenic plants expressing the bacterial enzyme ACC deaminase to accumulate Cd, Co, Cu, Ni, Pb, and Zn. *Journal of Biotechnology* 81:45-53
- Grichko VP and Glick BR. 2001. Amelioration of flooding stress by ACC deaminase-containing plant growth-promoting bacteria. *Plant Physiology and Biochemistry* 39:11–17

- Grimes DJ, Woese CR, MacDonell MT, Colwell RR. 1997. Systematic study of the genus *Vogesella* gen. nov. and its type species, *Vogesella indigofera* comb. nov. International Journal of Systematic Bacteriology 47:19-27
- Guindon S and Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology 52:696-704
- Gulati A, Rahi P, Vyas P. 2008. Characterization of phosphate-solubilizing fluorescent pseudomonads from the rhizosphere of seabuckthorn growing in the cold deserts of Himalayas. Current Microbiology 56:73-79
- Guo J, Xu N, Li Z, Zhang S, Wu J, Kim DH, Marma MS, Meng Q, Cao H, Li X, Shi S, Yu L, Kalachikov S, Russo JJ, Turro NJ, Ju J. 2008. Four-color DNA sequencing with 3'-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. Proceedings of the National Academy of Sciences of the United States of America 105:9145-9150
- Guo W, Wang Y, Song C, Yang C, Li Q, Li B, Su W, Sun X, Song D, Yang X, Wang S. 2011. Complete genome of *Pseudomonas mendocina* NK-01, which synthesizes medium-chain-length polyhydroxyalkanoates and alginate oligosaccharides. Journal of Bacteriology 193:3413-3414
- Guo Y, Zheng W, Rong X, Huang Y. 2008. A multilocus phylogeny of the *Streptomyces griseus* 16S rRNA gene clade: use of multilocus sequence analysis for streptomycete systematics 58:149-159
- Gupta CP, Dubey RC, Maheshwari DK. 2002. Plant growth enhancement and suppression *Macrophomina phaseolina* causing charcoal rot of peanut by fluorescent *Pseudomonas*. Biology and Fertility of Soils 35:399-405
- Gutierrez JC, Ramos F, Ortner L, Tortolero M. 1995. *nasST*, two genes involved in the induction

- of the assimilatory nitrite-nitrate reductase operon (*nasAB*) of *Azotobacter vinelandii*.  
Molecular Microbiology 18:579-591
- Gutiérrez-Mañero FJ, Ramos-Solano B, Probanza A, Mehouchi J, Tadeo FR, Talon M. 2001.  
The plant-growth-promoting rhizobacteria *Bacillus pumilus* and *Bacillus licheniformis*  
produce high amounts of physiologically active gibberellins. *Physiologia Plantarum*  
111:206-211
- Gyaneshwar P, Parekh LJ, Archana G, Poole PS, Collins MD, Hutson RA, Kumar GN. 1999.  
Involvement of a phosphate-starvation inducible glucose dehydrogenase in soil phosphate  
solubilization by *Enterobacter asburiae*. *FEMS Microbiology Letters* 171:223-229
- Gyaneshwar P, Kumar GN, Parekh LJ, Poole PS. 2002. Role of soil microorganisms in  
improving P nutrition of plants. *Plant and Soil* 245:83-93
- Haas D and Keel C. 2003. Regulation of antibiotic production in root-colonizing *Pseudomonas*  
spp. and relevance for biological control of plant disease. *Annual Review of*  
*Phytopathology* 41:117-153
- Haas D and Défago G. 2005. Biological control of soil-borne pathogens by fluorescent  
pseudomonads. *Nature Reviews Microbiology* 3:307-319
- Hacker J, Blum-Oehler G, Mühldorfer I, Tschäpe H. 1997. Pathogenicity islands of virulent  
bacteria: structure, function and impact on microbial evolution. *Molecular Microbiology*  
23:1089-1097
- Hahn HP. 1997. The type-4 pilus is the major virulence-associated adhesion of *Pseudomonas*  
*aeruginosa*-a review. *Gene* 192:99-108
- Halder AK, Misra AK, Chakrabarty PK. 1991. Solubilization of inorganic phosphates by  
*Bradyrhizobium*. *Indian Journal of Experimental Biology* 29:28-31

- Hall TA. 1999. BioEdit: A User-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41:95-98
- Han SH, Kim CH, Lee JH, Park JY, Cho SM, Park SK, Kim KY, Krishnan HB, Kim YC. 2008. Inactivation of pqq genes of *Enterobacter intermedium* 60-2G reduces antifungal activity and induction of systemic resistance. FEMS Microbiology Letters 282:140-146
- Hanahan D. 1983. Studies on transformation of *Escherichia coli* with plasmids. Journal of Molecular Biology 166:557-580
- Hantke K. 1981. Regulation of ferric iron transport in *Escherichia coli* K12: Isolation of a constitutive mutant. Molecular and General Genetics 182:288-292
- Hantke K. 2005. Bacterial zinc uptake and regulators. Current Opinion in Microbiology 8:196-202
- Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, Schork NJ, Murray SS, Topol EJ, Levy S, Frazer KA. 2009. Evaluation of next generation sequencing platforms for population targeted sequencing studies. Genome Biology 10:R32
- Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, Braslavsky I, Causey M, Colonell J, DiMeo J, Efcavitch JW, Giladi E, Gill J, Healy J, Jarosz M, Lapen D, Moulton K, Quake SR, Steinmann K, Thayer E, Tyurina A, Ward R, Weiss H, Xie Z. 2008. Single-molecule DNA sequencing of a viral genome. Science 320:106-109
- Hao X, Lin Y, Johnstone L, Baltrus DA, Miller SJ, Wei G, Rensing C. 2012. Draft genome sequence of plant growth-promoting Rhizobium *Mesorhizobium amorphae*, isolated from zinc-lean mine tailings. Journal of Bacteriology 194:736-737
- Hao Y, Charles TC, Glick BR. 2007. ACC deaminase from plant growth promoting bacteria affects crown gall development. Canadian Journal of Microbiology 53:1291-1299
- Hass D and Défago G. 2005. Biological control of soil-borne pathogens by fluorescent

- pseudomonads. *Nature Reviews Microbiology* 3:307-319
- He M, Li X, Guo L, Miller SJ, Rensing C, Wang G. 2010. Characterization and genomic analysis of chromate resistant and reducing *Bacillus cereus* strain SJ1. *BMC Microbiology* 10:221
- Hedden P and Kamiya Y. 1997. Gibberellin biosynthesis: enzymes, genes and their regulation. *Plant Molecular Biology* 48:431-460
- Heger M. 2011. New Paired-end protocol for Ion PGM increases accuracy when analyzed with SoftGenetics' algorithm. Retrieved Dec 29<sup>th</sup> 2011, from [http://www.genomeweb.com//node/1003786?hq\\_e=el&hq\\_m=1167919&hq\\_l=5&hq\\_v=94d10546c1](http://www.genomeweb.com//node/1003786?hq_e=el&hq_m=1167919&hq_l=5&hq_v=94d10546c1)
- Hemmerich C, Buechlein A, Podicheti R, Revanna KV, Dong Q. 2010. An Ergatis-based prokaryotic genome annotation web server. *Bioinformatics* 26:1122-1124
- Hider RC and Kong X. 2010. Chemistry and biology of siderophores. *Natural Product Reports* 27:637-657
- Hill WA, Bacon-Hill P, Crossman SM, Stevens C. 1983. Characterization of N<sub>2</sub>-fixing bacteria associated with sweet potato roots. *Canadian Journal of Microbiology* 29:860-862
- Hillier LW, Marth GT, Quinlan AR, Dooling D, Fewell G, Barnett D, Fox P, Glasscock JI, Hickenbotham M, Huang W, Magrini VJ, Richt RJ, Sander SN, Stewart DA, Stromberg M, Tsung EF, Wylie T, Schedl T, Wilson RK, Mardis ER. 2008. Whole-genome sequencing and variant discovery in *C. elegans*. *Nature Methods* 5:183-188
- Hilario E, Buckley T, Young J. 2004. Improved resolution on the phylogenetic relationships among *Pseudomonas* by the combined analysis of *atpD*, *carA*, *recA* and 16S rDNA. *Antonie van Leeuwenhoek* 86:51-64
- Hinsinger P. 2001. Bio-availability of soil inorganic P in the rhizosphere as affected by root induced chemical changes: a review. *Plant and Soil* 237:173-195

- Hohnadel D, Haas D, Meyer J-M. 1986. Mapping of mutations affecting pyoverdine production in *Pseudomonas aeruginosa*. *FEMS Microbiology Letters* 36:195-199
- Holt JG. 1994. *Bergey's Manual of Determinative Bacteriology* 9<sup>th</sup> edition. Williams & Wilkins, Baltimore
- Honma M and Shimomura T. 1978. Metabolism of 1-aminocyclopropane-1-carboxylic acid. *Agricultural and Biological Chemistry* 42:1825-1831
- Honma M. 1993. Stereospecific reaction of 1-aminocyclopropane-1-carboxylate deaminase. In: Pech JC, Latche A and Balague C (eds.) *Cellular and Molecular Aspects of the Plant Hormone Ethylene*, pp: 111-116 Kluwer Academic Publishers, Dordrecht, The Netherlands
- Hontzas N, Richardson AO, Belimov A, Safronova V, Abu-Omar MM, Glick BR. 2005. Evidence for horizontal transfer of 1-aminocyclopropane-1-carboxylate deaminase genes. *Applied and Environmental Microbiology* 71:7556-7558
- Hryniewicz MM and Kredich NM. 1991. The *cysP* promoter of *Salmonella typhimurium*: characterization of two binding sites for CysB protein, studies of in vivo transcription initiation, and demonstration of the anti-inducer effects of thiosulfate. *Journal of Bacteriology* 173:5876-5886
- Hsiao W, Wan I, Jones SJ, Brinkman FSL. 2003. IslandPath: aiding detection of genomic islands in prokaryotes. *Bioinformatics* 19:418-420
- Hu HC, Wang YY, Tsay YF. 2009. AtCIPK8, a CBL-interacting protein kinase, regulates the low-affinity phase of the primary nitrate response. *The Plant Journal* 57:264-278
- Huang XD, El-Alawi Y, Penrose DM, Glick BR, Greenberg BM. 2004a. Responses of three grass species to creosote during phytoremediation. *Environmental Pollution* 130:453-463
- Huang XD, El-Alawi Y, Penrose DM, Glick BR, Greenberg BM. 2004b. A multi-process phytoremediation system for removal of polycyclic aromatic hydrocarbons from

- contaminated soils. *Environmental Pollution* 130:465-476
- Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119
- Iavicoli A, Boutet E, Buchala A, Métraux JP. 2003. Induced systemic resistance in *Arabidopsis thaliana* in response to root inoculation with *Pseudomonas fluorescens* CHA0. *Molecular Plant-Microbe Interactions* 16:851-858
- Idriss EE, Makarewicz O, Farouk A, Rosner K, Greiner R, Bochow H, Richter T, Borriss R. 2002. Extracellular phytase activity of *Bacillus amyloliquefaciens* FZB45 contributes to its plant-growth-promoting effect. *Microbiology* 148:2097-2109
- Illmer P and Schinner F. 1992. Solubilization of inorganic phosphates by microorganisms isolated from forest soils. *Soil Biology & Biochemistry* 24:389-395
- Iniguez AL, Dong Y, Triplett EW. 2004. Nitrogen fixation in wheat provided by *Klebsiella pneumonia* 342. *Molecular Plant-Microbe Interaction* 17:1078-1085
- Ion Torrent. 2011a. Application Note. Ion Personal Genome Machine™ Performance Overview. Retrieved Dec 29<sup>th</sup> 2011, from [http://www.iontorrent.com/lib/images/PDFs/performance\\_overview\\_application\\_note\\_041211.pdf](http://www.iontorrent.com/lib/images/PDFs/performance_overview_application_note_041211.pdf)
- Ion Torrent. 2011b. Application Note. Ion semiconductor sequencing uniquely enables both accurate long reads and paired-end sequencing. Retrieved Dec 29<sup>th</sup> 2011, from [http://www.iontorrent.com/lib/images/PDFs/pe\\_appnote\\_v12b.pdf](http://www.iontorrent.com/lib/images/PDFs/pe_appnote_v12b.pdf)
- Jacobson MR, Brigle KE, Bennett LT, Setterquist RA, Wilson MS, Cash VL, Beynon J, Newton WE, Dean DR. 1989. Physical and genetic map of the major *nif* gene cluster from *Azotobacter vinelandii*. *Journal of Bacteriology* 171:1017-1027
- Jacobson CB, Pasternak JJ, Glick BR. 1994. Partial purification and characterization of 1-

- aminocyclopropane-1-carboxylate deaminase from the plant growth promoting rhizobacterium *Pseudomonas putida* GR12-2. Canadian Journal of Microbiology 40:1019-1025
- Jarvie T and Harkins T. 2008. 3K long-tag paired end sequencing with the Genome Sequencer FLX system. Nature Methods 5:i-ii
- Jayashree S, Vadivukkarasi P, Anand K, Kato Y, Seshadri S. 2011. Evaluation of pink-pigmented facultative methylotrophic bacteria for phosphate solubilization. Archives of Microbiology 193:543-552
- Jia YJ, Kakuta Y, Sugawara M, Igarashi T, Oki N, Kisaki M, Shoji T, Kanetuna Y, Horita T, Matsui H, Honma M. 1999. Synthesis and degradation of 1-aminocyclopropane-1-carboxylic acid by *Penicillium citrinum*. Bioscience, Biotechnology, and Biochemistry 63:542-549
- Jiang CY, Sheng XF, Qian M, Wang QY. 2008. Isolation and characterization of a heavy metal-resistant *Burkholderia* sp. from heavy metal-contaminated paddy field soil and its potential in promoting plant growth and heavy metal accumulation in metal-polluted soil. Chemosphere 72:157-164
- Jiménez JI, Miñambres B, Luis García J, Díaz E. 2004. Genomic insights in the metabolism of aromatic compounds in *Pseudomonas*. In: Ramos JL (ed.) *Pseudomonas: Biosynthesis of Macromolecules and Molecular Metabolism*, pp: 427 Kluwer Academic/Plenum Publishers, New York
- Joardar V, Lindeberg M, Jackson RW, Selengut J, Dodson R, Brinkac LM, Daugherty SC, DeBoy R, Durkin AS, Giglio MG, Madupu R, Nelson WC, Rosovitz MJ, Sullivan S, Crabtree J, Creasy T, Davidsen T, Haft DH, Zafar N, Zhou L, Halpin R, Holley T, Khouri H, Feldblyum T, White O, Fraser CM, Chatterjee AK, Cartinhour S, Schneider DJ,

- Mansfield J, Collmer A, Buell CR. 2005. Whole-genome sequence analysis of *Pseudomonas syringae* pv. *phaseolicola* 1448A reveals divergence among pathovars in genes involved in virulence and transposition. *Journal of Bacteriology* 187:6488-6498
- Joerger RD, Premakumar R, Bishop PE. 1986. Tn5-induced mutants of *Azotobacter vinelandii* affected in nitrogen fixation under Mo-deficient and Mo-sufficient conditions. *Journal of Bacteriology* 168:673–682
- Joerger RD and Bishop PE. 1988. Bacterial alternative nitrogen fixation systems. *Critical Reviews in Microbiology* 16:1-14
- Joerger RD, Jacobson MR, Premakumar R, Wolfinger ED, Bishop PE. 1989. Nucleotide-sequence and mutational analysis of the structural genes (*anfHDGK*) for the second alternative nitrogenase from *Azotobacter vinelandii*. *Journal of Bacteriology* 171:1075-1086
- Joerger RD, Loveless TM, Pau RN, Mitchenall LA, Simon BH and Bishop PE. 1990. Nucleotide-sequences and mutational analysis of the structural genes for nitrogenase-2 of *Azotobacter vinelandii*. *Journal of Bacteriology* 172:3400-3408
- Johnson JL, Palleroni N. 1989. Deoxyribonucleic acid similarities among *Pseudomonas* species. *International Journal of Systematic and Evolutionary Microbiology* 39:230-235
- Jones DL. 1998. Organic acids in the rhizosphere – a critical review. *Plant and Soil* 205:25-44
- Jongedijk E, Tigelaar H, van Roekel JSC, Bres-Vloemans SA, Dekker I, van den Elzen PJM, Cornelissen BJC, Melchers LS. 1995. Synergistic activity of chitinases and  $\beta$ -1,3-glucanases enhances fungal resistance in transgenic tomato plants. *Euphytica* 85:173-180
- Joo GJ, Kim YM, Lee IJ, Song KS, Rhee IK. 2004. Growth promotion of red pepper plug seedlings and the production of gibberellins by *Bacillus cereus*, *Bacillus macroides* and *Bacillus pumilus*. *Biotechnology Letters* 26:487-491

- Joo GJ, Kim YM, Kim JT, Rhee IK, Kim JH, Lee IJ. 2005. Gibberellins-producing rhizobacteria increase endogenous gibberellins content and promote growth of red peppers. *The Journal of Microbiology* 43:510-515
- Juretic D, Zoranic L, Zucic D. 2002. Basic charge clusters and predictions of membrane protein topology. *Journal of Chemical Information and Modeling* 42:620-632
- Kämpfer P, Falsen E, Busse HJ. 2008. Reclassification of *Pseudomonas mephitica* Claydon and Hammer 1939 as a later heterotypic synonym of *Janthinobacterium lividum* (Eisenberg 1891) De Ley et al. 1978. *International Journal of Systematic and Evolutionary Microbiology* 58:136-138
- Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Research* 32(Database issue):277-280
- Kang SM, Joo GJ, Hamayun M, Na CI, Shin DH, Kim HY, Hong JK, Lee IJ. 2009. Gibberellin production and phosphate solubilization by newly isolated strain of *Acinetobacter calcoaceticus* and its effect on plant growth. *Biotechnology Letters* 31:277-281
- Karp PD, Paley SM, Krummenacker M, Latendresse M, Dale JM, Lee TJ, Kaipa P, Gilham F, Spaulding A, Popescu L, Altman T, Paulsen I, Keseler IM, Caspi R. 2010. Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Briefings in Bioinformatics* 11:40-79
- Katiyar V and Goel R. 2004. Siderophore-mediated plant growth promotion at low temperature by mutant of fluorescent pseudomonad. *Plant Growth Regulation* 42:239-244
- Kende H. 1993. Ethylene biosynthesis. *Annual Review of Plant Physiology and Plant Molecular Biology* 44:283-307
- Kende H and Zeevaart JAD. 1997. The five “classical” plant hormones. *Plant Cell* 9:1197-1210
- Kennedy C and Dean D. 1992. The *nifU*, *nifS* and *nifV* gene products are required for activity of

all three nitrogenases of *Azotobacter vinelandii*. *Molecular and General Genetics* 231:494–498

Kertesz MA. 2004. Metabolism of sulphur-containing organic compounds. In: Ramos JL (ed.) *Pseudomonas: Biosynthesis of Macromolecules and Molecular Metabolism*, pp: 323-357  
Kluwer Academic/Plenum Publishers, New York

Khan MS, Zaidi A, Wani PA. 2007. Role of phosphate-solubilizing microorganisms in sustainable agriculture – a review. *Agronomy for Sustainable Development* 27:29-43

Kim K, Jordan D, McDonald G. 1998. *Enterobacter agglomerans*, phosphate solubilizing bacteria and microbial activity in soil: effect of carbon sources. *Soil Biology & Biochemistry* 30:995–1003

Kim WT, Silverstone A, Yip WK, Dong JG, Yang S F. 1992. Induction of 1-aminocyclopropane-1-carboxylate synthase mRNA by auxin in mung bean hypocotyls and cultured apple shoots. *Plant Physiology* 98:465–471

Kimbrell JA, Givan SA, Halgren AB, Creason AL, Mills DI, Banowetz GM, Armstrong DJ, Chang JH. 2010. An improved, high-quality draft genome sequence of the Germination-Arrest Factor-producing *Pseudomonas fluorescens* WH6. *BMC Genomics* 11:522

King RW and Evans LT. 2003. Gibberellins and flowering of grasses and cereals: prising open the lid of the “Florigen” black box. *Annual Review of Plant Physiology and Plant Molecular Biology* 54:307-328

Kircher M and Kelso J. 2010. High-throughput DNA sequencing – concepts and limitations. *Bioessays* 32:524-536

Kirner S, Hammer PE, Hill DS, Altmann A, Fischer I, Weislo LJ, Lanahan M, Van Pée KH, Ligon JM. 1998. Functions encoded by pyrrolnitrin biosynthetic genes from *Pseudomonas fluorescens*. *Journal of Bacteriology* 180:1939-1943

- Kishore AK, Pande S, Podile AR. 2006. *Pseudomonas aeruginosa* GSE 18 inhibits the cell wall degrading enzymes of *Aspergillus niger* and activates defence-related enzymes of groundnut in control of collar rot disease. *Australasian Plant Pathology* 35:259-263
- Klee HJ, Hayford MB, Kretzmer KA, Barry GF, Kishore GM. 1991. Control of ethylene synthesis by expression of a bacterial enzyme in transgenic tomato plants. *Plant Cell* 3:1187-1193
- Klee HJ and Kishore GM. 1992. Control of fruit ripening and senescence in plants. International Patent No. WO 92/12249. European Patent Office, World Intellectual Property Organization
- Klipp W, Masepohl B, Gallon JR, Newton WE (eds.). 2004. Genetics and regulation of nitrogen fixation in free-living bacteria. Kluwer Academic Publishers, Dordrecht, The Netherlands
- Kloda A, Petrov E, Meyer GR, Nguyen T, Hurst AC, Hool L, Martinac B. 2008. Mechanosensitive channel of large conductance. *The International Journal of Biochemistry & Cell Biology*. 40: 164-169
- Kloepper JW and Schroth MN. 1978. Plant growth-promoting rhizobacteria on radishes. In: *Proceedings of the Fourth International Conference on Plant Pathogenic Bacteria*. Gilbert-Clarey, Tours, France. Vol. 2, pp: 879-882
- Kloepper JW, Zablockovicz RM, Tipping EM, Lifshitz R. 1991. Plant growth promotion mediated by bacterial rhizosphere colonizers. In: Keister DL and Cregan PB (eds.) *The rhizosphere and plant growth*, pp: 315-326 Kluwer Academic Publishers, Dordrecht, The Netherlands
- Kpombekou K and Tabatabai MA. 1994. Effect of organic acids on release of phosphorus from phosphate rocks. *Soil Science* 158:442-453
- Kucey RMN. 1988. Plant growth altering effects of *Azospirillum brasilense* and *Bacillus C-11-25*

- on 2 wheat cultivars. *Journal of Applied Bacteriology* 64:187-195
- Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research* 35:3100-3108
- Lam HY, Clark MJ, Chen R, Chen R, Natsoulis G, O'Huallachain M, Dewey FE, Habegger L, Ashley EA, Gerstein MB, Butte AJ, Ji HP, Snyder M. 2011. Performance comparison of whole-genome sequencing platforms. *Nature Biotechnology* Dec 18, Epub ahead of print
- Lambrecht M, Okon Y, Vande Broek A, Vanderleyden J. 2000. Indole-3-acetic acid: a reciprocal signaling molecule in bacteria-plant interactions. *Trends in Microbiology* 8:298-300
- Lamont IL, Martin LW, Sims T, Scott A, Wallace M. 2006. Characterization of a gene encoding an acetylase required for pyoverdine synthesis in *Pseudomonas aeruginosa*. *Journal of Bacteriology* 188:3149-3152
- Langille MG and Brinkman FS. 2009. IslandViewer: an integrated interface for computational identification and visualization of genomic islands. *Bioinformatics* 25:664-665
- Langille MG, Hsiao WW, Brinkman FS. 2010. Detecting genomic islands using bioinformatics approaches. *Nature Reviews Microbiology* 8:373-382
- Latifi A, Winson MK, Foglino M, Bycroft BW, Stewart GSAB, Lazdunski L, Williams P. 1995. Multiple homologues of LuxR and LuxI control expression of virulence and secondary metabolites through quorum sensing in *Pseudomonas aeruginosa* PAO1. *Molecular Microbiology* 17:333-343
- Lawrence CB, Josten MHAJ, Tuzun S. 1996. Differential induction of pathogenesis-related proteins in tomato by *Alternaria solani* and the association of basic chitinase isozyme with resistance. *Physiological and Molecular Plant Pathology* 48:361-377
- Lazdunski AM, Ventre I, Sturgis J. 2004. Regulatory circuits and communication in Gram-

negative bacteria. *Nature Reviews Microbiology* 2:581-592

Leaungvutiviroj C, Ruangphisarn P, Hansanimitkul P, Shinkawa H, Sasaki K. 2010.

Development of a new biofertilizer with a high capacity for N<sub>2</sub> fixation, phosphate and potassium solubilization and auxin production. *Bioscience, Biotechnology, and Biochemistry* 74:1098-1101

Lechner S, Mayr R, Francis KP, Prüss, BM, Kaplan T, Wiessner-Gunkel E, Stewart GS, Scherer

S. 1998. *Bacillus weihenstephanensis* sp. nov. is a new psychrotolerant species of the *Bacillus cereus* group. *International Journal of Systematic and Evolutionary Microbiology* 48:1373-1382

Lee DG, Urbach JM, Wu G, Liberati NT, Feinbaum RL, Miyata S, Diggins LT, He J, Saucier M,

Déziel M, Friedman L, Li L, Grills G, Montgomery K, Kucherlapati R, Rahme LG, Ausubel FM. 2006. Genomic analysis reveals that *Pseudomonas aeruginosa* virulence is combinatorial. *Genome Biology* 7:R90

Lee SW, Glickmann E, Cooksey DA. 2001. Chromosomal locus for cadmium resistance in

*Pseudomonas putida* consisting of a cadmium-transporting ATPase and a MerR family response regulator. *Applied and Environmental Microbiology* 67:1437-1444

Leeman M, Van Pelt JA, Den Ouden FM, Heinsbroek M, Bakker PAHM, Schippers B. 1995.

Induction of systemic resistance against Fusarium wilt of radish by lipopolysaccharides of *Pseudomonas fluorescens*. *Phytopathology* 85:1021-1027

Leeman M, Van Pelt JA, Hendrickx MJ, Scheffer RJ, Bakker PAHM, Schippers B. 1995.

Biocontrol of Fusarium wilt of radish in commercial greenhouse trials by seed treatment with *Pseudomonas fluorescens* WCS374. *Phytopathology* 85:1301-1305

Leeman M, Den Ouden FM, Van Pelt JA, Dirkx FPM, Steijl H, Bakker PAHM, Schippers B.

1996. Iron availability affects induction of systemic resistance against Fusarium wilt of

- radish by *Pseudomonas fluorescens*. *Phytopathology* 86:149-155
- Lehoux DE, Sanschagrín F, Levesque RC. 2000. Genomics of the 35-kb *pvd* locus and analysis of novel *pvdIJK* genes implicated in pyoverdine biosynthesis in *Pseudomonas aeruginosa*. *FEMS Microbiology Letters* 190:141-146
- Leonard PM, Smits SHJ, Sedelnikova SE, Brinkman AB, de Vos WM, van der Oost J, Rice DW, Rafferty JB. 2001. Crystal structure of the Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus*. *EMBO Journal* 20:990–997
- Leoni L, Orsi N, de Lorenzo V, Visca P. 2000. Functional analysis of PvdS, an iron starvation sigma factor of *Pseudomonas aeruginosa*. *Journal of Bacteriology* 182:1481-1491
- Letham DS. 1963. Zeatin, a factor inducing cell division from *Zea mays*. *Life Sciences* 8:569-573
- Letham DS. 1973. Cytokinins from *Zea mays*. *Phytochemistry* 12:2445-2455
- Levene MJ, Korlach J, Turner SW, Foquet M, Craighead HG, Webb WW. 2003. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* 299:682-686
- Li J, Ovakim DH, Charles TC, Glick BR. 2000. An ACC deaminase minus mutant of *Enterobacter cloacae* UW4 no longer promotes root elongation. *Current Microbiology* 41:101-105
- Li J and Glick BR. 2001. Transcriptional regulation of the *Enterobacter cloacae* UW4 1-aminocyclopropane-1-carboxylate (ACC) deaminase gene (*acdS*). *Canadian Journal of Microbiology* 47:259-267
- Lichter A, Barash I, Valinsky L, Manulis S. 1995. The genes involved in cytokinin biosynthesis in *Erwinia herbicola* pv. *gypsophylae*: characterization and role in gall formation. *Journal of Bacteriology* 177:4457-4465
- Linnemannstöns P, Voß T, Hedden P, Gaskin P, Tudzynski B. 1999. Deletions in the gibberellin biosynthesis gene cluster of *Gibberella fujikuroi* by restriction enzyme-mediated integration

- and conventional transformation-mediated mutagenesis. *Applied and Environmental Microbiology* 65:2558–2564
- Liu ST, Perry KL, Schardl CL, Kado CI. 1982. *Agrobacterium* Ti plasmid indoleacetic acid gene is required for crown gall oncogenesis. *Proceedings of the National Academy of Sciences of the United States of America* 79:2812-2816
- Liu ST, Lee LY, Tai CY, Hung CH, Chang YS, Wolfram JH, Rogers R, Goldstein AH. 1992. Cloning of an *Erwinia herbicola* gene necessary for gluconic acid production and enhanced mineral phosphate solubilization in *Escherichia coli* HB101: nucleotide sequence and probable involvement in biosynthesis of the coenzyme pyrroloquinoline quinone. *Journal of Bacteriology* 174:5814–5819
- Lowe TM and Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* 25:955-964
- Lynch JM and Whipps JM. 1990. Substrate flow in the rhizosphere. *Plant and Soil* 129: 1-10
- Ma W, Sebastianova SB, Sebastian J, Burd GI, Guinel FC, Glick BR. 2003. Prevalence of 1-aminocyclopropane-1-carboxylate deaminase in *Rhizobium* spp. *Antonie Van Leeuwenhoek* 83:285-291
- MacMillan J. 2002. Occurrence of gibberellins in vascular plants, fungi, and bacteria. *Journal of Plant Growth Regulation* 20:387-442
- Madhaiyan M, Poonguzhali S, Ryu J, Sa T. 2006. Regulation of ethylene levels in canola (*Brassica campestris*) by 1-aminocyclopropane-1-carboxylate deaminase-containing *Methylobacterium fujisawaense*. *Planta* 224:268-278
- Madison LL and Huisman GW. 1999. Metabolic engineering of poly(3-hydroxyalkanoates): from DNA to plastic. *Microbiology and Molecular Biology Reviews* 63:21-53

- Maiden MCJ. 2006. Multilocus sequence typing of bacteria. *Annual Review of Microbiology* 60:561-588
- Makarewicz O, Dubrac S, Msadek T, Borriss R. 2006. Dual role of the PhoP~P response regulator: *Bacillus amyloliquefaciens* FZB45 phytase gene transcription is directed by positive and negative interactions with the *phyC* promoter. *Journal of Bacteriology* 188:6953-6965
- Makarewicz O, Neubauer S, Preusse C, Borriss R. 2008. Transition state regulator AbrB inhibits transcription of *Bacillus amyloliquefaciens* FZB45 phytase through binding at two distinct sites located within the extended *phyC* promoter region. *Journal of Bacteriology* 190:6467-6474
- Malboobi MA, Behbahani M, Madani H, Owlia P, Deljou A, Yakhchali B, Moradi M, Hassanabadi H. 2009. Performance evaluation of potent phosphate solubilizing bacteria in potato rhizosphere. *World Journal of Microbiology and Biotechnology* 25:1479-1484
- Manero FJG, Acero N, Lucas JA, Probanza A. 1996. The influence of native rhizobacteria on European alder (*Alnus glutinosa*(L.) Gaertn) growth. II. Characterisation and biological assays of metabolites from growth promoting and growth inhibiting bacteria. *Plant and Soil* 182:67-74
- Manulis S, Valinski L, Gafni Y, Hershenhorn J. 1991. Indole-3-acetic acid biosynthetic pathways in *Erwinia herbicola* in relation to pathogenicity on *Gypsophila paniculata*. *Physiological and Molecular Plant Pathology* 39:161-171
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB,

- McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376-380
- Marín A and Xia X. 2008. GC skew in protein-coding genes between the leading and lagging strands in bacterial genomes: New substitution models incorporating strand bias. *Journal of Theoretical Biology* 253:508-513
- Martínez-Noël G, Curatti L, Hernandez JA, Rubio LM. 2011. NifB and NifEN protein levels are regulated by ClpX2 under nitrogen fixation conditions in *Azotobacter vinelandii*. *Molecular Microbiology* 79:1182-1193
- Martín Y, González YV, Cabrera E, Rodríguez C, Siverio JM. 2011. NprI Ser/Thr protein kinase links nitrogen source quality and carbon availability with the yeast nitrate transporter (Ynt1) levels. *The Journal of Biological Chemistry* 286:27225-27235
- Masson-Boivin C, Giraud E, Perret X, Batut J. 2009. Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? *Trends in Microbiology* 17:458-466
- Matias SR, Pagano MC, Carvalho-Muzzi F, Oliveira CA, Almeida-Carneiro A, Horta SN, Scotti MR. 2009. Effect of rhizobia, mycorrhizal fungi and phosphate-solubilizing microorganisms in the rhizosphere of native plants used to recover an iron ore area in Brazil. *European Journal of Soil Biology* 45:259-266
- Matilla MA, Pizarro-Tobias P, Roca A, Fernández M, Duque E, Molina L, Wu X, van der Lelie D, Gómez MJ, Segura A, Ramos JL. 2011. Complete genome of the plant growth-promoting rhizobacterium *Pseudomonas putida* BIRD-1. *Journal of Bacteriology* 193:1290
- Maurhofer M, Hase C, Meuwly P, Métraux JP, Défago G. 1994. Induction of systemic resistance

- of tobacco to tobacco necrosis virus by the root-colonizing *Pseudomonas fluorescens* strain CHA0: Influence of the *gacA* gene and of pyoverdine production. *Phytopathology* 84:139-146
- Mavrodi DV, Blankenfeldt W, Thomashow LS. 2006. Phenazine compounds in fluorescent *Pseudomonas* spp.: biosynthesis and regulation. *Annual Review of Phytopathology* 44:417-445
- Mavrodi DV, Joe A, Mavrodi OV, Hassan KA, Weller DM, Paulsen IT, Loper JE, Alfano JR, Thomashow LS. 2011. Structural and functional analysis of the type III secretion system from *Pseudomonas fluorescens* Q8r1-96. *Journal of Bacteriology* 193:177-189
- Mavrodi OV, Mavrodi DV, Weller DM, Thomashow LS. 2006. Role of *ptsP*, *orfT*, and *sss* recombinase genes in root colonization by *Pseudomonas fluorescens* Q8r1-96. *Applied and Environmental Microbiology* 72:7111-7122
- Mayak S, Tirosh T, Glick BR. 2001. Stimulation of the growth of tomato, pepper and mung bean plants by the plant growth-promoting bacterium *Enterobacter cloacae* CAL3. *Biological Agriculture and Horticulture* 19:261-274
- Mayak S, Tirosh T, Glick BR. 2004a. Plant growth-promoting bacteria that confer resistance to water stress in tomato and pepper. *Plant Science* 166:525-530
- Mayak S, Tirosh T, Glick BR. 2004b. Plant growth-promoting bacteria confer resistance in tomato plants to salt stress. *Plant Physiology and Biochemistry* 42:565-572
- Mazzola M and White FF. 1994. A mutation in the indole-3-acetic acid biosynthesis pathway of *Pseudomonas syringae* pv. *syringae* affects growth in *Phaseolus vulgaris* and syringomycin production. *Journal of Bacteriology* 176:1374-1382
- McCutcheon JP and Moran NA. 2007. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *Proceedings of the National Academy of Sciences of the United*

States of America 104:19392-19397

- McMorran BJ, Kumara HMCS, Sullivan K, Lamont IL. 2001. Involvement of a transformylase enzyme in siderophore synthesis in *Pseudomonas aeruginosa*. *Microbiology* 147:1517-1524
- Mende K, Homann V, Tudzynske B. 1997. The geranylgeranyl diphosphate synthase gene of *Gibberella fujikuroi*: isolation and expression. *Molecular and General Genetics* 255:96-105
- Merriman TR, Merriman ME, Lamont IL. 1995. Nucleotide sequence of *pvdD*, a pyoverdine biosynthetic gene from *Pseudomonas aeruginosa*: PvdD has similarity to peptide synthetases. *Journal of Bacteriology* 177:252-258
- Metzker ML. 2009. Sequencing in real time. *Nature Biotechnology* 27:150-151
- Metzker ML. 2010. Sequencing technologies-the next generation. *Nature Reviews Genetics* 11:31-46
- Meyer JB, Frapolli M, Keel C, 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* 77:7345-7354
- Meziane H, Van der Sluis I, Van Loon LC, Höfte M, Bakker PAHM. 2005. Determinants of *Pseudomonas putida* WCS358 involved in inducing systemic resistance in plants. *Molecular Plant Pathology* 6:177-185
- Miethke M and Marahiel MA. 2007. Siderophore-based iron acquisition and pathogen control. *Microbiology and Molecular Biology Reviews* 71:413-451
- Miller CD, Pettee B, Zhang C, Pabst M, McLean JE, Anderson AJ. 2009. Copper and cadmium: responses in *Pseudomonas putida* KT2440. *Letters in Applied Microbiology* 49:775-783
- Miller MB and Bassler BL. 2001. Quorum sensing in bacteria. *Annual Review of Microbiology* 55:165-199
- Milner J, Silo-Suh L, Lee JC, He H, Calrdy J, Handelsman J. 1996. Production of kanosamine by

*Bacillus cereus* UW85. *Applied and Environmental Microbiology* 62:3061-3065

- Minami R, Uchiyama K, Murakami T, Kawai J, Mikami K, Yamada T, Yokoi D, Ito H, Matsui H, Honma M. 1998. Properties, sequence, and synthesis in *Escherichia coli* of 1-aminocyclopropane-1-carboxylate deaminase from *Hansenula saturnus*. *Journal of Biochemistry* 123:1112-1118
- Moir JWB and Wood NJ. 2001. Nitrate and nitrite transport in bacteria. *Cellular and Molecular Life Sciences* 58:215-224
- Molina L, Constantinescu F, Michel L, Reimann C, Duffy B, Défago G. 2003. Degradation of pathogen quorum-sensing molecules by soil bacteria: a preventive and curative biological control mechanism. *FEMS Microbiology Ecology* 45:71-81
- Molina MA, Ramos JL, Espinosa-Urgel M. 2006. A two-partner secretion system is involved in seed and root colonization and iron uptake by *Pseudomonas putida* KT2440. *Environmental Microbiology* 8:639-647
- Montanez A, Abreu C, Gill PR, Hardarson G, Sicardi M. 2009. Biological nitrogen fixation in maize (*Zea mays* L.) by <sup>15</sup>N isotope-dilution and identification of associated culturable diazotrophs. *Biology and Fertility of Soils* 45:253-263
- Morgan PW and Gausman HW. 1966. Effects of ethylene on auxin transport. *Plant Physiology* 41:45-52
- Morgan PW and Drew MC. 1997. Ethylene and plant response to stress. *Physiologia Plantarum* 100:620-630
- Morohoshi T, Someya N, Ikeda T. 2009. Novel *N*-acylhomoserine lactone-degrading bacteria isolated from the leaf surface of *Solanum tuberosum* and their quorum-quenching properties. *Bioscience, Biotechnology, and Biochemistry* 73:2124-2127
- Morrone D, Chambers J, Lowry L, Kim G, Anterola A, Bender K, Peters RJ. 2009. Gibberellin

- biosynthesis in bacteria: separate *ent*-copalyl diphosphate and *ent*-kaurene synthases in *Bradyrhizobium japonicum*. FEBS Letters 583:475-480
- Mossialos D, Ochsner U, Baysse C, Chablain P, Pirnay JP, Koedam N, Budzikiewicz H, Fernandez DU, Schafer M, Ravel J, Cornelis P. 2002. Identification of new, conserved, non-ribosomal peptide synthetases from fluorescent pseudomonads involved in the biosynthesis of the siderophore pyoverdine. Molecular Microbiology 45:1673-685
- Nahas E. 1996. Factors determining rock phosphate solubilization by microorganisms isolated from soil. World Journal of Microbiology and Biotechnology 12:567-572
- Navarro C, Wu LF, Mandrand-Berthelot MA. 1993. The *nik* operon of *Escherichia coli* encodes a periplasmic binding-protein-dependent transport system for nickel. Molecular Microbiology 9:1181-1191
- Neeraja C, Anil K, Purushotham P, Suma K, Sarma P, Moerschbacher BM, Podile AR. 2010. Biotechnological approaches to develop bacterial chitinases as a bioshield against fungal diseases of plants. Critical Reviews in Biotechnology 30:231-241
- Nelson EB. 2004. Microbial dynamics and interactions in the spermosphere. Annual Review of Phytopathology 42:271-309
- Nelson KE, Weinel C, Paulsen IT, Dodson RJ, Hilbert H, Martins dos Santos VAP, Fouts DE, Gill SR, Pop M, Holmes M, Brinkac L, Beanan M, DeBoy RT, Daugherty S, Kolonay J, Madupu R, Nelson W, White O, Peterson J, Khouri H, Hance I, Chris Lee P, Holtzapple E, Scanlan D, Tran K, Moazzez A, Utterback T, Rizzo M, Lee K, Kosack D, Moestl D, Wedler H, Lauber J, Stjepandic D, Hoheisel J, Straetz M, Heim S, Kiewitz C, Eisen J, Timmis KN, Düsterhöft A, Tümmeler B, Fraser CM. 2002. Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. Environmental Microbiology 4:799-808

- Neuman DS, Rood SB, Smith BA. 1990. Does cytokinin transport from root-shoot in the xylem sap regulate leaf responses to root hypoxia? *Journal of Experimental Botany* 41:1325-1331
- Nie L, Shah S, Burd GI, Dixon DG, Glick BR. 2002. Phytoremediation of arsenate contaminated soil by transgenic canola and the plant growth-promoting bacterium *Enterobacter cloacae* CAL2. *Plant Physiology and Biochemistry* 40:355–361
- Nies A, Nies DH, Silver S. 1990. Nucleotide sequence and expression of a plasmid-encoded chromate resistance determinant from *Alcaligenes eutrophus*. *The Journal of Biological Chemistry* 265:5648-5653
- Nieto KF and Frankenberger WT. 1989. Biosynthesis of cytokinins by *Azotobacter chroococcum*. *Soil Biology and Biochemistry* 21:967-972
- Nonaka G, Blankschien M, Herman C, Gross CA, Rhodius VA. 2006. Regulon and promoter analysis of the E. coli heat-shock factor,  $\sigma^{32}$ , reveals a multifaceted cellular response to heat stress. *Genes & Development* 20:1776-1789
- Nooden LD and Letham DS. 1993. Cytokinin metabolism and signaling in the soybean plant. *Australian Journal of Plant Physiology* 20:639-645
- Nowak-Thompson B, Chaney N, Wing JS, Gould SJ, Loper JE. 1999. Characterization of the pyoluteorin biosynthetic gene cluster of *Pseudomonas fluorescens* Pf-5. *Journal of Bacteriology* 181:2166-2174
- Ochsner UA, Vasil AI, Vasil ML. 1995. Role of the ferric uptake regulator of *Pseudomonas aeruginosa* in the regulation of siderophore and exotoxin A expression: purification and activity on iron-regulated promoters. *Journal of Bacteriology* 177:7194-7201
- Oglesby-Sherrouse AG and Vasil ML. 2010. Characterization of a heme-regulated non-coding RNA encoded by the *prrF* locus of *Pseudomonas aeruginosa*. *PLoS ONE* 5: e9930

- Oliveira ALM, Urquiga S, Dobereiner J, Baldani JI. 2002. The effect of inoculating endophytic N<sub>2</sub>-fixing bacteria on micropropagated sugarcane plants. *Plant and Soil* 242:205-215
- Oliveira ALM, Stoffels M, Schmid M, Reis VM, Baldani JI, Hartmann A. 2009. Colonization of sugarcane plantlets by mixed inoculations with diazotrophic bacteria. *European Journal of Soil Biology* 45:106-113
- Omar SA. 1998. The role of rock-phosphate-solubilizing fungi and vesicular-arbuscular-mycorrhiza (VAM) in growth of wheat plants fertilized with rock phosphate. *World Journal of Microbiology and Biotechnology* 14:211-218
- Ortet P, Barakat M, Lalaouna D, Fochesato S, Barbe V, Vacherie B, Santaella C, Heulin T, Achouak W. 2011. Complete genome sequence of a beneficial plant root-associated bacterium, *Pseudomonas brassicacearum*. *Journal of Bacteriology* 193:3146
- Ortíz-Castro R, Valencia-Cantero E, López-Bucio J. 2008. Plant growth promotion by *Bacillus megaterium* involves cytokinin signaling. *Plant Signaling & Behavior* 3:263-265
- Ostrowski J and Kredich NM. 1989. Molecular characterization of the *cysJIIH* promoters of *Salmonella typhimurium* and *Escherichia coli*: regulation by *cysB* protein and *N*-acetyl-L-serine. *Journal of Bacteriology* 171:130-140
- Ostrowski J and Kredich NM. 1990. In vitro interactions of CysB protein with the *cysJIIH* promoter of *Salmonella typhimurium*: inhibitory effects of sulfide. *Journal of Bacteriology* 172:779-785
- Ouzounis C and Sander C. 1991. A structure-derived sequence pattern for the detection of type I copper binding domains in distantly related proteins. *FEBS Letters* 279:73-78
- Palleroni NJ, Kunisawa R, Contopoulou R, Doudoroff M. 1973. Nucleic acid homologies in the genus *Pseudomonas*. *International Journal of Systematic and Evolutionary Microbiology* 23:333-339

- Palleroni NJ and Bradbury JF. 1993. *Stenotrophomonas*, a new bacterial genus for *Xanthomonas maltophilia* (Hugh 1980) Swings et al. 1983. International Journal of Systematic and Evolutionary Microbiology 43:606-609
- Palumbo JD, Yuen GY, Jochum CC, Tatum K, Kobayashi DY. 2005. Mutagenesis of  $\beta$ -1,3-glucanase genes in *Lysobacter enzymogenes* strain C3 results in reduced biological control activity toward *Bipolaris* leaf spot of tall fescue and *Pythium* Damping-off of sugar beet. Phytopathology 95:701-707
- Park KH, Lee CY, Son HJ. 2009. Mechanism of insoluble phosphate solubilization by *Pseudomonas fluorescens* RAF15 isolated from ginseng rhizosphere and its plant growth-promoting activities. Letters in Applied Microbiology 49:222-228
- Park KH, Lee OM, Jung HI, Jeong JH, Jeon YD, Hwang DY, Lee CY, Son HJ. 2010. Rapid solubilization of insoluble phosphate by a novel environmental stress-tolerant *Burkholderia vietnamiensis* M6 isolated from ginseng rhizospheric soil. Applied Microbiology and Biotechnology 86:947-955
- Passador L, Cook JM, Gambello MJ, Rust L, Iglewski BH. 1993. Expression of *Pseudomonas aeruginosa* virulence genes requires cell-to-cell communication. Science 260:1127-1130.
- Patten CL and Glick BR. 1996. Bacterial biosynthesis of indole-3-acetic acid. Canadian Journal of Microbiology 42:207-220
- Patten CL and Glick BR. 2002. Role of *Pseudomonas putida* indoleacetic acid in development of the host plant root system. Applied and Environmental Microbiology 68:3795-3801
- Paul MJ, Primavesi LF, Jhurrea D, Zhang Y. 2008. Trehalose metabolism and signaling. Annual Review of Plant Biology 59:417-441
- Paula MA, Reis VM, Döbereiner J. 1991. Interactions of *Glomus clarum* with *Acetobacter diazotrophicus* in infection of sweet potato (*Ipomoea batatas*), sugarcane (*Saccharum* spp.),

and sweet sorghum (*Sorghum vulgare*). *Biology and Fertility of Soils* 11:111-115

- Paulsen IT, Press CM, Ravel J, Kobayashi DY, Myers GSA, Mavrodi DV, DeBoy RT, Seshadri R, Ren Q, Madupu R, Dodson RJ, Durkin AS, Brinkac LM, Daugherty SC, Sullivan SA, Rosovitz MJ, Gwinn ML, Zhou L, Schneider DJ, Cartinhour SW, Nelson WC, Weidman J, Watkins K, Tran K, Khouri H, Pierson EA, Pierson III LS, Thomashow LS, Loper JE. 2005. Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. *Nature Biotechnology* 23: 873-878
- Peçonek J, Gruber C, Gallego V, Ventosa A, Busse HJ, Kämpfer P, Radax C, Stan-Lotter H. 2006. Reclassification of *Pseudomonas beijerinckii* (Hof, 1935) as *Chromohalobacter beijerinckii* comb. nov., and emended description of the species. *International Journal of Systematic and Evolutionary Microbiology* 56:1953-1957
- Pei AY, Oberdorf WE, Nossa CW, Agarwal A, Chokshi P, Gerz EA, Jin Z, Lee P, Yang L, Poles M, Brown SM, Sotero S, DeSantis T, Brodie E, Nelson K, Pei Z. 2010. Diversity of 16S rRNA genes within individual prokaryotic genomes. *Applied and Environmental Microbiology* 76:3886-3897
- Peix A, Valverde A, Rivas R, Igual JM, Ramírez-Bahena MH, Mateos PF, Santa-Regina I, Rodríguez-Barrueco C, Martínez-Molina E, Velázquez E. 2007. Reclassification of *Pseudomonas aurantiaca* as a synonym of *Pseudomonas chlororaphis* and proposal of three subspecies, *P. chlororaphis* subsp. *chlororaphis* subsp. nov., *P. chlororaphis* subsp. *aureofaciens* subsp. nov., comb. nov. and *P. chlororaphis* subsp. *aurantiaca* subsp. nov., comb. nov. *International Journal of Systematic and Evolutionary Microbiology* 57:1286-1290
- Peix A, Ramírez-Bahena MH, Velázquez E. 2009. Historical evolution and current status of the taxonomy of genus *Pseudomonas*. *Infection, Genetics and Evolution* 9:1132-1147

- Peng G, Zhang W, Luo H, Xie H, Lai W, Tan Z. 2009. *Enterobacter oryzae* sp. nov., a nitrogen-fixing bacterium isolated from the wild rice species *Oryza latifolia*. International Journal of Systematic and Evolutionary Microbiology 59:1650-1655
- Penrose DM and Glick BR. 2001. Levels of ACC and related compounds in exudate and extracts of canola seeds treated with ACC deaminase containing plant growth-promoting bacteria. Canadian Journal of Microbiology 47:368-372
- Peplow M. 2005. Farms spew out nitrogen oxides. Nature, doi:10.1038/news050718-15
- Perret X, Staehelin C, Broughton WJ. 2000. Molecular basis of symbiotic promiscuity. Microbiology and Molecular Biology Reviews 64:180-201
- Ramírez M, Valderrama B, Arrendondo-Peter R, Soberón M, Nora J, Hernández G. 1999. *Rhizobium etli* genetically engineered for the heterologous expression of *Vitreoscilla* sp. hemoglobin: effects on free-living and symbiosis. Molecular Plant-Microbe Interactions 12:1008-1015
- Persello-Cartieaux F, David P, Sarrobert C, Thibaud MC, Achouak W, Robaglia C, Nussaume L. 2001. Utilization of mutants to analyze the interaction between *Arabidopsis thaliana* and its naturally root-associated *Pseudomonas*. Planta 212:190-198
- Persello-Cartieaux F, Nussaume L, Robaglia C. 2003. Tales from the underground: molecular plant-rhizobacteria interactions. Plant, Cell & Environment 26:189-199
- Pharis RP and King RW. 1985. Gibberellins and reproductive development in seed plants. Annual Review of Plant Physiology 36:517-568
- Phillips DA and Torrey JG. 1972. Studies on cytokinin production by *Rhizobium*. Plant Physiology 49:11-15
- Piccoli P, Masciarelli O and Bottini R. 1996. Metabolism of 17,17[H-2(2)]-gibberellins A(4), A(9) and A(20) by *Azospirillum lipoferum* in chemically-defined culture medium.

Symbiosis 21:263-274

Piccoli P, Lucangeli CD, Schneider G, Bottini R. 1997. Hydrolysis of [17,17-H-2(2)]gibberellin A(20) glucoside and [17,17-H-2(2)]gibberellin A(20) glucosyl ester by *Azospirillum lipoferum* cultured in a nitrogen-free biotin-based chemically-defined medium. *Plant Growth Regulation* 23:179-182

Pierik R, Tholen D, Poorter H, Visser EJW, Voesenek LACJ. 2006. The Janus face of ethylene: growth inhibition and stimulation. *Trends in Plant Science* 11:176-183

Pieterse CMJ, Van Wees SCM, Van Pelt JA, Knoester M, Laan R, Gerrits H, Weisbeek PJ, Van Loon LC. 1998. A novel signaling pathway controlling induced systemic resistance in *Arabidopsis*. *Plant Cell* 10:1571-1580

Piłsyk S and Paszewski A. 2009. Sulfate permeases-phylogenetic diversity of sulfate transporter. *Acta Biochimica Polonica* 56:375-384

Pimentel BE, Moreno-Sánchez R, Cervantes C. 2002. Efflux of chromate by *Pseudomonas aeruginosa* cells expressing the ChrA protein. *FEMS Microbiology Letters* 212:249-254

Pliego C, Kamilova F, Lugtenberg B. 2011. Plant growth-promoting bacterial: fundamentals and exploitation. In: Maheshwari DK (ed.) *Bacterial in Agrobiolgy: Crop Ecosystems*, pp: 295-343 Springer-Verlag, Berlin, Heidelberg

Podile AR and Prakash AP. 1996. Lysis and biological control of *Aspergillus niger* by *Bacillus subtilis* AF1. *Canadian Journal of Microbiology* 42:533-538

Poole K. 2001. Multidrug efflux pumps and antimicrobial resistance in *Pseudomonas aeruginosa* and related organisms. *Journal of Molecular Microbiology and Biotechnology* 3:255-264

Posada D. 2008. jModelTest: Phylogenetic Model Averaging. *Molecular Biology and Evolution* 25:1253-1256

Prayitno J, Rolfe BG, Mathesius U. 2006. The Ethylene-insensitive sickle mutant of *Medicago*

*truncatula* shows altered auxin transport regulation during nodulation. *Plant Physiology* 142:168–180

Preston GM, Bertrand N, Rainey PB. 2001. Type III secretion in plant growth-promoting *Pseudomonas fluorescens* SBW25. *Molecular Microbiology* 41:999-1014

Prieto MA, de Eugenio LI, Galán B, Luengo JM, Witholt B. 2007. Synthesis and degradation of polyhydroxyalkanoates. In: Ramos JL and Filloux A (eds.) *Pseudomonas: A Model System in Biology*, pp: 397-428 Springer, Dordrecht, London

Pukatzke S, Ma AT, Sturtevant D, Krastins B, Sarracino D, Nelson WC, Heidelberg JF, Mekalanos JJ. 2006. Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* host model system. *Proceedings of the National Academy of Sciences of the United States of America* 103:1528-1533

Quadroni M, James P, Dainese-Hatt P, Kertesz MA. 1999. Proteome mapping, mass spectrometric sequencing and reverse transcriptase-PCR for characterization of the sulfate starvation-induced response in *Pseudomonas aeruginosa* PAO1. *European Journal of Biochemistry* 266:986-996

Quaranta D, McCarty R, Bandarian V, Rensing C. 2007. The copper-inducible *cin* operon encodes an unusual methionine-rich azurin-like protein and a Pre-Q<sub>0</sub> reductase in *Pseudomonas putida* KT2440. *Journal of Bacteriology* 189:5361-5371

Quaranta D, McEvoy MM, Rensing C. 2009. Site-directed mutagenesis identifies a molecular switch involved in copper sensing by the histidine kinase CinS in *Pseudomonas putida* KT2440. *Journal of Bacteriology* 191:5304-5311

Ran LX, Li ZN, Wu GJ, Van Loon LC, Bakker PAHM. 2005. Induction of systemic resistance against bacterial wilt in *Eucalyptus urophylla* by fluorescent *Pseudomonas* spp. *European Journal of Plant Pathology* 113:59-70

- Ray J, Bagyaraj DJ, Manjunath A. 1981. Influence of soil inoculation vesicular-arbuscular mycorrhiza and a phosphate-dissolving bacterium on plant growth and  $^{32}\text{P}$ -uptake. *Soil Biology and Biochemistry* 13:105-108
- Reed MLE and Glick BR. 2005. Growth of canola (*Brassica napus*) in the presence of plant growth-promoting bacteria and either copper or polycyclic aromatic hydrocarbons. *Canadian Journal of Microbiology* 51:1061–1069
- Reis VM, Estrada-de Los SP, Tenorio-Salgado S, Vogel J, Stoffels M, Guyon S, Mavingui P, Baldani VL, Schmid M, Baldani JI, Balandreau J, Hartmann A, Caballero-Mellado J. 2004. *Burkholderia tropica* sp. nov., a novel nitrogen-fixing, plant-associated bacterium. *International Journal of Systematic and Evolutionary Microbiology* 54:2155-2162
- Ren Y and West CA. 1992. Elicitation of diterpene biosynthesis in rice (*Oryza sativa* L.) by chitin. *Plant Physiology* 99:1169-1178
- Rezzonico F, Défago G, Moënne-Loccoz Y. 2004. Comparison of ATPase-encoding type III secretion system *hrcN* genes in biocontrol fluorescent *Pseudomonads* and in phytopathogenic proteobacteria. *Applied and Environmental Microbiology* 70:5119-5131
- Rezzonico F, Binder C, Defago G, Moenne-Loccoz Y. 2005. The type III secretion system of biocontrol *Pseudomonas fluorescens* KD targets the phytopathogenic Chromista *Pythium ultimum* and promotes cucumber protection. *Molecular Plant-Microbe Interactions* 18:991-1001
- Rivas R, García-Fraile P, Peix A, Mateos PF, Martínez-Molina E, Velázquez E. 2007. *Alcanivorax balearicus* sp. nov., isolated from Lake Martel. *International Journal of Systematic and Evolutionary Microbiology* 57:1331-1335

- Robison MM, Shah S, Tamot B, Paule KP, Moffatt BB, Glick BR. 2001. Reduced symptoms of Verticillium wilt in transgenic tomato expressing a bacterial ACC deaminase. *Molecular Plant Pathology* 2:135-145
- Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS. 2003. Comparative genomics of the vitamin B<sub>12</sub> metabolism and regulation in prokaryotes. *The Journal of Biological Chemistry* 278:41148-41159
- Rodríguez H and Fraga R. 1999. Phosphate solubilizing bacteria and their role in plant growth promotion. *Biotechnology Advances* 17:319-339
- Rodríguez H, Vessely S, Shah S, Glick BR. 2008. Effect of a nickel-tolerant ACC deaminase-producing *Pseudomonas* strain on growth of nontransformed and transgenic canola plants. *Current Microbiology* 57:170-174
- Rodríguez-Salazar J, Suárez R, Caballero-Mellado J, Iturriaga G. 2009. Trehalose accumulation in *Azospirillum brasilense* improves drought tolerance and biomass in maize plants. *FEMS Microbiology Letters* 296:52-59
- Rojas A, Holguin G, Glick BR, Bashan Y. 2001. Synergism between *Phyllobacterium* sp. (N<sub>2</sub>-fixer) and *Bacillus licheniformis* (P-solubilizer), both from a semiarid mangrove rhizosphere. *FEMS Microbiology Ecology* 35:181-187
- Rombel IT, McMorran BJ, Lamont IL. 1995. Identification of a DNA sequence motif required for expression of iron-regulated genes in pseudomonads. *Molecular and General Genetics* 246:519-528
- Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyren P. 1996. Real-time DNA sequencing using detection of pyrophosphate release. *Analytical Biochemistry* 242:84-89
- Ronaghi M, Uhlén M, Nyren P. 1998. A sequencing method based on real-time pyrophosphate. *Science* 281:363-365

- Rossolini GM, Shippa S, Riccio ML, Berlutti F, Macaskie LE, Thaller MC. 1998. Bacterial nonspecific acid phosphatases: physiology, evolution, and use as tools in microbial biotechnology. *Cellular and Molecular Life Sciences* 54:833–50
- Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M, Dimalanta E, Dressman D, Kasinskas R, Sokolsky T, Fidanza JA, Namsaraev E, McKernan KJ, Williams A, Roth GT, Bustillo J. 2011. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475:348-352
- Roy PH, Tetu SG, Larouche A, Elbourne L, Tremblay S, Ren Q, Dodson R, Harkins D, Shay R, Watkins K, Mahamoud Y, Paulsen IT. 2010. Complete genome sequence of the multiresistant taxonomic outlier *Pseudomonas aeruginosa* PA7. *PLoS One* 5:e8842
- Rubio LM and Ludden PW. 2008. Biosynthesis of the iron-molybdenum cofactor of nitrogenase. *Annual Review of Microbiology* 62:93-111
- Ryu CM, Farag MA, Hu CH, Reddy MS, Wei HX, Paré PW, Kloepper JW. 2003. Bacterial volatiles promote growth in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* 100:4927-4932
- Ryu CM, Farag MA, Hu CH, Reddy MS, Kloepper JW, Paré PW. 2004. Bacterial volatiles induce systemic resistance in *Arabidopsis*. *Plant Physiology* 134:1017-1026
- Sage CI, Ando T, Kouno K, Nagaoka T. 1998. Relative importance of protons and solution calcium concentration in phosphate rock dissolution by organic acids. *Soil Science & Plant Nutrition* 44:617-625
- Sahin F, Çakmakçi R, Kantar F. 2004. Sugar beet and barley yields in relation to inoculation with

N<sub>2</sub>-fixing and phosphate solubilizing bacteria. *Plant and Soil* 265:123-129

Saleh SS and Glick BR. 2001. Involvement of *gacS* and *rpoS* in enhancement of the plant growth-promoting capabilities of *Enterobacter cloacae* CAL2 and UW4. *Canadian Journal of Microbiology* 47:698-705

Saleh S, Huang X-D, Greenberg BM, Glick BR. 2004. Phytoremediation of persistent organic contaminants in the environment. In: Singh A, Ward O (eds.) *Soil Biology: vol. 1. Applied Bioremediation and Phytoremediation*, pp: 115-134 Springer-Verlag, Berlin

Sandy M and Butler A. 2009. Microbial iron acquisition: marine and terrestrial siderophores. *Chemical Reviews* 109:4580-4595

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* 74:5463-5467

Saravanakumar D and Samiyappan R. 2006. ACC deaminase from *Pseudomonas fluorescens* mediated saline resistance in groundnut (*Arachis hypogea*) plants. *Journal of Applied Microbiology* 102:1283-1292

Satomi M, Kimura B, Hamada T, Harayama S, Fujii T. 2002. Phylogenetic study of the genus *Oceanospirillum* based on 16S rRNA and *gyrB* genes: emended description of the genus *Oceanospirillum*, description of *Pseudospirillum* gen. nov., *Oceanobacter* gen. nov. and *Terasakiella* gen. nov. and transfer of *Oceanospirillum jannaschii* and *Pseudomonas stanieri* to *Marinobacterium* as *Marinobacterium jannaschii* comb. nov. and *Marinobacterium stanieri* comb. nov. *International Journal of Systematic and Evolutionary Microbiology* 52:739-747

Schadt EE, Turner S, Kasarskis A. 2010. A window into third-generation sequencing. *Human Molecular Genetics* 19:R227-R240

- Schnitzler JP, Madlung J, Rose A, Seitz HU. 1992. Biosynthesis of p-hydroxybenzoic acid in elicitor-treated carrot cell cultures. *Planta* 188:594-600
- Seefeldt LC, Hoffman BM, Dean DR. 2009. Mechanism of Mo-dependent nitrogenase. *Annual Review of Biochemistry* 78:701-722
- Segers P, Vancanneyt M, Pot B, Torck U, Hoste B, Dewettinck D, Falsen E, Kersters K, de Vos P. 1994. Classification of *Pseudomonas diminuta* Leifson and Hugh 1954 and *Pseudomonas vesicularis* Büsing, Döll, and Freytag 1953 in *Brevundimonas* gen. nov. as *Brevundimonas diminuta* comb. nov. and *Brevundimonas vesicularis* comb. nov., respectively. *International Journal of Systematic Bacteriology* 44:499-510
- Sekine M, Ichikawa T, Kuga N, Kobayashi M, Sakurai A, Syono K. 1988. Detection of the IAA biosynthetic pathway from tryptophan via indole-3-acetamide in *Bradyrhizobium* spp. *Plant and Cell Physiology* 29:867-874
- Sela-Buurlage MB, Ponstein AS, Bres-Vloemans SA, Melchers LS, Elzen PJM, Cornelissen BJC. 1993. Only specific tobacco (*Nicotiana tabacum*) chitinases and  $\beta$ -1,3-glucanases exhibit antifungal activity. *Plant Physiology* 101:857-863
- Self WT, Grunden AM, Hasona A, Shanmugam KT. 2001. Molybdate transport. *Research in Microbiology* 152:311-321
- Setubal JC, Dos Dantos P, Goldman BS, Ertesvag H, Espin G, Rubio LM, Valla S, Almeida NF, Balasubramanian D, Cromes L, Curatti L, Du Z, Godsy E, Goodner B, Hellner-Burris K, Hernandez JA, Houmiel K, Imperial J, Kennedy C, Larson TJ, Latreille P, Ligon LS, Lu J, Maerk M, Miller NM, Norton S, O'Carroll IP, Paulsen I, Raulfs EC, Roemer R, Rosser J, Segura D, Slater S, Stricklin SL, Studholme DJ, Sun J, Viana CJ, Wallin E, Wang B, Wheeler C, Zhu H, Dean DR, Dixon R, Wood D. 2009. Genome sequence of *Azotobacter vinelandii*, an obligate aerobe specialized to support diverse anaerobic metabolic processes.

Journal of Bacteriology 191:4534-4545

- Sevilla M, Burris RH, Gunapala N, Kennedy C. 2001. Comparison of benefit to sugarcane plant growth and  $^{15}\text{N}_2$  incorporation following inoculation of sterile plants with *Acetobacter diazotrophicus* wild-type and Nif-mutant strains. *Molecular Plant-Microbe Interactions* 14:358-366
- Shah S, Li J, Moffat BA, Glick BR. 1998. Isolation and Characterization of ACC deaminase genes from two different plant growth promoting rhizobacteria. *Canadian Journal of Microbiology* 44:833-843
- Sharma A, Johria BN, Sharma AK, Glick BR. 2003. Plant growth-promoting bacterium *Pseudomonas* sp. strain GRP3 influences iron acquisition in mung bean (*Vigna radiate* L. Wilzeck). *Soil Biology & Biochemistry* 35:887-894
- Sharma V, Kumar V, Archana G, Kumar GN. 2005. Substrate specificity of glucose dehydrogenase (GDH) of *Enterobacter asburiae* PSI3 and rock phosphate solubilization with GDH substrates as C sources. *Canadian Journal of Microbiology* 51:477-482
- Sheehy RE, Honma M, Yamada M, Sasaki T, Martineau B, Hiatt WR. 1991. Isolation, sequence, and expression in *Escherichia coli* of the *Pseudomonas* sp. strain ACP gene encoding 1-aminocyclopropane-1-carboxylate deaminase. *Journal of Bacteriology* 173:5260-5265
- Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM. 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 309:1728-1732
- Shendure J and Ji H. 2008. Next-generation DNA sequencing. *Nature Biotechnology* 26:1135-1145
- Siddiqui IA and Shoukat SS. 2003. Suppression of root-knot disease by *Pseudomonas fluorescens* CHA0 in tomato: Importance of bacterial secondary metabolite, 2,4-diacetylphloroglucinol.

Soil Biology & Biochemistry 35:1615-1623

Silby MW, Cerdeño-Tárraga AM, Vernikos GS, Giddens SR, Jackson RW, Preston GM, Zhang XX, Moon CD, Gehrig SM, Godfrey SAC, Knight CG, Malone JG, Robinson Z, Spiers AJ, Harris S, Challis GL, Yaxley AM, Harris D, Seeger K, Murphy L, Rutter S, Squares R, Quail MA, Saunders E, Mavromatis K, Brettin TS, Bentley SD, Hothersall J, Stephens E, Thomas CM, Parkhill J, Levy SB, Rainey PB, Thomson NR. 2009. Genomic and genetic analyses of diversity and plant interactions *Pseudomonas fluorescens*. *Genome Biology* 10:R51

Silo-Suh LA, Lethbridge BJ, Raffel SJ, He H, Clardy J, Handelsman J. 1994. Biological activities of two fungistatic antibiotics produced by *Bacillus cereus* UW85. *Applied and Environmental Microbiology* 60:2023-2030

Silo-Suh LA, Stabb EV, Raffel SJ, Handelsman J. 1988. Target range of zwittermicin A, an aminopolyol antibiotic from *Bacillus cereus*. *Current Microbiology* 37:6-11

Simons M, van der Bij AJ, de Weger LA, Wijffelman CA, Lugtenberg BJ. 1996. Gnotobiotic system for studying rhizosphere colonization by plant growth-promoting *Pseudomonas* bacteria. *Molecular Plant-Microbe Interactions* 9:600-607

Smith EE, Sims EH, Spencer DH, Kaul R, Olson MV. 2005. Evidence for diversifying selection at the pyoverdine locus of *Pseudomonas aeruginosa*. *Journal of Bacteriology* 187:2138-2147

Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SBH, Hood LE. 1986. Fluorescence detection in automated DNA sequence analysis. *Nature* 321:674-679

Sohm JA, Webb EA, Capone DG. 2011. Emerging patterns of marine nitrogen fixation. *Nature Reviews Microbiology* 9:499-508

- Song CX, Clark TA, Lu XY, Kislyuk A, Dai Q, Turner SW, He C, Korlach J. 2011. Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine. *Nature Methods* Nov 20 Epub ahead of print
- Spaepen S, Vanderleyden J, Remans R. 2007. Indole-3-acetic acid in microbial and microorganism-plant signaling. *FEMS Microbiology Reviews* 31:425-448
- Stearns J and Glick BR. 2003. Transgenic plants with altered ethylene biosynthesis or perception. *Biotechnology Advances*. 21:193-210
- Stearns JC, Shah S, Greenberg BM, Dixon DG, Glick BR. 2005. Tolerance of transgenic canola expressing 1-aminocyclopropane-1-carboxylic acid deaminase to growth inhibition by nickel. *Plant Physiology and Biotechnology* 43:701-708
- Steenhoudt O and Vanderleyden J. 2000. *Azospirillum*, a free-living nitrogen-fixing bacterium closely associated with grasses: genetic, biochemical and ecological aspects. *FEMS Microbiology Reviews* 24:487-506
- Steinbüchel A and Lütke-Eversloh T. 2003. Metabolic engineering and pathway construction for biotechnological production of relevant polyhydroxyalkanoates in microorganisms. *Biochemical Engineering Journal* 16:81-96
- Stevenson FJ. 1986. Cycles of soil carbon, nitrogen, phosphorus, sulfur. In: *Micronutrients*, Wiley, New York
- Sticher L, Mauch-Mani B, Métraux JP. 1997. Systemic acquired resistance. *Annual Review of Phytopathology* 35:235-270
- Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrenner P, Hickey MJ, Brinkman FS, Hufnagle WO, Kowalik DJ, Lagrou M, Garber RL, Goltry L, Tolentino E, Westbrook-Wadman S, Yuan Y, Brody LL, Coulter SN, Folger KR, Kas A, Larbig K, Lim R, Smith K, Spencer D, Wong GK, Wu Z, Paulsen IT, Reizer J, Saier MH, Hancock RE, Lory S, Olson

- MV. 2000. Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature* 406:959-964
- Strom MS and Lory S. 1993. Structure-function and biogenesis of the type IV pili. *Annual Review of Microbiology* 30:565–596
- Strzelczyk E, Kampert M, Li CY. 1994. Cytokinin-like substances and ethylene production by *Azospirillum* in media with different carbon sources. *Microbiological Research* 149:55-60
- Suttle JC. 1988. Effect of ethylene treatment on polar IAA transport, net IAA uptake and specific binding of *N*-1-naphthylphthalamic acid in tissues and microsomes isolated from etiolated pea epicotyls. *Plant Physiology* 88:795–799
- Suzuki S, Ono N, Furusawa C, Ying BW, Yomo T. 2011. Comparison of sequence reads obtained from three Next-Generation Sequencing platforms. *PLoS ONE* 6:e19534
- Swerdlow H, Wu SL, Harke H, Dovichi NJ. 1990. Capillary gel electrophoresis for DNA sequencing. Laser-induced fluorescence detection with the sheath flow cuvette. *Journal of Chromatography* 516:61-67
- Taiz L and Zeiger E. 2010. *Plant physiology*, 5<sup>th</sup> edition, pp: 545-648. Sinauer Associates Inc., Sunderland, Massachusetts, USA
- Taghavi S, van der Lelie D, Hoffman A, Zhang YB, Walla MD, Vangronsveld J, Newman L, Monchy S. 2010. Genome sequence of the plant growth promoting endophytic bacterium *Enterobacter* sp. 638. *PLoS Genetics* 6:e1000943
- Tanimoto E. 1987. Gibberellin-dependent root elongation in *Lactuca sativa*: recovery from growth retardant-suppressed elongation with thickening by low concentration of GA<sub>3</sub>. *Plant and Cell Physiology* 28:963-973
- Tao GC, Tian SJ, Cai MY, Xie GH. 2008. Phosphate-solubilizing and mineralizing abilities of bacteria isolated from soils. *Pedosphere* 18:515-523

- Tate R, Riccio A, Laccarino M, Patriarca EJ. 1997. A *cysG* mutant strain of *Rhizobium etli* pleiotropically defective in sulfate and nitrate assimilation. *Journal of Bacteriology* 179:7343-7350
- Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families. *Science* 278:631-637
- Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41
- Teale WD, Paponov IA, Palme K. 2006. Auxin in action: signaling, transport and the control of plant growth and development. *Nature Reviews Molecular Cell Biology* 7:847-859
- Templeton DM (ed.). 2002. *Molecular and Cellular Iron Transport*. Marcel Dekker Inc., New York, USA
- Theunis M, Kobayashi H, Broughton WJ, Prinsen E. 2004. Flavonoids, NodD1, NodD2, and *Nod*-box NB15 modulate expression of the *y4wEFG* locus that is required for indole-3-acetic acid synthesis in *Rhizobium* sp. strain NGR234. *Molecular Plant-Microbe Interactions* 17:1153-1161
- Thiara AS and Cundliffe E. 1988. Cloning and characterization of a DNA gyrase B gene from *Streptomyces sphaeroides* that confers resistance to novobiocin. *EMBO Journal* 7:2255-2259
- Thiara AS and Cundliffe E. 1989. Interplay of novobiocin-resistant and -sensitive DNA gyrase activities in self-protection of the novobiocin producer, *Streptomyces sphaeroides*. *Gene* 81:65-72
- Timmusk S, Nicander B, Granhall U, Tillberg E. 1999. Cytokinin production by *Paenobacillus*

*polymyxa*. Soil Biology & Biochemistry 31:1847-1852

Todorovic B and Glick BR. The interconversion of ACC deaminase and D-cysteine desulphydrase by directed mutagenesis. Planta 229:193-205

Toklikishvili N, Dandurishvili N, Vainstein A, Tediashvili M, Giorgobiani N, Lurie S, Szegedi E, Glick BR, Chernin L. 2010. Inhibitory effect of ACC deaminase-producing bacteria on crown gall formation in tomato plants infected by *Agrobacterium tumefaciens* or *A. vitis*. Plant Pathology 59:1023-1030

Tomkinson AE, Vijayakumar S, Pascal JM, Ellenberger T. DNA ligases: structure, reaction mechanism and function. 2006. Chemical Reviews 106:687-699

Toro M, Azcon R, Barea J. 1997. Improvement of arbuscular mycorrhiza development by inoculation of soil with phosphate solubilizing rhizobacteria to improve rock phosphate bioavailability (<sup>32</sup>P) and nutrient cycling. Applied and Environmental Microbiology 63:4408-4412

Townsend AR and Howarth RW. 2010. Fixing the global nitrogen problem. Scientific American 302:64-71

Tsuda M, Miyazaki H, Nakazawa T. 1995. Genetic and physical mapping of genes involved in pyoverdinin production in *Pseudomonas aeruginosa* PAO1. Journal of Bacteriology 177:423-431

Tudzynski B, Mihlan M, Rojas MC, Linnemannstons P, Gaskin P, Hedden P. 2003. Characterization of the final two genes of the gibberellin biosynthesis gene cluster of *Gibberella fujikuroi* *des* and *P450-3* encode GA4 desaturase and the 13-hydroxylase, respectively. The Journal of Biological Chemistry 278:28635-28643

Turnbull GA, Morgan JAW, Whipps JM, Saunders JR. 2001a. The role of motility in the in vitro attachment of *Pseudomonas putida* PaW8 to wheat roots. FEMS Microbiology Ecology

- Turnbull GA, Morgan JAW, Whipps JM, Saunders JR. 2001b. The role of bacterial motility in the survival and spread of *Pseudomonas fluorescens* in soil and in the attachment and colonization of wheat roots. *FEMS Microbiology Ecology* 36:21-31
- Turner BL, Paphazy MJ, Haygarth PM, McKelvie ID. 2002. Inositol phosphates in the environment. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 357:449-469
- Uchiumi T, Ohwada T, Itakura M, Mitsui H, Nukui N, Dawadi P, Kaneko T, Tabata S, Yokoyama T, Tejima K, Saeki K, Omori H, Hayashi M, Maekawa T, Sriprang R, Murooka Y, Tajima S, Simomura K, Nomura M, Suzuki A, Shimoda Y, Sioya K, Abe M, Minamisawa K. 2004. Expression islands clustered on the symbiosis island of the *Mesorhizobium loti* genome. *Journal of Bacteriology* 186:2439-2448
- Uroz S, Dessaux Y, Oger P. 2009. Quorum sensing and quorum quenching: the yin and yang of bacterial communication. *Chembiochem* 10:205-216
- Usdin K. 2008. The biological effects of simple tandem repeats: Lessons from the repeat expansion diseases. *Genome Research* 18:1011-1019
- Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H, Zeng K, Malek JA, Costa G, McKeman K, Sidow A, Fire A, Johnson SM. 2008. A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Research* 18:1051-1063
- Valverde A, Burgos A, Fiscella T, Rivas R, Velázquez E, Rodríguez-Barrueco C, Cervantes E, Chamber M, Igual JM. 2006a. Differential effects of coinoculations with *Pseudomonas jessenii* PS06 (a phosphate-solubilizing bacterium) and *Mesorhizobium cicero* C-2/2 strains on the growth and seed yield of chickpea under greenhouse and field conditions. *Plant and*

Soil 287:43-50

- Valverde A, Igual JM, Peix A, Cervantes E, Velázquez E. 2006b. *Rhizobium lusitanum* sp. nov. a bacterium that nodulates *Phaseolus vulgaris*. International Journal of Systematic and Evolutionary Microbiology 56:2631-2637
- Van Domselaar GH, Stothard P, Shrivastava S, Cruz JA, Guo A, Dong X, Lu P, Szafron D, Greiner R, Wishart DS. 2005. BASys: a web server for automated bacterial genome annotation. Nucleic Acids Research 33 (Web Server issue):W455-459
- Van Loon LC and Glick BR. 2004. Increased plant fitness by rhizobacteria. In: Sandermann H (ed.) Molecular Ecotoxicology of Plants, pp: 177-207 Springer-Verlag, Berlin.
- Van Loon LC, Bakker PAHM, Pieterse CMJ. 1997. Mechanisms of PGPR-induced resistance against pathogens. In: Ogoshi A, Kobayashi K, Homma Y, Kodama F, Kondo N, Akino S (eds.) Plant Growth-Promoting Rhizobacteria: Present Status and Future Prospects, pp: 50-57 OECD, Paris
- Van Peer R, Niemann GJ, Schippers B. 1991. Induced resistance and phytoalexin accumulation in biological control of Fusarium wilt of carnation by *Pseudomonas* sp. strain WCS417r. Phytopathology 81:728-734
- Van Peer R and Schippers B. 1992. Lipopolysaccharides of plant-growth promoting *Pseudomonas* sp. strain WCS417r induce resistance in carnation to Fusarium wilt. Netherlands Journal of Plant Pathology 98:129-139
- Van Rij ET, Wesselink M, Chin-A-Woeng TFC, Bloemberg GV, Lugtenberg BJJ. 2004. Influence of environmental conditions on the production of phenazine-1-carboxamide by *Pseudomonas chlororaphis* PCL1391. Molecular Plant-Microbe Interactions 17:557-566
- Van Wees SCM, Pieterse CMJ, Trijssenaar A, Van 't Westende Y, Hartog F, Van Loon LC. 1997. Differential induction of systemic resistance in *Arabidopsis* by biocontrol bacteria.

- Vance CP. 2001. Symbiotic nitrogen fixation and phosphorus acquisition. Plant nutrient in a world of declining renewable resources. *Plant Physiology* 127:390-397
- Vandenende CS, Vlasschaert M, Seah SY. 2004. Functional characterization of an aminotransferase required for pyoverdine siderophore biosynthesis in *Pseudomonas aeruginosa* PAO1. *Journal of Bacteriology* 186:5596-5602
- Vansuyt G, Robin A, Briat JF, Curie C, Lemanceau P. 2007. Iron acquisition from Fe-pyoverdine by *Arabidopsis thaliana*. *Molecular Plant-Microbe Interactions* 4:441-447
- Vasanthakumar A and McManus PS. 2004. Indole-3-acetic acid-producing bacteria are associated with cranberry stem gall. *Phytopathology* 94:1164-1171
- Verlinden RAJ, Hill DJ, Kenward MA, Williams CD, Radecka I. 2007. Bacterial synthesis of biodegradable polyhydroxyalkanoates. *Journal of Applied Microbiology* 102:1437-1449
- Visca P, Serino L, Orsi N. 1992. Isolation and characterization of *Pseudomonas aeruginosa* mutants blocked in the synthesis of pyoverdine. *Journal of Bacteriology* 174:5727-5731
- Visca P, Imperl F, Lamont IL. 2007a. Pyoverdine siderophores: from biogenesis to biosignificance. *Trends in Microbiology* 15:22-30
- Visca P, Franceso I, Lamont IL. 2007b. Pyoverdine synthesis and its regulation in fluorescent *Pseudomonads*. In: Varma A, Chincholkar SB (eds.) *Microbial Siderophores*. 12:135-163 Springer, New York
- Viti C, Decorosi F, Mini A, Tatti E, Giovannetti L. 2009. Involvement of the *oscA* gene in the sulphur starvation response and in Cr(VI) resistance in *Pseudomonas corrugate* 28. *Microbiology* 155:95-105
- Vodovar N, Vallenet D, Cruveiller S, Rouy Z, Barbe V, Acosta C, Cattolico L, Jubin C, Lajus A, Segurens B, Vacherie B, Wincker P, Weissenbach J, Lemaitre B, Médigue C, Boccard F.

2006. Complete genome sequence of the entomopathogenic and metabolically versatile soil bacterium *Pseudomonas entomophila*. *Nature Biotechnology* 24:673-679
- Vogeli U, Meins F, Boller T. 1988. Co-ordinated regulation of chitinase and  $\beta$ -1,3-glucanase in bean leaves. *Planta* 174:364-372
- Vyas P and Gulati A. 2009. Organic acid production *in vitro* and plant growth promotion in maize under controlled environment by phosphate-solubilizing fluorescent *Pseudomonas*. *BMC Microbiology* 9:174
- Waack S, Keller O, Asper R, Brodag T, Damm C, Fricke WF, Surovcik K, Meinicke P, Merkl R. 2006. Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *BMC Bioinformatics* 16:142
- Wang C, Knill E, Glick BR, D efago G. 2000. Effect of transferring 1-aminocyclopropane-1-carboxylic acid (ACC) deaminase genes into *Pseudomonas fluorescens* strain CHA0 and its *gacA* derivative CHA96 on their growth promoting and disease-suppressive capacities. *Canadian Journal of Microbiology* 46:898–907
- Wang R, Guegler K, Labrie ST, Crawford NM. 2000. Genomic analysis of a nutrient response in *Arabidopsis* reveals diverse expression patterns and novel metabolic and potential regulatory genes induced by nitrate. *The Plant Cell* 12:1491-1509
- Wei G, Kloepper JW, Tuzun S. 1991. Induction of systemic resistance of cucumber to *Colletotrichum orbiculare* by select strains of plant growth-promoting rhizobacteria. *Phytopathology* 81:1508-1512
- Welbaum G, Sturz AV, Dong Z, Nowak J. 2004. Fertilizing soil microorganisms to improve productivity of agroecosystems. *Critical Reviews in Plant Sciences* 23:175-193
- Weyens N, Boulet J, Adriaensen D, Timmermans JP, Prinsen E, Van Oevelen S, D’Haen J, Smeets K, van der Lelie D, Taghavi S, Vangronsveld J. 2011. Contrasting colonization and

plant growth promoting capacity between wild type and a gfp-derivative of the endophyte *Pseudomonas putida* W619 in hybrid poplar. Plant and Soil, Online First, 9 June 2011

Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song XZ, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Margulies M, Weinstock GM, Gibbs RA, Rothberg JM. The complete genome of an individual by massively parallel DNA sequencing. Nature 452:872-876

Wilderman PJ, Sowa NA, FitzGerald DJ, FitzGerald PC, Gottesman S, Ochsner UA, Vasil ML. 2004. Identification of tandem duplicate regulatory small RNAs in *Pseudomonas aeruginosa* involved in iron homeostasis. Proceedings of the National Academy of Sciences of the United States of America 101:9792-9797

Wilhelm S, Gdynia A, Tielen P, Rosenau F, Jaeger KE. 2007. The autotransporter esterase EstA of *Pseudomonas aeruginosa* is required for rhamnolipid production, cell motility, and biofilm formation. Journal of Bacteriology 189:6695-6703

Willems A, Falsen E, Pot B, Jantzen E, Hoste B, Vandamme P, Gillis M, Kersters K, de Ley J. 1990. *Acidovorax*, a new genus for *Pseudomonas facilis*, *Pseudomonas delafieldii*, E. Falsen (EF) group 13, EF group 16, and several clinical isolates, with the species *Acidovorax facilis* comb. nov., *Acidovorax delafieldii* comb. nov., and *Acidovorax temperans* sp. nov. International Journal of Systematic and Evolutionary Microbiology 40:384-398

Willems A, Goor, M, Thielemans S, Gillis M, Kersters K, de Ley J. 1992. Transfer of several phytopathogenic *Pseudomonas* species to *Acidovorax* as *Acidovorax avenae* subsp. *avenae* subsp. nov., comb. nov., *Acidovorax avenae* subsp. *citrulli*, *Acidovorax avenae* subsp. *cattleyae*, and *Acidovorax konjaci*. International Journal of Systematic and Evolutionary

Microbiology 42:107-119

- Wilson MJ and Lamont IL. 2000. Characterization of an ECF sigma factor protein from *Pseudomonas aeruginosa*. Biochemical and Biophysical Research Communications 273:578-583
- Wink J, Kroppenstedt RM, Seibert G, Stackebrandt E. 2003. *Actinomadura namibiensis* sp. nov. International Journal of Systematic and Evolutionary Microbiology 53:721-724
- Winstanley C, Langille MGI, Fothergill JL, Kukavica-Ibrulj I, Paradis-Bleau C, Sanschagrín F, Thomson NR, Winsor GL, Quail MA, Lennard N, Bignell A, Clarke L, Seeger K, Saunders D, Harris D, Parkhill J, Hancock REW, Brinkman FSL, Levesque RC. 2009. Newly introduced genomic prophage islands are critical determinants of in vivo competitiveness in the Liverpool Epidemic Strain of *Pseudomonas aeruginosa*. Genome Research 19:12-23
- Woese CR, Stackebrandt E, Weisburg WG, Paster BJ, Madigan MT, Fowler VJ, Hahn CM, Blanz P, Gupta R, Nealson KH, Fox GE. 1984. The phylogeny of purple bacteria: The alpha subdivision. Systematic and Applied Microbiology 5:315-326
- Wu X, Monchy S, Taghavi S, Zhu W, Ramos J, van der Lelie D. 2011. Comparative genomics and functional analysis of niche-specific adaptation in *Pseudomonas putida*. FEMS Microbiology Reviews 35:299-323
- Xie H, Pasternak JJ, Glick BR. 1996. Isolation and characterization of mutants of the plant growth-promoting rhizobacterium *Pseudomonas putida* GR12-2 that overproduce indoleacetic acid. Current Microbiology 32:67-71
- Yabuuchi, E, Kosako Y, Oyaizu H, Yano I, Hotta H, Hashimoto Y, Ezaki T, Arakawa M. 1992. Proposal of *Burkholderia* gen. nov. and transfer of seven species of the genus *Pseudomonas* homology group II to the new genus, with the type species *Burkholderia cepacia* (Palleroni and Holmes 1981) comb. nov. Microbiology and Immunology 36:1251-1275

- Yabuuchi E, Kosako Y, Yano I, Hotta H, Nishiuchi Y. 1995. Transfer of two *Burkholderia* and an *Alcaligenes* species to *Ralstonia* gen. Nov.: Proposal of *Ralstonia pickettii* (Ralston, Palleroni and Doudoroff 1973) comb. Nov., *Ralstonia solanacearum* (Smith 1896) comb. nov. and *Ralstonia eutropha* (Davis 1969) comb. nov. *Microbiology and Immunology* 39:897-904
- Yan Y, Ping S, Peng J, Han Y, Li L, Yang J, Dou Y, Li Y, Fan H, Fan Y, Li D, Zhan Y, Chen M, Lu W, Zhang W, Cheng Q, Jin Q, Lin M. 2010. Global transcriptional analysis of nitrogen fixation and ammonium repression in root-associated *Pseudomonas stutzeri* A1501. *BMC Genomics* 11:11
- Yan Y, Yang J, Dou Y, Chen M, Ping S, Peng J, Lu W, Zhang W, Yao Z, Li H, Liu W, He S, Geng L, Zhang X, Yang F, Yu H, Zhan Y, Li D, Lin Z, Wang Y, Elmerich C, Lin M, Jin Q. 2008. Nitrogen fixation island and rhizosphere competence traits in the genome of root-associated *Pseudomonas stutzeri* A1501. *Proceedings of the National Academy of Sciences of the United States of America* 105:7564-7569
- Ye J, Yang HC, Rosen BP, Bhattacharjee H. 2007. Crystal structure of the flavoprotein ArsH from *Sinorhizobium meliloti*. *FEBS Letters* 581:3996-4000
- Yee TW and Smith DW. 1990. *Pseudomonas* chromosomal replication origins: A bacterial class distinct from *Escherichia coli*-type origins. *Proceedings of the National Academy of Sciences of the United States of America* 87:1278-1282
- Yeramian E and Buc H. 1999. Tandem repeats in complete bacterial genome sequences: sequence and structural analyses for comparative studies. *Research in Microbiology* 150:745-754
- Yi Y, Huang W, Ge Y. 2008. Exopolysaccharide: a novel important factor in the microbial dissolution of tricalcium phosphate. *World Journal of Microbiology and Biotechnology*

24:1059-1065

- Yu H, Tang H, Wang L, Yao Y, Wu G, Xu P. 2011. Complete genome sequence of the nicotine-degrading *Pseudomonas putida* strain S16. *Journal of Bacteriology* 193:5541-5542
- Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FS. 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 26:1608-1615
- Yura T, Nagai H, Mori H. 1993. Regulation of the heat-shock response in bacteria. *Annual Review of Microbiology* 47:321-350
- Zaidi A, Khan MS, Ahemad M, Oves M. 2009. Plant growth promotion by phosphate solubilizing bacteria. *Acta Microbiologica et Immunologica Hungarica* 56:263-28
- Zurdo-Piñeiro JL, Rivas R, Trujillo ME, Vizcaíno N, Carrasco JA, Chamber M, Palomares A, Mateos PF, Martínez-Molina E, Velázquez E. 2007. *Ochrobactrum cytisi* sp. nov., isolated from nodules of *Cytisus scoparius* in Spain. *International Journal of Systematic and Evolutionary Microbiology* 57:784-788
- Zylstra GJ and Gibson DT. 1989. Toluene degradation by *Pseudomonas putida* F1. *The Journal of Biological Chemistry* 264:14940-14946

Appendix A. Examples of primers used for gaps closure.

<b>Oligo name</b>	<b>Sequence</b>	<b>Estimated Gap Size (bp)</b>
1-5	GACTATGCCGAGCGTACTGG	2619
1-3	GCCTGCTGGATGACAAACAC	
2-5	CAAATGGCTCCACAACCTCGC	341
2-3	CATCCACTACACCGCCACC	
3-5	GCAGGTCTGGAACGAAGGTC	1357
3-3	GTGGGCAAGTTCAGGGTTT	
4-5	GGGTCATCGGCACTTCTTCG	68
4-3	CGCCAGCGTTCAGCAAGT	
5-5	CGCCTGGGTTTCGGTGTT	1197
5-3	AGCGGCACTGAACCCTGA	
6-5	GGATGGCGTGCGTTTCTA	4578
6-3	CGTTTGAGGTGTATGGAGGC	
7-5	GGCTCGGCACAAATACTCCA	1197
7-3	CAACGCCGATGCCGAAAA	
8-5	GGTCGTGCTGGTGTGTTG	17
8-3	GACCGCACCCACGCTAAA	
9-5	GCCCAAAGCGGTCAGTCA	5220
9-3	CGCTGCTGAGAATGTGGG	
10-5	GGCGTGCCGTTGTTTCATT	226
10-3	TGGTTGGCTTGGGCGATA	
11-5	GAGTTCGTGCGGCGTTTG	20
11-3	CGCCGACTCCAGCAAGAT	
12-5	GCCCTGTCAGCCATTCGT	1361
12-3	AGCACGCCACCTTCACGAC	
13-5	GGCGTTGTCCATCCACCA	547
13-3	CGTGACGACCAAGGGCAAT	
14-5	CAAATGCCTGCCGTGCTG	1245
14-3	GGGCGTTTGATTCTGCTGT	
15-5	ATGAGCAGCAGCCCAACC	5349
15-3	CAGGTAATGGCGGCAACA	
16-5	GGAGTTTGCACCGCTTGA	1628
16-3	CGAATCTGAAGGAAGCCCGA	
17-5	AGACTCTGCTGCTGGTTCGC	351
17-3	GCAGCGTGACCAGACCGATT	
18-5	GCGGGCAGGGAAGTAATCG	590
18-3	CGCTTTCGCCGCAACTTCT	
19-5	CGGCGAGTTGCGGTATCTG	20
19-3	CGCCTTCTGCTCGTTTGG	
20-5	CGGGGTTTACAGTTCGTTCC	826

20-3	CCCTTGAATGCTTGCGTGT	
21-5	GCTGCCCAACGACTGAATAG	460
21-3	GCTGGTGACCCGAATGAGTA	
22-5	CCAAAGGAGGTGTCGGGAGA	1245
22-3	TGGTGATGGGCGATGTGA	
23-5	AACTGGGCGGGCGTGGACAT	300
23-3	GCCTGCCCTATCTGTCGCTGA	
24-5	CACTGCCGCACGACCGAC	499
24-3	CCTGACCCGAGCCAAATCC	
25-5	CGGACAGCACCCAATACCAG	425
25-3	GCCCTGGACCATCAAGACC	
26-5	CGAGAAGCCGCAGGTGTAGG	396
26-3	AGCGAGCACCGAAAACCC	
27-5	GCTGGTGGTGCCGAGTGAGT	1245
27-3	TAACCGCCGCTGCCTTTGG	
28-5	GGATGATGAAGTCGCCAAGCC	2372
28-3	GACGGACGGGCTTGTGGC	
29-5	TCTTTCTCGCCCTGACGC	475
29-3	TGATTCCGTTTGGTGAGCC	
30-5	GCACGCAAGACTCCTCAAACA	516
30-3	CGTTGAGCGAGCGGGAAAT	
31-5	GCGGTCAGCAGCAATCCA	20
31-3	GCCATTGCCTGCTACACCA	
32-5	CTGGGTTTGTGTGTGATGGG	1613
32-3	AGGTTACTGGGAAGGGTTGG	
33-5	CTGTGGAAGCAAACGGAGAT	4381
33-3	GCTCCGCCTTGAACATCG	
34-5	TGATGGCTGCCCTTGGTG	4526
34-3	CCAAAGGGTTCGGCGTCA	
35-5	TGGTGGCGGTGATGGTGC	360
35-3	GCCTGGAAATACTCGCTGGTCG	
36-5	CGAAGAGGGCAGGACGAT	1613
36-3	CAGCAGTTTGAGCCGTGGT	
37-5	CTACAGTTTCAGCGTGGGCA	469
37-3	GTTGCGTGCCAGTGGTGTA	
38-5	AAACCTGTGCCTGCGTATCA	352
38-3	GCTGGCGGACCGTTTGTG	
39-5	CAAGGAGGCGTTGCTGGTG	102
39-3	GCGTGCCGACCATTACCCT	
40-5	GCATTATTTCCGAGCCGCAAC	20
40-3	CGCTTTGACTCATACGCCCTACC	

41-5	ACCTCTGAAGCCTGGACG	169
41-3	TCGCATCTGTAGGAGCATC	
42-5	TGGCGGTTTCGGCTATCAA	3306
42-3	ACTGCTCTGCGTCGGGTTTCG	
43-5	GTGAGCGGACAACATTCG	1414
43-3	GCGAATGTTTGCCCTGTT	
44-5	GGCACAGGATGGAGGCAAGG	595
44-3	CGGGCAAGGCATTGGTACTC	
45-5	CGTCGGAAATCAGCCACAGG	351
45-3	CCCAATGCGTGCTAAATCCC	
46-5	GCCACTTCTTCACGCAATACCG	2490
46-3	TTCCTCGCCAGCCAGACC	
47-5	AGAAGCCCTCAACCTGTGCC	3643
47-3	GCCTTCTCCGCAACCTCATCC	
48-5	CCCTGCTGCTGTTTGTCTCCG	378
48-3	TGCTGCCCAGGAACAACCC	
49-5	TCACCGTCAACGCCAACA	994
49-3	GGTGACGCCAGGGACATCG	
50-5	TCGGTGAGTCGTTGCTGCC	4753
50-3	GGGCGATGAACGAAGGCAAT	
51-5	TTGCGGTCCAGGGAGTTGC	443
51-3	GCGGCGAGTTGTTGTTGA	
52-5	TGGAGGAAACCCGCAAGA	3241
52-3	GCAAAGCCTATCGGTTCAGC	
53-5	GCCGTTCTGGTTGGTCTTG	2795
53-3	CTGGAAATGGGCGACACG	
54-5	TTGCCTTGCTGCGGTTCA	456
54-3	GGTTGAGGGTGGCGGTGTAG	
55-5	GGCGATGTAGGTTTGTCCG	428
55-3	GGCACTTTCGCTTCAACCTC	
56-5	GGATTTAGTCGTGGAAGCC	1245
56-3	CGGTATGGCGTGAGGAGA	
57-5	GCATAACCATACGCCCCGACAA	370
57-3	CTGCCGCTGCTGCCCAAG	
58-5	GCCACCTCGGACAGTCGCTT	153
58-3	ACGCCAGCCAGGAACAGA	
59-5	CCGCTACCTCTGCTCAACCTG	1502
59-3	GCCCACTGGAATGCCTGTAAT	
60-5	ACACGCAACCTGAGGAATAC	47
60-3	GCTACGGCGGTGACGAGT	
61-5	CCGAGCAACAAACCACCC	439

61-3	CCAGTGAGGCATTGAAGACG	
62-5	CGGCGATTTGTGGTTTGGC	2272
62-3	GGATTGGCAGTTCGTGGC	
63-5	AGCCCGCCGAGGTGTTTA	344
63-3	GTCTGACACTCCGCCTTCG	
64-5	CACCACGCAACGGAACCA	6032
64-3	CGGTGACGAGCGAATGGAC	
65-5	GTGGTGCTCGGCTCGTTT	169
65-3	CCGACGCCGAGGAGTTGT	
66-5	CCTTGTGGTGCCTGTGATGG	1357
66-3	ATGGGCTCGCAGTTCGTC	
67-5	GGACGAACACGGCGAAACA	676
67-3	GAGCAATGACGGGCGAGTG	
68-5	GCCTGTTTGCTGGGTTTCG	249
68-3	CTTCCCGCAGTTTGTTTCGTT	
69-5	GCTGGACTAACTGCTGACCG	385
69-3	CACGGAACTCCTCAACCC	
70-5	TCGGCTTGGGCTTGAGGT	430
70-3	GCACCGACACCGAAGCCTAT	
71-5	TGTAGAACTCGGGCAAGGACG	3435
71-3	GCCGCTCCTGTTGCTCCAC	
72-5	CGCACCACGCTGACGAAG	388
72-3	TGACGAGGACGGCGATAGA	
73-5	CGGTCGTCAGTAGCGTCTCG	2537
73-3	GCTGACAGTGGACGAGACGC	
74-5	TCGCAATCATCATCGCAGCA	582
74-3	GATGACCCTGGCGAAGATGG	
75-5	CAGGGAAGCCGCACTCATC	81
75-3	TGTCGGCGTGGCAGATTTT	
76-5	TCCCGCAACTGGTCAAGG	707
76-3	GCAACTCAGCCAGGGTCGTCT	
77-5	CCAAATCCAACACCTGACCCA	1225
77-3	CCTGGACGGCGATGGAAA	
78-5	CAGGTTGCCGATGTTATGCG	683
78-3	GGCGTGGAAGTGGCGAAC	
79-5	GCGGTGAAGGATTGCGAACTG	1245
79-3	CGCATCACCGACTTCAGCA	
80-5	CACCTGCTGAGCCACGACTG	1245
80-3	AGCGGCAGGGAGTTGGAGT	
81-5	CGAGGTCCTGAGCGGTGAT	393
81-3	GGGCGAACGGCATCTCAA	

82-5	TGTCCCAAGGGCAGAAGC	
82-3	GGGAAGGAACACTAAATGGCTC	1361
83-5	AGGCGGAAGTGGCGGCTT	
83-3	TCAGAAGGGCGACGAGCAG	20
84-5	TTCCACCTGCTGACCACCA	
84-3	GGTTTCGGCGGGATTGGT	1124
85-5	GGGCGAGTTCTGGCTTACG	
85-3	TGTGCGTGCCGACCTCAA	5670
86-5	CGCAACAAACCGAACCTG	
86-3	GACTCGCAAGCGGTGTTT	359
87-5	CAGGTCGGTATGGATGGTG	
87-3	GTTACGGTGGCAAGGGTG	308
88-5	CGCTCGGGCACTTATCTCG	
88-3	CTTCTCGGCTCGGGTGGC	5370
89-5	CGGAAGTATGGACGGCTAC	
89-3	CGATGGCTTTGAAGTAGAGGG	459
90-5	GCTCAGGAACGGCACCAT	
90-3	GGGCAAGCAAAGCGAGTG	494
91-5	AATCGCTCGCCTCGTTCA	
91-3	CTTCTGGCTCACGACCTGCT	290
92-5	ACCTCAGCCGACAGCAATC	
92-3	GCTGGCTGGCGGAAGAAA	1361
93-5	GGAAGGGATGCGGGTGTT	
93-3	GCAACCAGTTCAAAGGCTCCA	1357
94-5	CGCCACGCTGTTCAAGGA	
94-3	TTGCGGGCACTGAACTGG	320
95-5	ACGAACAGGCGGAGATGC	
95-3	TGCCTCGGCTATCTGCTTC	1361
96-5	GGCGGTGCTGTAGGAAA	
96-3	ACGAACCGAAGCCCGACA	1197
97-5	AATACGGGCACGGCTCC	
97-3	CTGGAGTTGGCGGAAAGC	470
98-5	CTGGCGTGCTTTGATTGC	
98-3	GGTGGTTCCAACGCTCTATG	4982
99-5	GGCAGCCTTGATTGATTCGG	
99-3	CCCGAGCGTGGACGAAAA	4097
100-5	CGACAGACCGGTCCAGAC	
100-3	GAACAGCGTGGCCGAGAT	10000

Appendix B. Examples of agarose gel pictures for filling gaps between the contigs.

The number showing above each lane indicates the gap number. M: DNA ladder.

