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 *Pseudomonas putida* 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 *Pseudomonas* genomes ....................................... 54
3.5 Analysis of Heat Shock Protein Genes and σ^{32} promoters ....................................... 54

4 Results .................................................................................................................................. 55

4.1 General Genome Features .......................................................................................... 55
4.2 Genes Involved in Plant Growth Promotion and *P. putida* 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
List of Figures

Figure 4-1 Circular genome map of \textit{P. putida} UW4 ..............................................................56

Figure 4-2 Nucleotide sequence of UW4 \textit{dnaA} upstream region analysis........................................63

Figure 4-3 Genomic islands of \textit{P. putida} UW4 ...........................................................................65

Figure 4-4 \textit{P. putida} UW4 predicted protein localizations...............................................................84

Figure 4-5 Pyoverdine synthesis genes in \textit{P. putida} UW4 ..............................................................87

Figure 4-6 \textit{P. putida} UW4 pyoverdine II biosynthesis pathway .......................................................90

Figure 4-7 Amino acids sequence alignment of \textit{pvdYII} in UW4 and \textit{P. aeruginosa} .....................92

Figure 4-8 \textit{P. putida} UW4 IAA biosynthesis pathways ..................................................................95

Figure 4-9 \textit{P. putida} UW4 trehalose biosynthesis pathways .........................................................97

Figure 4-10 \textit{P. putida} UW4 acetoin synthesis pathway.................................................................100

Figure 4-11 \textit{P. putida} UW4 4-hydroxybenzoate synthesis pathway .............................................102

Figure 4-12 \textit{P. putida} UW4 PHA biosynthesis and degradation pathways......................................108

Figure 4-13 Benzoate degradation pathway in \textit{P. putida} UW4 .......................................................110

Figure 4-14 4-hydroxybenzoate and L-quinate degradation pathways in \textit{P. putida} 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 \textit{P. putida} UW4 ............................................121

Figure 4-17 CueAR copper resistance system in \textit{P. putida} UW4 ....................................................124

Figure 4-18 CopABCD copper resistance system in \textit{P. putida} UW4 .............................................127
Figure 4-19 CinQARS copper resistance system in *P. putida* 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 *P. putida* 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 *E. coli* K12 ............................................................................................................. 147

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

*putida* KT2440 ................................................................................................................................... 149

Figure 4-26 Genetic organization of *oscA* in the genome of *P. putida* UW4 ................................. 151

Figure 4-27 Schematic overview of metabolic pathways and transport systems in UW4 ........... 154

Figure 4-28 Phylogenetic tree of 21 different *Pseudomonas* species ................................................. 166

Figure 4-29 Comparative syntenic line plots of the complete six-frame translations of the whole
genome sequences of UW4 with other *P. fluorescens* and *P. putida* genomes .................. 183

Figure 4-30 Maximum likelihood phylogenetic tree of 16S rRNA sequences from completely

sequenced *Pseudomonas* genomes .................................................................................................... 186

Figure 4-31 16S rRNA gene sequences alignment of UW4, KT2440, SBW25, and the type strain

of *P. putida* IAM1236 (D84020) and *P. fluorescens* IAM12022 (D84013) ......................... 188

Figure 4-32 HSPs maximum likelihood trees ...................................................................................... 194

Figure 4-33 *σ*


promoter sequence alignment of HSPs .......................................................................................... 203
List of Tables

Table 1-1 Top Twenty Sequenced Bacterial Genomes as of May 25th, 2012 ........................................ 2

Table 4-1 General Features of *P. putida* UW4 Genome ................................................................... 58

Table 4-2 COG Functional Categories of *P. putida* UW4 ................................................................. 59

Table 4-3 *P. putida* UW4 Pseudogenes ............................................................................................. 60

Table 4-4 *P. putida* UW4 tRNAs ........................................................................................................ 61

Table 4-5 Genomic Islands of *P. putida* UW4 Predicted by IslandViewer ........................................... 67

Table 4-6 *P. putida* UW4 IS Elements ................................................................................................ 72

Table 4-7 *P. putida* UW4 Phage Related CDSs ................................................................................. 73

Table 4-8 Tandem Repeats Identified in *P. putida* UW4 ................................................................. 74

Table 4-9 *P. putida* UW4 Protein Localizations ................................................................................. 83

Table 4-10 Genes Associated With Pyoverdine Synthesis in *P. putida* UW4 .............................. 89

Table 4-11 *P. putida* 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 *P. putida* UW4 ............................ 117

Table 4-14 Protein Secretion Systems in *P. putida* UW4 ............................................................... 159

Table 4-15 Type IV Pilus Genes in *P. putida* UW4 ....................................................................... 165

Table 4-16 Putative Orthologous Relations Between UW4 And Completely Sequenced

*Pseudomonas* genomes .................................................................................................................. 168
Table 4-17 Predicted CDSs That Share Similarities With Other *Pseudomonas* sp..........................169

Table 4-18 Putative Unique CDSs In *P. putida* 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 *Pseudomonas* Genomes.................................................................208
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2ODDs</td>
<td>2-oxoglutarate dependent dioxygenases</td>
</tr>
<tr>
<td>3-HPP</td>
<td>3-hydroxyphenylpropionate</td>
</tr>
<tr>
<td>4-Cl-IAA</td>
<td>4-chloroindole-3-acetic acid</td>
</tr>
<tr>
<td>ABC</td>
<td>ATP-binding-cassette</td>
</tr>
<tr>
<td>ACC</td>
<td>1-aminocyclopropane-1-carboxylate</td>
</tr>
<tr>
<td>AcdB</td>
<td>glycerophosphoryl diester phosphodiesterase</td>
</tr>
<tr>
<td>acdR</td>
<td>ACC deaminase regulatory gene</td>
</tr>
<tr>
<td>acdS</td>
<td>ACC deaminase structural gene</td>
</tr>
<tr>
<td>ACT</td>
<td>artemis comparison tool</td>
</tr>
<tr>
<td>AHLs</td>
<td>acyl homoserine lactones</td>
</tr>
<tr>
<td>AIs</td>
<td>autoinducers</td>
</tr>
<tr>
<td>aLRT</td>
<td>approximate likelihood ratio test</td>
</tr>
<tr>
<td>APS</td>
<td>adenosine-5’-phosphosulfate</td>
</tr>
<tr>
<td>ATs</td>
<td>autotrasporters</td>
</tr>
<tr>
<td>BASys</td>
<td>bacterial annotation system</td>
</tr>
<tr>
<td>CCD</td>
<td>charge-coupled device</td>
</tr>
<tr>
<td>CDS</td>
<td>coding sequence</td>
</tr>
<tr>
<td>COG</td>
<td>cluster of orhologous group</td>
</tr>
<tr>
<td>CRP</td>
<td>cAMP receptor protein</td>
</tr>
<tr>
<td>Cy</td>
<td>cyanine</td>
</tr>
<tr>
<td>dATPαS</td>
<td>deoxyadenosine α-thiotriphosphate</td>
</tr>
<tr>
<td>ddNTP</td>
<td>dideoxynucleotide</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxyribonucleotide</td>
</tr>
<tr>
<td>ECF</td>
<td>extracytoplasmic function</td>
</tr>
<tr>
<td>ent-K</td>
<td>ent-kaurene</td>
</tr>
<tr>
<td>EPS</td>
<td>exopolysaccharides</td>
</tr>
<tr>
<td>FeMoCo</td>
<td>iron molybdenum cofactor</td>
</tr>
<tr>
<td>FNR</td>
<td>fumarate and nitrate reduction protein</td>
</tr>
<tr>
<td>Fur</td>
<td>ferric uptake regulator</td>
</tr>
<tr>
<td>GAs</td>
<td>Gibberellins</td>
</tr>
<tr>
<td>GDH</td>
<td>glucose dehydrogenase</td>
</tr>
<tr>
<td>GGPP</td>
<td>geranyl geranyl diphosphate</td>
</tr>
<tr>
<td>GI</td>
<td>genomic island</td>
</tr>
<tr>
<td>GTR</td>
<td>general time-reversible model</td>
</tr>
<tr>
<td>HCN</td>
<td>hydrogen cyanide</td>
</tr>
<tr>
<td>Hcp</td>
<td>haemolysin coregulated protein</td>
</tr>
<tr>
<td>HMA</td>
<td>heavy-metal-associated</td>
</tr>
<tr>
<td>HR</td>
<td>hypersensitive response</td>
</tr>
<tr>
<td>HSP</td>
<td>heat shock protein</td>
</tr>
<tr>
<td>IAA</td>
<td>indole-3-acetic acid</td>
</tr>
<tr>
<td>IAAld</td>
<td>indole-3-acetaldehyde</td>
</tr>
<tr>
<td>IaaH</td>
<td>IAM hydrolase</td>
</tr>
<tr>
<td>IaaM</td>
<td>tryptophan-2-monooxygenase</td>
</tr>
<tr>
<td>IAM</td>
<td>indole-3-acetamide</td>
</tr>
<tr>
<td>IAN</td>
<td>indole-3-acetonitrile</td>
</tr>
<tr>
<td>IBA</td>
<td>indole-3-butyric acid</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>IpDC</td>
<td>indole-3-pyruvate decarboxylase</td>
</tr>
<tr>
<td>IPyA</td>
<td>indole-3-pyruvate</td>
</tr>
<tr>
<td>IS</td>
<td>insertion sequence</td>
</tr>
<tr>
<td>ISGA</td>
<td>integrative services for genomic analysis</td>
</tr>
<tr>
<td>ISR</td>
<td>induced systemic resistance</td>
</tr>
<tr>
<td>JA</td>
<td>jasmonic acid</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani</td>
</tr>
<tr>
<td>LPS</td>
<td>lipopolysaccharides</td>
</tr>
<tr>
<td>Lrp</td>
<td>leucine-responsive regulatory protein</td>
</tr>
<tr>
<td>MEME</td>
<td>multiple em for motif elicitation</td>
</tr>
<tr>
<td>ML</td>
<td>maximum likelihood</td>
</tr>
<tr>
<td>N₂</td>
<td>dinitrogen</td>
</tr>
<tr>
<td>NCBI</td>
<td>National Center for Biotechnology Information</td>
</tr>
<tr>
<td>NGS</td>
<td>next generation sequencing</td>
</tr>
<tr>
<td>nif</td>
<td>nitrogen fixation</td>
</tr>
<tr>
<td>NR</td>
<td>non-redundant</td>
</tr>
<tr>
<td>NRPS</td>
<td>non-ribosomal peptide synthetases</td>
</tr>
<tr>
<td>oriC</td>
<td>origin of replication</td>
</tr>
<tr>
<td>P</td>
<td>phosphorous</td>
</tr>
<tr>
<td>PAF</td>
<td><em>Pseudomonas</em> Agar F</td>
</tr>
<tr>
<td>PAP2</td>
<td>phosphatidic acid phosphatase</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PGM</td>
<td>personal genome machine</td>
</tr>
<tr>
<td>PGPB</td>
<td>plant growth-promoting bacteria</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>PGPR</td>
<td>plant growth-promoting rhizobacteria</td>
</tr>
<tr>
<td>PPP</td>
<td>pentose phosphate pathway</td>
</tr>
<tr>
<td>PQQ</td>
<td>pyrroloquinoline quinone</td>
</tr>
<tr>
<td>PHAs</td>
<td>Polyhydroxyalkanoates</td>
</tr>
<tr>
<td>PHB</td>
<td>poly-hydroxybutyrate</td>
</tr>
<tr>
<td>Pre-Q&lt;sub&gt;0&lt;/sub&gt;</td>
<td>7-cyano-7-deazaguanine</td>
</tr>
<tr>
<td>Prodigal</td>
<td>prokaryotic dynamic programming genefinding algorithm</td>
</tr>
<tr>
<td>PSB</td>
<td>phosphorous solubilizing bacteria</td>
</tr>
<tr>
<td>QS</td>
<td>Quorum sensing</td>
</tr>
<tr>
<td>RNAP</td>
<td>RNA polymerase</td>
</tr>
<tr>
<td>rnm</td>
<td>ribosomal RNA</td>
</tr>
<tr>
<td>SA</td>
<td>salicylic acid</td>
</tr>
<tr>
<td>SAM</td>
<td>S-adenosylmethionine</td>
</tr>
<tr>
<td>SAR</td>
<td>systemic acquired resistance</td>
</tr>
<tr>
<td>SBL</td>
<td>sequencing by ligation</td>
</tr>
<tr>
<td>SBS</td>
<td>sequencing by synthesis</td>
</tr>
<tr>
<td>Sec</td>
<td>general secretory pathway</td>
</tr>
<tr>
<td>SMRT</td>
<td>single molecule real time</td>
</tr>
<tr>
<td>SOLiD</td>
<td>support oligonucleotide ligation detection</td>
</tr>
<tr>
<td>T1SS</td>
<td>type I secretion system</td>
</tr>
<tr>
<td>T2SS</td>
<td>type II secretion system</td>
</tr>
<tr>
<td>T3SS</td>
<td>type III secretion system</td>
</tr>
<tr>
<td>T5SS</td>
<td>type V secretion system</td>
</tr>
<tr>
<td>T6SS</td>
<td>type VI secretion system</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>TAM</td>
<td>tryptamine</td>
</tr>
<tr>
<td>Tat</td>
<td>twin arginine translocation</td>
</tr>
<tr>
<td>TCA</td>
<td>tricarboxylic acid</td>
</tr>
<tr>
<td>TPS</td>
<td>two-partner secretion system</td>
</tr>
<tr>
<td>TSB</td>
<td>tryptic soy broth</td>
</tr>
<tr>
<td>TSO</td>
<td>tryptophan side-chain oxidase</td>
</tr>
<tr>
<td>VgrG</td>
<td>valine-glycine repeat protein</td>
</tr>
<tr>
<td>ZMW</td>
<td>zero-mode waveguide</td>
</tr>
</tbody>
</table>
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 25th, 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>
<thead>
<tr>
<th>Organisms</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Escherichia coli</em></td>
<td>60</td>
</tr>
<tr>
<td><em>Chlamydia trachomatis</em></td>
<td>59</td>
</tr>
<tr>
<td><em>Bacillus cereus</em></td>
<td>40</td>
</tr>
<tr>
<td><em>Helicobacter pylori</em></td>
<td>39</td>
</tr>
<tr>
<td><em>Staphylococcus aureus</em></td>
<td>33</td>
</tr>
<tr>
<td><em>Salmonella enterica</em></td>
<td>31</td>
</tr>
<tr>
<td><em>Streptococcus pneumoniae</em></td>
<td>24</td>
</tr>
<tr>
<td><em>Mycobacterium tuberculosis</em></td>
<td>21</td>
</tr>
<tr>
<td><em>Bacillus thuringiensis</em></td>
<td>19</td>
</tr>
<tr>
<td><em>Clostridium difficile</em></td>
<td>18</td>
</tr>
<tr>
<td><em>Listeria monocytogenes</em></td>
<td>16</td>
</tr>
<tr>
<td><em>Neisseria meningitidis</em></td>
<td>16</td>
</tr>
<tr>
<td><em>Streptococcus pyogenes</em></td>
<td>16</td>
</tr>
<tr>
<td><em>Clostridium beijerinckii</em></td>
<td>15</td>
</tr>
<tr>
<td><em>Buchnera aphidicola</em></td>
<td>14</td>
</tr>
<tr>
<td><em>Corynebacterium pseudotuberculosis</em></td>
<td>14</td>
</tr>
<tr>
<td><em>Rickettsia prowazekii</em></td>
<td>14</td>
</tr>
<tr>
<td><em>Sulfolobus islandicus</em></td>
<td>14</td>
</tr>
<tr>
<td><em>Acinetobacter baumannii</em></td>
<td>13</td>
</tr>
<tr>
<td><em>Corynebacterium diphtheriae</em></td>
<td>13</td>
</tr>
<tr>
<td><em>Streptococcus suis</em></td>
<td>13</td>
</tr>
</tbody>
</table>
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 ~ 400-1200 mg·kg⁻¹ (Rodríguez and Fraga 1999). However, soluble P that is available for plants under most soil conditions is only ~ 1 mg·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, phosphodiesters, 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).


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; Kpomblekou 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⁺ 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 Eσ^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, EσRNA 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 EσRNA at the -10 region due to the formation of a PhoP~P-RNA polymerase 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; Leaungvutiviroj 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$_2$). 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$_2$ 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).

\[ \text{N}_2 + 8\text{e}^- + 8\text{H}^+ + 16\text{MgATP} \rightarrow 2\text{NH}_3 + \text{H}_2 + 16\text{MgADP} + 16\text{P}_i \]

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 ($\text{nif}$) 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 $\text{nif}$ 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 $\text{nifA, nifL, nifD, nifK, nifH, nifF, nifJ, nifQ, nifB, nifN, nifE, nifV, nifM, nifS, nifW, nifZ, nifT, nifY, nifU,}$ and $\text{nifX}$. NifA is a positive regulatory protein, which turns on the transcription of all of the $\text{nif}$ 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 α and β 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 H₂. This can be achieved by introducing hydrogenase gene to a diazotroph. Hydrogenase can take up H₂ 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. vinelandii. A. vinelandii 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. vinelandii (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 \( \text{Fe(OH)}_3=10^{-39} \)) 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 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 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 Fe\(^{3+}\) to 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\textsuperscript{5}-oxygenase that is responsible for converting L-ornithine to L-N\textsuperscript{5}-hydroxyornithine (Ge and Seah 2006), which can be subsequently converted to L-N\textsuperscript{5}-formyl-N\textsuperscript{5}-hydroxyornithine by the product of pvdF (McMorran et al. 2001). Then PvdI and PvdJ can incorporate the substrate L-N\textsuperscript{5}-formyl-N\textsuperscript{5}-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\textsuperscript{2+} 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 \textit{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$^{2+}$ 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-diacyetyl 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). Zwittermicin 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 β-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 β-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 endoprotease 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.
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. 1993; Sticher et al. 1997; Pieterse et al. 1998).
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-monoxygenase (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; Ortiz-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.*
*Arabidopsis 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$_1$, GA$_3$, GA$_4$, GA$_9$, GA$_{19}$ and GA$_{20}$ 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 C20 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 2ODD-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 α-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 α-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 Archaebacteria 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 μm in length and 2 μm 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 μg/ml), erythromycin (64 μg/ml), novobiocin (256 μg/ml), but is sensitive to kanamycin (4 μg/ml) and tetracycline (8 μg/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* DH5α, *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’ α-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 α-thiotriphosphate (dATPαS), 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 µm 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 (Chaiet 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 Ix, 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 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 AgNO₃, 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 φ29 DNA polymerase is chosen for this sequencing platform because it can incorporate phospholinked dNTPs efficiently into the growing strand. In addition, φ29 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 N6-methyladenine, 5-methylcytosine, 5-hydroxymethylcytosine, and N4-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, O6-methylguanine, 1-methyladenine, O4-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 × 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 × 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 σ^{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® Genomic DNA purification kit (Promega, Madison, WI, USA) according to the manufacturer’s instructions. *E. coli* DH5α (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 µg/mL; tetracycline, 15 µg/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 µg 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× 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 µg 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 ≤1 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 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 σ^{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, σ^{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

<table>
<thead>
<tr>
<th>Features</th>
<th>Chromosome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size (bp)</td>
<td>6,183,388</td>
</tr>
<tr>
<td>G+C content (%)</td>
<td>60.05</td>
</tr>
<tr>
<td>Number of CDSs</td>
<td>5431</td>
</tr>
<tr>
<td>Total CDSs size (bp)</td>
<td>5,406,063</td>
</tr>
<tr>
<td>Coding %</td>
<td>87.4</td>
</tr>
<tr>
<td>Average CDS length (nt)</td>
<td>995</td>
</tr>
<tr>
<td>Pseudogenes</td>
<td>17</td>
</tr>
<tr>
<td>tRNAs</td>
<td>72</td>
</tr>
<tr>
<td>rRNA genes (clusters)</td>
<td>22 (7)</td>
</tr>
<tr>
<td>Number of genes with assigned function</td>
<td>4154 (76%)</td>
</tr>
<tr>
<td>Number of genes without assigned function</td>
<td>1277 (24%)</td>
</tr>
<tr>
<td>Number of predicted enzymes</td>
<td>1551 (29%)</td>
</tr>
</tbody>
</table>
Table 4-2. COG Functional Categories of *P. putida* UW4

<table>
<thead>
<tr>
<th>Functional Category</th>
<th>UW4</th>
</tr>
</thead>
<tbody>
<tr>
<td>A RNA processing and modification</td>
<td>1</td>
</tr>
<tr>
<td>B Chromatin structure and dynamics</td>
<td>3</td>
</tr>
<tr>
<td>C Energy production and conversion</td>
<td>288</td>
</tr>
<tr>
<td>D Cell cycle control, cell division, chromosome partitioning</td>
<td>37</td>
</tr>
<tr>
<td>E Amino acid transport and metabolism</td>
<td>500</td>
</tr>
<tr>
<td>F Nucleotide transport and metabolism</td>
<td>90</td>
</tr>
<tr>
<td>G Carbohydrate transport and metabolism</td>
<td>232</td>
</tr>
<tr>
<td>H Coenzyme transport and metabolism</td>
<td>156</td>
</tr>
<tr>
<td>I Lipid transport and metabolism</td>
<td>255</td>
</tr>
<tr>
<td>J Translation, ribosomal structure and biogenesis</td>
<td>174</td>
</tr>
<tr>
<td>K Transcription</td>
<td>431</td>
</tr>
<tr>
<td>L Replication, recombination and repair</td>
<td>177</td>
</tr>
<tr>
<td>M Cell wall/membrane/envelope biogenesis</td>
<td>262</td>
</tr>
<tr>
<td>N Cell motility</td>
<td>114</td>
</tr>
<tr>
<td>O Posttranslational modification, protein turnover, chaperones</td>
<td>179</td>
</tr>
<tr>
<td>P Inorganic ion transport and metabolism</td>
<td>229</td>
</tr>
<tr>
<td>Q Secondary metabolites biosynthesis, transport and catabolism</td>
<td>102</td>
</tr>
<tr>
<td>R General function prediction only</td>
<td>440</td>
</tr>
<tr>
<td>S Function unknown</td>
<td>373</td>
</tr>
<tr>
<td>T Signal transduction mechanisms</td>
<td>231</td>
</tr>
<tr>
<td>U Intracellular trafficking, secretion, and vesicular transport</td>
<td>43</td>
</tr>
<tr>
<td>V Defense mechanisms</td>
<td>62</td>
</tr>
<tr>
<td>- Not in COGs</td>
<td>1052</td>
</tr>
<tr>
<td>Total</td>
<td>5431</td>
</tr>
</tbody>
</table>
### Table 4-3. *P. putida* UW4 Pseudogenes

<table>
<thead>
<tr>
<th>locus_tag</th>
<th>product</th>
<th>pseudo</th>
</tr>
</thead>
<tbody>
<tr>
<td>PputUW4_2113</td>
<td>long-chain fatty acid transporter</td>
<td>C-terminus missing fragment</td>
</tr>
<tr>
<td>PputUW4_2521</td>
<td>AraC family transcriptional regulator</td>
<td>C-terminus missing fragment</td>
</tr>
<tr>
<td>PputUW4_3551</td>
<td>IS630 family transposase, truncated</td>
<td>C-terminus missing fragment</td>
</tr>
<tr>
<td>PputUW4_5207</td>
<td>glutamine amidotransferase</td>
<td>C-terminus missing fragment</td>
</tr>
<tr>
<td>PputUW4_0114</td>
<td>von Willebrand factor, type A</td>
<td>N-terminus missing fragment</td>
</tr>
<tr>
<td>PputUW4_1252</td>
<td>methyltransferase</td>
<td>N-terminus missing fragment</td>
</tr>
<tr>
<td>PputUW4_2026</td>
<td>ATP-dependent DNA ligase LigD</td>
<td>N-terminus missing fragment</td>
</tr>
<tr>
<td>PputUW4_2127</td>
<td>IS30 family transposase</td>
<td>N-terminus missing fragment</td>
</tr>
<tr>
<td>PputUW4_2373</td>
<td>IclR family transcriptional regulator</td>
<td>N-terminus missing fragment</td>
</tr>
<tr>
<td>PputUW4_2519</td>
<td>xylulose kinase</td>
<td>N-terminus missing fragment</td>
</tr>
<tr>
<td>PputUW4_3361</td>
<td>phage integrase</td>
<td>N-terminus missing fragment</td>
</tr>
<tr>
<td>PputUW4_3383</td>
<td>LuxR family transcriptional regulator</td>
<td>N-terminus missing fragment</td>
</tr>
<tr>
<td>PputUW4_4664</td>
<td>type IV pilus-associated protein</td>
<td>N-terminus missing fragment</td>
</tr>
<tr>
<td>PputUW4_1487</td>
<td>transcriptional regulator</td>
<td>N- and C-terminus missing fragment</td>
</tr>
<tr>
<td>PputUW4_2143</td>
<td>hypothetical protein</td>
<td>N- and C-terminus missing fragment</td>
</tr>
<tr>
<td>PputUW4_2427</td>
<td>LysR family transcriptional regulator</td>
<td>N- and C-terminus missing fragment</td>
</tr>
<tr>
<td>PputUW4_4694</td>
<td>glutamate dehydrogenase</td>
<td>N- and C-terminus missing fragment</td>
</tr>
<tr>
<td>tRNA</td>
<td>Number</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Cys</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Gln</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Met</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Sec</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Trp</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>
(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.
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>
<thead>
<tr>
<th>Genomic Islands</th>
<th>Size (bp)</th>
<th>Locus ID</th>
<th>Product</th>
<th>GC content</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 743,662-749,401 5,739</td>
<td>PputUW4_0641</td>
<td>ATPase</td>
<td>53.94</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PputUW4_0642</td>
<td>McrBC 5-methylcytosine restriction system component-like protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PputUW4_0643</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PputUW4_0644</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PputUW4_0645</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PputUW4_0646</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PputUW4_0647</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PputUW4_0648</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 780,455-791,108 10,653</td>
<td>PputUW4_0666</td>
<td>hypothetical protein</td>
<td>55.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PputUW4_0667</td>
<td>putative helicase</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PputUW4_0668</td>
<td>DNA helicase</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PputUW4_0669</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PputUW4_0670</td>
<td>phage integrase</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PputUW4_0671</td>
<td>chemotaxis protein MotB-related protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PputUW4_0672</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 797,778-809,172 11,394</td>
<td>PputUW4_0675</td>
<td>SNF2 family DNA/RNA helicase</td>
<td>54.74</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PputUW4_0676</td>
<td>5-methylcytosine-specific restriction protein C</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PputUW4_0677</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PputUW4_0678</td>
<td>Phage integrase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 1,078,659-1,088,425 9,766</td>
<td>PputUW4_0917</td>
<td>NAD-dependent formate dehydrogenase gamma subunit</td>
<td>52.94</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PputUW4_0918</td>
<td>type III restriction protein res subunit</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PputUW4_0919</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PputUW4_1368</td>
<td>putative lipoprotein</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PputUW4_1369</td>
<td>Integration host factor subunit beta</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 1,598,659-1,609,449 10,790</td>
<td>PputUW4_1370</td>
<td>chain length determinant family protein</td>
<td>44.61</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PputUW4_1371</td>
<td>acetyltransferase</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PputUW4_1372</td>
<td>lipopolysaccharide biosynthesis protein RffA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gene ID</td>
<td>Product Description</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>---------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_1373</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_1374</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_1375</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_1376</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_1377</td>
<td>glycosyl transferases group 1 family protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_1378</td>
<td>dTDP-4-dehydrorhamnose reductase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2095</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2096</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2097</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2098</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2099</td>
<td>diguanylate cyclase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2100</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2101</td>
<td>acetyltransferase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2102</td>
<td>MarR family transcriptional regulator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2103</td>
<td>AraC family transcriptional regulator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2118</td>
<td>quinone oxidoreductase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2119</td>
<td>methionine sulfoxide reductase A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2120</td>
<td>RND family efflux transporter, MFP subunit</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2121</td>
<td>ABC transporter family protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2122</td>
<td>peptide ABC transporter permease</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2123</td>
<td>peptide ABC transporter permease</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2124</td>
<td>IS1182 family transposase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2125</td>
<td>ThiF family protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2126</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2127</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2128</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2129</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2130</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2131</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2132</td>
<td>IS1182 family transposase, truncated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2133</td>
<td>IS1182 family transposase, truncated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2134</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gene ID</td>
<td>Description</td>
<td>Start</td>
<td>End</td>
<td>Length</td>
</tr>
<tr>
<td>------------</td>
<td>--------------------------------------------------</td>
<td>-------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>PputUW4_2135</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2136</td>
<td>IS1182 family transposase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2137</td>
<td>polysaccharide deacetylase family protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2138</td>
<td>GCN5-related N-acetyltransferase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2139</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2140</td>
<td>HxIR family transcriptional regulator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2141</td>
<td>nitrilase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2142</td>
<td>glutathione-dependent formaldehyde-activatingGFA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2167</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2168</td>
<td>integrase family protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2169</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2170</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2171</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_219</td>
<td>cytosine/purines uracil thiamine allantoinpermease</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_220</td>
<td>histone deacetylase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_221</td>
<td>LysR family transcriptional regulator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_222</td>
<td>lysine exporter protein LysE/YggA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_234</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_235</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2600</td>
<td>2,4-dihydroxyacetophenone dioxygenase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2601</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2602</td>
<td>short chain dehydrogenase family protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2603</td>
<td>thiolase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2604</td>
<td>AMP-dependent synthetase and ligase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2652</td>
<td>LysR family transcriptional regulator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2653</td>
<td>amino acid transporter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2654</td>
<td>cystathionine gamma-synthase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2655</td>
<td>AraC family transcriptional regulator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chromosome Position</td>
<td>Start (bp)</td>
<td>End (bp)</td>
<td>Length (bp)</td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>------------</td>
<td>----------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>3,176,210</td>
<td>3,186,955</td>
<td>10,745</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>3,737,337</td>
<td>3,743,674</td>
<td>6,337</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>3,966,752</td>
<td>3,970,895</td>
<td>4,143</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>4,116,806</td>
<td>4,123,634</td>
<td>6,828</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>4,159,348</td>
<td>4,182,589</td>
<td>23,241</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Description</th>
<th>Location</th>
<th>Proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>PputUW4_2727</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2728</td>
<td>CadC family transcriptional regulator</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2729</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2730</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2731</td>
<td>AraC family transcriptional regulator</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2732</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2733</td>
<td>OmpA-like transmembrane domain protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_2734</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3195</td>
<td>two component LuxR family transcriptional regulator</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3196</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3197</td>
<td>IS110 family transposase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3198</td>
<td>XRE family transcriptional regulator</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3199</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3200</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3201</td>
<td>3-oxoadipate enol-lactonase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3202</td>
<td>aldehyde dehydrogenase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3416</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3417</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3418</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3419</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3420</td>
<td>flagella synthesis regulator FleN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3421</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3545</td>
<td>glycerol-3-phosphate cytidyltransferase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3546</td>
<td>polysaccharide biosynthesis protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3547</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3548</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3549</td>
<td>glucosyltransferase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3550</td>
<td>glucosyltransferase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3551</td>
<td>integrase family protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3552</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3553</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3554</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3555</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3556</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3557</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3558</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3559</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3590</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3591</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3592</td>
<td>transposon resolvase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3593</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3594</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3595</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3596</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3597</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3598</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3599</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3600</td>
<td>NAD-dependent DNA ligase LigA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3601</td>
<td>cell division protein ZipA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3602</td>
<td>chromosome partition protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3603</td>
<td>GntR family transcriptional regulator</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3604</td>
<td>xanthine dehydrogenase small subunit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_3605</td>
<td>xanthine dehydrogenase, molybdopterin-bindingsubunit B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_4192</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_4193</td>
<td>cytosine-specific methyltransferase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_4194</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_4195</td>
<td>histidine kinase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_4196</td>
<td>hypothetical protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_4197</td>
<td>5-methylcytosine-specific restriction enzyme</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_4198</td>
<td>DNA mismatch endonuclease Vsr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_4611</td>
<td>paraquat-inducible protein A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_4612</td>
<td>paraquat-inducible protein A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PputUW4_4613</td>
<td>sulfite oxidase subunit YedZ</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

18 4,818,892 4,826,804 7,912 44.48

19 5,281,019 5,288,766 7,748 52.04
<table>
<thead>
<tr>
<th>Family/Group</th>
<th>P. putida UW4 CDS ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS110</td>
<td>PputUW4_0474 PputUW4_0841 PputUW4 2025 PputUW4 4591 PputUW4 5362</td>
</tr>
<tr>
<td>IS110/IS1111</td>
<td>PputUW4_0318 PputUW4_1254 PputUW4_1997 PputUW4_3132 PputUW4_3197 PputUW4_3451 PputUW4_4228 PputUW4_4445 PputUW4_5136 PputUW4_5212 PputUW4_5335</td>
</tr>
<tr>
<td>IS1182</td>
<td>PputUW4_1560 PputUW4_2124 PputUW4_2136 PputUW4_2164 PputUW4_2166 PputUW4_2215 PputUW4_2223</td>
</tr>
<tr>
<td>IS3/IS3</td>
<td>PputUW4_0151 PputUW4_1502 PputUW4_1775 PputUW4_2015 PputUW4_3143 PputUW4_3925 PputUW4_4241 PputUW4_4349</td>
</tr>
<tr>
<td>Truncated</td>
<td></td>
</tr>
<tr>
<td>IS110/IS1111</td>
<td>PputUW4_3001</td>
</tr>
<tr>
<td>IS1182</td>
<td>PputUW4_2132 PputUW4_2133</td>
</tr>
<tr>
<td>IS30</td>
<td>PputUW4_2127</td>
</tr>
<tr>
<td>IS5/IS427</td>
<td>PputUW4_3360</td>
</tr>
<tr>
<td>IS630</td>
<td>PputUW4_3551</td>
</tr>
</tbody>
</table>
Table 4-7. *P. putida* UW4 Phage Related CDSs

<table>
<thead>
<tr>
<th>locus_tag</th>
<th>product</th>
</tr>
</thead>
<tbody>
<tr>
<td>PputUW4_0456</td>
<td>HflK protein</td>
</tr>
<tr>
<td>PputUW4_0457</td>
<td>HflC protein</td>
</tr>
<tr>
<td>PputUW4_0670</td>
<td>phage integrase</td>
</tr>
<tr>
<td>PputUW4_0678</td>
<td>phage integrase</td>
</tr>
<tr>
<td>PputUW4_1008</td>
<td>GtrA family protein</td>
</tr>
<tr>
<td>PputUW4_1009</td>
<td>bactoprenol glucosyl transferase</td>
</tr>
<tr>
<td>PputUW4_1029</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1125</td>
<td>phage repressor</td>
</tr>
<tr>
<td>PputUW4_1558</td>
<td>phage integrase family site specific recombinase</td>
</tr>
<tr>
<td>PputUW4_1764</td>
<td>phage repressor</td>
</tr>
<tr>
<td>PputUW4_1775</td>
<td>integrase catalytic protein</td>
</tr>
<tr>
<td>PputUW4_2419</td>
<td>cointegrate resolution protein T</td>
</tr>
<tr>
<td>PputUW4_2420</td>
<td>cointegrate resolution protein S</td>
</tr>
<tr>
<td>PputUW4_2781</td>
<td>phage integrase</td>
</tr>
<tr>
<td>PputUW4_3269</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3361</td>
<td>phage integrase</td>
</tr>
<tr>
<td>PputUW4_3592</td>
<td>transposon resolvase</td>
</tr>
<tr>
<td>PputUW4_4196</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_5301</td>
<td>prophage antirepressor</td>
</tr>
</tbody>
</table>
Table 4-8. Tandem Repeats Identified in *P. putida* UW4

<table>
<thead>
<tr>
<th>Start</th>
<th>End</th>
<th>Period Size</th>
<th>Consensus Size</th>
<th>Copy Number</th>
<th>Affected Gene or Region</th>
<th>Gene Product</th>
<th>Repeats Relative to CDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>43924</td>
<td>43964</td>
<td>21</td>
<td>22</td>
<td>1.9</td>
<td>41 oligopeptidase A</td>
<td>inside</td>
<td>inside</td>
</tr>
<tr>
<td>121176</td>
<td>121255</td>
<td>27</td>
<td>28</td>
<td>3</td>
<td>114 (pseudo) von Willebrand factor, type A</td>
<td>inside</td>
<td>inside</td>
</tr>
<tr>
<td>121830</td>
<td>121859</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>114 (pseudo) von Willebrand factor, type A</td>
<td>inside</td>
<td>inside</td>
</tr>
<tr>
<td>124695</td>
<td>129038</td>
<td>318</td>
<td>317</td>
<td>13.7</td>
<td>114 (pseudo), 115, 116 conserved repeat domain protein; conserved repeat domain protein</td>
<td>inside</td>
<td>inside</td>
</tr>
<tr>
<td>129077</td>
<td>129901</td>
<td>300</td>
<td>300</td>
<td>2.8</td>
<td>116 conserved repeat domain protein</td>
<td>inside</td>
<td>inside</td>
</tr>
<tr>
<td>130252</td>
<td>130291</td>
<td>6</td>
<td>6</td>
<td>6.7</td>
<td>116 conserved repeat domain protein</td>
<td>inside</td>
<td>inside</td>
</tr>
<tr>
<td>448990</td>
<td>449029</td>
<td>21</td>
<td>21</td>
<td>1.9</td>
<td>396 dihydrolipoamide acetyltransferase</td>
<td>inside</td>
<td>inside</td>
</tr>
<tr>
<td>571073</td>
<td>571412</td>
<td>84</td>
<td>84</td>
<td>4</td>
<td>497 hypothetical protein</td>
<td>inside</td>
<td>inside</td>
</tr>
<tr>
<td>634896</td>
<td>634947</td>
<td>15</td>
<td>14</td>
<td>3.7</td>
<td>555 acetyl-CoA carboxylase biotin carboxyl carrier protein subunit</td>
<td>inside</td>
<td>inside</td>
</tr>
<tr>
<td>634923</td>
<td>634952</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>555 acetyl-CoA carboxylase biotin carboxyl carrier protein subunit</td>
<td>inside</td>
<td>inside</td>
</tr>
<tr>
<td>1058842</td>
<td>1059321</td>
<td>123</td>
<td>123</td>
<td>3.9</td>
<td>903 hypothetical protein</td>
<td>inside</td>
<td>inside</td>
</tr>
<tr>
<td>1237244</td>
<td>1237308</td>
<td>21</td>
<td>20</td>
<td>3.1</td>
<td>1043 methyl-accepting chemotaxis protein</td>
<td>inside</td>
<td>inside</td>
</tr>
<tr>
<td>1373898</td>
<td>1373930</td>
<td>16</td>
<td>17</td>
<td>2</td>
<td>1170 hypothetical protein</td>
<td>inside</td>
<td>inside</td>
</tr>
<tr>
<td>1408421</td>
<td>1408452</td>
<td>12</td>
<td>12</td>
<td>2.7</td>
<td>1199 hypothetical protein</td>
<td>inside</td>
<td>inside</td>
</tr>
<tr>
<td>1479472</td>
<td>1479511</td>
<td>15</td>
<td>15</td>
<td>2.7</td>
<td>1266 hypothetical protein</td>
<td>inside</td>
<td>inside</td>
</tr>
<tr>
<td>1479931</td>
<td>1480007</td>
<td>36</td>
<td>36</td>
<td>2.1</td>
<td>1267 hypothetical protein</td>
<td>inside</td>
<td>inside</td>
</tr>
<tr>
<td>Start Position</td>
<td>End Position</td>
<td>Coverage</td>
<td>Exon Prediction</td>
<td>Protein Name</td>
<td>Location</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>--------------</td>
<td>----------</td>
<td>-----------------</td>
<td>--------------------------------------------------</td>
<td>----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1807605</td>
<td>1807634</td>
<td>15</td>
<td></td>
<td>Non-ribosomal peptide synthetase</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1865671</td>
<td>1865696</td>
<td>12</td>
<td></td>
<td>Hpt sensor hybrid histidine kinase</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1920541</td>
<td>1921009</td>
<td>39</td>
<td></td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1939911</td>
<td>1939955</td>
<td>21</td>
<td></td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1940032</td>
<td>1940078</td>
<td>24</td>
<td></td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2029994</td>
<td>2030045</td>
<td>27</td>
<td></td>
<td>heme peroxidase</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2030037</td>
<td>2030826</td>
<td>273</td>
<td></td>
<td>heme peroxidase</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2030234</td>
<td>2030299</td>
<td>27</td>
<td></td>
<td>heme peroxidase</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2030533</td>
<td>2030592</td>
<td>27</td>
<td></td>
<td>heme peroxidase</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2067404</td>
<td>2067444</td>
<td>20</td>
<td></td>
<td>AFG1-like ATPase</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2222958</td>
<td>2223231</td>
<td>39</td>
<td></td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2248389</td>
<td>2248418</td>
<td>15</td>
<td></td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2248428</td>
<td>2248469</td>
<td>15</td>
<td></td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2248458</td>
<td>2248521</td>
<td>30</td>
<td></td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2259311</td>
<td>2259560</td>
<td>84</td>
<td></td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2449938</td>
<td>2450234</td>
<td>33</td>
<td></td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2450214</td>
<td>2450295</td>
<td>18</td>
<td></td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2450218</td>
<td>2450303</td>
<td>36</td>
<td></td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2532903</td>
<td>2532935</td>
<td>6</td>
<td></td>
<td>cyclic beta 1-2 glucan synthetase</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2544850</td>
<td>2544899</td>
<td>18</td>
<td></td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2544895</td>
<td>2544959</td>
<td>9</td>
<td></td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2562691</td>
<td>2563936</td>
<td>276</td>
<td></td>
<td>sensor histidine kinase</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2600012</td>
<td>2600043</td>
<td>16</td>
<td></td>
<td>RipR family transcriptional regulator</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2665103</td>
<td>2665175</td>
<td>27</td>
<td></td>
<td>hemolysin-type calcium-binding region</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Start Position</td>
<td>End Position</td>
<td>Length</td>
<td>Score</td>
<td>Description</td>
<td>Inside Location</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>--------------</td>
<td>--------</td>
<td>-------</td>
<td>---------------------------------------------------------------</td>
<td>-----------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2665366</td>
<td>2665503</td>
<td>54</td>
<td>2.6</td>
<td>hemolysin-type calcium-binding region</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2665925</td>
<td>2665996</td>
<td>27</td>
<td>2.7</td>
<td>hemolysin-type calcium-binding region</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2682575</td>
<td>2682603</td>
<td>14</td>
<td>2</td>
<td>3-oxoacyl-[acyl-carrier-protein] reductase</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2801188</td>
<td>2801654</td>
<td>123</td>
<td>3.8</td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2975571</td>
<td>2975623</td>
<td>24</td>
<td>2.2</td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2975578</td>
<td>2975624</td>
<td>24</td>
<td>2</td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3112030</td>
<td>3112056</td>
<td>12</td>
<td>2.2</td>
<td>cation efflux family protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3144781</td>
<td>3144819</td>
<td>18</td>
<td>2.4</td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3154032</td>
<td>3154065</td>
<td>15</td>
<td>2.3</td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3176471</td>
<td>3176601</td>
<td>67</td>
<td>1.9</td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3510124</td>
<td>3510909</td>
<td>300</td>
<td>2.6</td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3511360</td>
<td>3512139</td>
<td>375</td>
<td>2.1</td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3641058</td>
<td>3641162</td>
<td>48</td>
<td>2.2</td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3641142</td>
<td>3641185</td>
<td>18</td>
<td>2.4</td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3644694</td>
<td>3645032</td>
<td>126</td>
<td>2.7</td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3644757</td>
<td>3644959</td>
<td>63</td>
<td>3.2</td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3738308</td>
<td>3738659</td>
<td>78</td>
<td>4.5</td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3741377</td>
<td>3741429</td>
<td>27</td>
<td>2</td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3741468</td>
<td>3741516</td>
<td>24</td>
<td>2</td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3846098</td>
<td>3846150</td>
<td>15</td>
<td>3.5</td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3846136</td>
<td>3846204</td>
<td>18</td>
<td>3.7</td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3846167</td>
<td>3846212</td>
<td>21</td>
<td>2.2</td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3846236</td>
<td>3846265</td>
<td>9</td>
<td>3.3</td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3846422</td>
<td>3846586</td>
<td>36</td>
<td>4.6</td>
<td>hypothetical protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4050739</td>
<td>4050793</td>
<td>24</td>
<td>2.3</td>
<td>copper-resistance protein CopA</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4051392</td>
<td>4051496</td>
<td>39</td>
<td>2.7</td>
<td>copper resistance protein</td>
<td>inside</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

76
<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Function</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>4155899</td>
<td>hypothetical protein inside</td>
<td></td>
</tr>
<tr>
<td>4165503</td>
<td>hypothetical protein inside</td>
<td></td>
</tr>
<tr>
<td>4216459</td>
<td>hypothetical protein inside</td>
<td></td>
</tr>
<tr>
<td>4394420</td>
<td>dihydrolipoamide succinyltransferase inside</td>
<td></td>
</tr>
<tr>
<td>4421694</td>
<td>urea transporter inside</td>
<td></td>
</tr>
<tr>
<td>4449929</td>
<td>chemotaxis sensor histidine kinase CheA inside</td>
<td></td>
</tr>
<tr>
<td>4469374</td>
<td>flagellar assembly protein H inside</td>
<td></td>
</tr>
<tr>
<td>4531905</td>
<td>TrkH family potassium uptake protein inside</td>
<td></td>
</tr>
<tr>
<td>4560584</td>
<td>ABC transporter, transmembrane region inside</td>
<td></td>
</tr>
<tr>
<td>4654006</td>
<td>ribonuclease E inside</td>
<td></td>
</tr>
<tr>
<td>4654015</td>
<td>ribonuclease E inside</td>
<td></td>
</tr>
<tr>
<td>4654106</td>
<td>ribonuclease E inside</td>
<td></td>
</tr>
<tr>
<td>4718832</td>
<td>GTP diphosphokinase inside</td>
<td></td>
</tr>
<tr>
<td>4767521</td>
<td>hypothetical protein inside</td>
<td></td>
</tr>
<tr>
<td>4769292</td>
<td>hypothetical protein inside</td>
<td></td>
</tr>
<tr>
<td>4769427</td>
<td>hypothetical protein inside</td>
<td></td>
</tr>
<tr>
<td>4785020</td>
<td>hypothetical protein inside</td>
<td></td>
</tr>
<tr>
<td>4850751</td>
<td>lysine exporter protein LysE/YggA inside</td>
<td></td>
</tr>
<tr>
<td>4930376</td>
<td>TolA colicin import membrane protein inside</td>
<td></td>
</tr>
<tr>
<td>4930413</td>
<td>TolA colicin import membrane protein inside</td>
<td></td>
</tr>
<tr>
<td>4930462</td>
<td>TolA colicin import membrane protein inside</td>
<td></td>
</tr>
<tr>
<td>Start</td>
<td>End</td>
<td>Gene Length</td>
</tr>
<tr>
<td>----------</td>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>4930518</td>
<td>4930582</td>
<td>24</td>
</tr>
<tr>
<td>4930537</td>
<td>4930591</td>
<td>24</td>
</tr>
<tr>
<td>4930623</td>
<td>4930735</td>
<td>18</td>
</tr>
<tr>
<td>5038760</td>
<td>5038790</td>
<td>9</td>
</tr>
<tr>
<td>5316884</td>
<td>5316909</td>
<td>12</td>
</tr>
<tr>
<td>5439683</td>
<td>5439712</td>
<td>15</td>
</tr>
<tr>
<td>5439985</td>
<td>5440013</td>
<td>12</td>
</tr>
<tr>
<td>5550454</td>
<td>5550510</td>
<td>27</td>
</tr>
<tr>
<td>5727150</td>
<td>5727183</td>
<td>12</td>
</tr>
<tr>
<td>5797511</td>
<td>5797551</td>
<td>21</td>
</tr>
<tr>
<td>5803893</td>
<td>5803923</td>
<td>12</td>
</tr>
<tr>
<td>5821054</td>
<td>5821090</td>
<td>12</td>
</tr>
<tr>
<td>5864843</td>
<td>5864867</td>
<td>12</td>
</tr>
<tr>
<td>5914851</td>
<td>5915078</td>
<td>48</td>
</tr>
<tr>
<td>5919408</td>
<td>5919444</td>
<td>15</td>
</tr>
<tr>
<td>5971992</td>
<td>5972025</td>
<td>15</td>
</tr>
<tr>
<td>5997825</td>
<td>5997883</td>
<td>30</td>
</tr>
<tr>
<td>5997897</td>
<td>5998087</td>
<td>24</td>
</tr>
<tr>
<td>6033004</td>
<td>6033039</td>
<td>18</td>
</tr>
<tr>
<td>6069812</td>
<td>6069848</td>
<td>15</td>
</tr>
<tr>
<td>1948571</td>
<td>1948892</td>
<td>133</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>3574756</td>
<td>3575299</td>
<td>260</td>
</tr>
<tr>
<td>3768552</td>
<td>3768922</td>
<td>123</td>
</tr>
<tr>
<td>3956354</td>
<td>3956411</td>
<td>17</td>
</tr>
<tr>
<td>4134951</td>
<td>4135470</td>
<td>284</td>
</tr>
<tr>
<td>4785054</td>
<td>4785456</td>
<td>123</td>
</tr>
<tr>
<td>1954356</td>
<td>1955116</td>
<td>263</td>
</tr>
<tr>
<td>3156644</td>
<td>3157707</td>
<td>123</td>
</tr>
<tr>
<td>3176187</td>
<td>3176540</td>
<td>123</td>
</tr>
<tr>
<td>3235498</td>
<td>3235719</td>
<td>123</td>
</tr>
<tr>
<td>3593725</td>
<td>3594335</td>
<td>136</td>
</tr>
<tr>
<td>3680599</td>
<td>3680921</td>
<td>133</td>
</tr>
<tr>
<td>4046312</td>
<td>4046894</td>
<td>123</td>
</tr>
<tr>
<td>3729837</td>
<td>3730254</td>
<td>133</td>
</tr>
<tr>
<td>4858915</td>
<td>4859409</td>
<td>112</td>
</tr>
<tr>
<td>305869</td>
<td>306071</td>
<td>79</td>
</tr>
<tr>
<td>1002063</td>
<td>1002092</td>
<td>14</td>
</tr>
<tr>
<td>Start Position</td>
<td>End Position</td>
<td>Start Range</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>1374095</td>
<td>1374169</td>
<td>19</td>
</tr>
<tr>
<td>1824758</td>
<td>1824840</td>
<td>18</td>
</tr>
<tr>
<td>1824936</td>
<td>1824975</td>
<td>19</td>
</tr>
<tr>
<td>1916677</td>
<td>1916736</td>
<td>6</td>
</tr>
<tr>
<td>2176903</td>
<td>2177229</td>
<td>129</td>
</tr>
<tr>
<td>2187851</td>
<td>2187894</td>
<td>17</td>
</tr>
<tr>
<td>2196382</td>
<td>2196412</td>
<td>14</td>
</tr>
<tr>
<td>2229075</td>
<td>2229120</td>
<td>15</td>
</tr>
<tr>
<td>2354303</td>
<td>2354650</td>
<td>134</td>
</tr>
<tr>
<td>2355824</td>
<td>2356096</td>
<td>134</td>
</tr>
<tr>
<td>2375568</td>
<td>2375860</td>
<td>131</td>
</tr>
<tr>
<td>2388278</td>
<td>2388645</td>
<td>133</td>
</tr>
<tr>
<td>2436141</td>
<td>2436172</td>
<td>15</td>
</tr>
<tr>
<td>2450520</td>
<td>2450623</td>
<td>33</td>
</tr>
<tr>
<td>2471329</td>
<td>2471355</td>
<td>6</td>
</tr>
<tr>
<td>2546335</td>
<td>2546418</td>
<td>18</td>
</tr>
<tr>
<td>2546336</td>
<td>2546400</td>
<td>21</td>
</tr>
<tr>
<td>2546351</td>
<td>2546388</td>
<td>15</td>
</tr>
<tr>
<td>2546683</td>
<td>2546763</td>
<td>9</td>
</tr>
<tr>
<td>2546786</td>
<td>2546882</td>
<td>33</td>
</tr>
<tr>
<td>2546929</td>
<td>2547042</td>
<td>33</td>
</tr>
<tr>
<td>2546930</td>
<td>2547092</td>
<td>66</td>
</tr>
<tr>
<td>2546944</td>
<td>2546985</td>
<td>21</td>
</tr>
<tr>
<td>2547120</td>
<td>2547179</td>
<td>18</td>
</tr>
<tr>
<td>2565588</td>
<td>2565646</td>
<td>11</td>
</tr>
<tr>
<td>2646854</td>
<td>2647217</td>
<td>123</td>
</tr>
<tr>
<td>2707912</td>
<td>2707959</td>
<td>16</td>
</tr>
<tr>
<td>2744769</td>
<td>2744999</td>
<td>123</td>
</tr>
<tr>
<td>2759436</td>
<td>2759466</td>
<td>15</td>
</tr>
<tr>
<td>2888440</td>
<td>2888575</td>
<td>68</td>
</tr>
<tr>
<td>Start</td>
<td>End</td>
<td>Length</td>
</tr>
<tr>
<td>---------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>2994262</td>
<td>2994913</td>
<td>123</td>
</tr>
<tr>
<td>3022913</td>
<td>3023148</td>
<td>125</td>
</tr>
<tr>
<td>3110014</td>
<td>3110072</td>
<td>8</td>
</tr>
<tr>
<td>3129909</td>
<td>3129952</td>
<td>15</td>
</tr>
<tr>
<td>3138106</td>
<td>3138445</td>
<td>123</td>
</tr>
<tr>
<td>3279148</td>
<td>3279185</td>
<td>19</td>
</tr>
<tr>
<td>3354637</td>
<td>3354674</td>
<td>19</td>
</tr>
<tr>
<td>3408820</td>
<td>3408861</td>
<td>18</td>
</tr>
<tr>
<td>3421369</td>
<td>3422102</td>
<td>123</td>
</tr>
<tr>
<td>3432149</td>
<td>3432476</td>
<td>123</td>
</tr>
<tr>
<td>3577587</td>
<td>3577730</td>
<td>68</td>
</tr>
<tr>
<td>3577672</td>
<td>3578137</td>
<td>123</td>
</tr>
<tr>
<td>3591390</td>
<td>3591761</td>
<td>125</td>
</tr>
<tr>
<td>3654879</td>
<td>3654911</td>
<td>17</td>
</tr>
<tr>
<td>3655013</td>
<td>3655045</td>
<td>17</td>
</tr>
<tr>
<td>3671332</td>
<td>3671585</td>
<td>123</td>
</tr>
<tr>
<td>3698756</td>
<td>3699197</td>
<td>169</td>
</tr>
<tr>
<td>3751285</td>
<td>3751347</td>
<td>19</td>
</tr>
<tr>
<td>3756009</td>
<td>3756048</td>
<td>19</td>
</tr>
<tr>
<td>3784412</td>
<td>3784469</td>
<td>28</td>
</tr>
<tr>
<td>4128189</td>
<td>4128229</td>
<td>18</td>
</tr>
<tr>
<td>4731531</td>
<td>4731562</td>
<td>16</td>
</tr>
<tr>
<td>4764716</td>
<td>4764748</td>
<td>17</td>
</tr>
<tr>
<td>4772590</td>
<td>4772615</td>
<td>9</td>
</tr>
<tr>
<td>5015287</td>
<td>5015322</td>
<td>18</td>
</tr>
<tr>
<td>5262394</td>
<td>5262754</td>
<td>141</td>
</tr>
<tr>
<td>5755701</td>
<td>5755743</td>
<td>21</td>
</tr>
<tr>
<td>6127819</td>
<td>6127848</td>
<td>14</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Protein Localization</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytoplasmic</td>
<td>2509</td>
</tr>
<tr>
<td>Cytoplasmic membrane</td>
<td>1259</td>
</tr>
<tr>
<td>Periplasmic</td>
<td>176</td>
</tr>
<tr>
<td>Outer membrane</td>
<td>115</td>
</tr>
<tr>
<td>Extracellular</td>
<td>44</td>
</tr>
<tr>
<td>Unknown</td>
<td>1328</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>5431</strong></td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>CDS ID</th>
<th>Gene</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>1484</td>
<td>pvdYII</td>
<td>Hydroxyornithine acetylase</td>
</tr>
<tr>
<td>1485</td>
<td>pvdS</td>
<td>ECF iron sigma factor</td>
</tr>
<tr>
<td>1486</td>
<td>pvdL</td>
<td>Non-ribosomal peptide synthetase</td>
</tr>
<tr>
<td>1503</td>
<td>pvdH</td>
<td>Aminotransferase</td>
</tr>
<tr>
<td>1504</td>
<td>mbtH</td>
<td>Unknown</td>
</tr>
<tr>
<td>1549</td>
<td>pvdG</td>
<td>Thioesterase</td>
</tr>
<tr>
<td>1550</td>
<td>pvdI</td>
<td>Non-ribosomal peptide synthetase</td>
</tr>
<tr>
<td>1551</td>
<td>syrP</td>
<td>Pyoverdine biosynthesis regulatory protein</td>
</tr>
<tr>
<td>1552</td>
<td>pvdD</td>
<td>Non-ribosomal peptide synthetase</td>
</tr>
<tr>
<td>1553</td>
<td>pvdJ</td>
<td>Non-ribosomal peptide synthetase</td>
</tr>
<tr>
<td>1569</td>
<td>fpvA</td>
<td>TonB-dependent pyoverdine receptor</td>
</tr>
<tr>
<td>1570</td>
<td>pvdE</td>
<td>ABC transporter</td>
</tr>
<tr>
<td>1571</td>
<td>pvdO</td>
<td>Unknown</td>
</tr>
<tr>
<td>1572</td>
<td>pvdN</td>
<td>Aminotransferase</td>
</tr>
<tr>
<td>1573</td>
<td>pvdM</td>
<td>Dipeptidase</td>
</tr>
<tr>
<td>1574</td>
<td>pvdP</td>
<td>Unknown</td>
</tr>
<tr>
<td>1575</td>
<td></td>
<td>Outer membrane efflux protein</td>
</tr>
<tr>
<td>1576</td>
<td></td>
<td>ABC transporter</td>
</tr>
<tr>
<td>1577</td>
<td></td>
<td>Membrane efflux protein</td>
</tr>
<tr>
<td>1578</td>
<td>fpvI</td>
<td>ECF sigma factor required for expression of fpvA</td>
</tr>
<tr>
<td>1579</td>
<td>pvdA</td>
<td>L-ornithine hydroxylase</td>
</tr>
<tr>
<td>2459</td>
<td>pvdQ</td>
<td>Acylase</td>
</tr>
</tbody>
</table>
Fig. 4-6. *P. putida* UW4 predicted pyoverdine biosynthesis pathway.
L-glutamate  L-tyrosine  serine  L-arginine  L-2,4-diaminobutanoate

L-aspartate-semialdehyde

diaminobutyrate-2-oxoglutarate aminotransferase

L-ornithine

L-ornithine \( N^5 \)-oxygenase

\( pvdA \)

\( N^5 \)-hydroxy-L-ornithine

hydroxyornithine acetylase

\( pvdYII \)

L-threonine

L-lysine

Non-ribosomal peptide synthetase: \( pvdD, pvdI, pvdJ, pvdL \)

Pyoverdine

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

Maltose $\xrightarrow{\text{Trehalose synthase TreS}}$ Trehalose

```
PputUW4_2805
```

**TreY-TreZ pathway**

Maltodextrin $\xrightarrow{\text{Maltooligosyltrehalose synthase TreY}}$ Maltooligosyltrehalose

```
PputUW4_2797
```

Maltooligosyltrehalose $\xrightarrow{\text{Maltooligosyltrehalose trehalohydrolase TreZ}}$ Trehalose

```
PputUW4_2795
```
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 acetyl-CoA 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.
Pentose phosphate pathway

D-erythrose-4-phosphate
PputUW4_1418 Phospho-2-dehydro-3-deoxyheptonate aldolase

3-deoxy-D-arabino-heptulosonate-7-phosphate
PputUW4_0352 3-dehydroquinate synthase

3-dehydroquinate
PputUW4_0556 PputUW4_4725 3-dehydroquinate dehydratase

3-dehydroshikimate

Quinate/shikimate dehydrogenase  Shikimate 5-dehydrogenase
PputUW4_4965 PputUW4_0022 and PputUW4_4726

Shikimate
PputUW4_0351 Shikimate kinase

Shikimate-3-phosphate
PputUW4_1365 3-phosphoshikimate 1-carboxyvinyltransferase

5-enolpyruvyl-shikimate-3-phosphate
PputUW4_3706 Chorismate synthase

Chorismate
PputUW4_5356 Chorismate lyase

4-hydroxybenzoate
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.

<table>
<thead>
<tr>
<th>PputUW4</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>0125</td>
<td>multidrug efflux MFS transporter</td>
</tr>
<tr>
<td>0126</td>
<td>multidrug efflux MFS membrane fusion protein</td>
</tr>
<tr>
<td>0127</td>
<td>multidrug efflux MFS outer membrane protein</td>
</tr>
<tr>
<td>0135</td>
<td>fusaric acid resistance protein</td>
</tr>
<tr>
<td>0137</td>
<td>multidrug resistance efflux pump protein</td>
</tr>
<tr>
<td>0138</td>
<td>multidrug efflux outer membrane protein</td>
</tr>
<tr>
<td>0180</td>
<td>multidrug RND transporter, membrane fusion protein</td>
</tr>
<tr>
<td>0181</td>
<td>multidrug RND transporter, membrane fusion protein</td>
</tr>
<tr>
<td>0182</td>
<td>acriflavin resistance protein</td>
</tr>
<tr>
<td>0503</td>
<td>small multidrug resistance protein SugE</td>
</tr>
<tr>
<td>1069</td>
<td>multidrug RND transporter, membrane fusion protein</td>
</tr>
<tr>
<td>1070</td>
<td>acriflavin resistance protein</td>
</tr>
<tr>
<td>1214</td>
<td>multidrug efflux RND outer membrane protein</td>
</tr>
<tr>
<td>1215</td>
<td>multidrug efflux RND inner membrane transporter</td>
</tr>
<tr>
<td>1216</td>
<td>multidrug efflux RND membrane fusion protein</td>
</tr>
<tr>
<td>1594</td>
<td>EmrB/QacA family drug resistance transporter</td>
</tr>
<tr>
<td>1595</td>
<td>multidrug resistance efflux protein</td>
</tr>
<tr>
<td>1897</td>
<td>multidrug resistance efflux protein</td>
</tr>
<tr>
<td>1898</td>
<td>EmrB/QacA subfamily drug resistance transporter</td>
</tr>
<tr>
<td>2114</td>
<td>multidrug efflux system outer membrane protein</td>
</tr>
<tr>
<td>2115</td>
<td>RND family efflux transporter MFP subunit</td>
</tr>
<tr>
<td>2116</td>
<td>acriflavin resistance protein</td>
</tr>
<tr>
<td>2144</td>
<td>acriflavin resistance protein</td>
</tr>
<tr>
<td>2145</td>
<td>RND family efflux transporter MFP subunit</td>
</tr>
<tr>
<td>2146</td>
<td>RND efflux system outer membrane lipoprotein</td>
</tr>
<tr>
<td>2443</td>
<td>multidrug efflux system transmembrane protein</td>
</tr>
<tr>
<td>2444</td>
<td>multidrug efflux RND membrane fusion protein</td>
</tr>
<tr>
<td>2536</td>
<td>fusaric acid resistance protein</td>
</tr>
<tr>
<td>2538</td>
<td>multidrug resistance efflux pump protein</td>
</tr>
<tr>
<td>2539</td>
<td>multidrug efflux system outer membrane protein</td>
</tr>
<tr>
<td>2540</td>
<td>RND multidrug efflux transporter</td>
</tr>
<tr>
<td>2541</td>
<td>RND family efflux transporter MFP subunit</td>
</tr>
<tr>
<td>2815</td>
<td>multidrug efflux system outer membrane protein</td>
</tr>
<tr>
<td>2816</td>
<td>multidrug efflux system transmembrane protein</td>
</tr>
<tr>
<td>2817</td>
<td>multidrug efflux system transmembrane protein</td>
</tr>
<tr>
<td>2818</td>
<td>multidrug efflux system inner membrane protein</td>
</tr>
<tr>
<td>2954</td>
<td>ABC efflux system outer membrane protein</td>
</tr>
<tr>
<td>2955</td>
<td>ABC efflux system ATP-binding protein</td>
</tr>
<tr>
<td>2956</td>
<td>ABC efflux system permease</td>
</tr>
<tr>
<td>Line</td>
<td>Description</td>
</tr>
<tr>
<td>------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>3471</td>
<td>multidrug resistance efflux pump</td>
</tr>
<tr>
<td>3472</td>
<td>EmrB/QacA family drug resistance transporter</td>
</tr>
<tr>
<td>3533</td>
<td>multidrug ABC transporter ATPase/permease</td>
</tr>
<tr>
<td>3536</td>
<td>multidrug efflux transporter</td>
</tr>
<tr>
<td>3756</td>
<td>small multidrug resistance protein</td>
</tr>
<tr>
<td>3988</td>
<td>multidrug ABC transporter ATP-binding protein/permease</td>
</tr>
<tr>
<td>4061</td>
<td>ABC-type multidrug transport system, permease component</td>
</tr>
<tr>
<td>4062</td>
<td>ABC-type multidrug transport system, ATPase component</td>
</tr>
<tr>
<td>4383</td>
<td>RND family multidrug transporter membrane fusion protein</td>
</tr>
<tr>
<td>4384</td>
<td>RND multidrug efflux transporter</td>
</tr>
<tr>
<td>4491</td>
<td>fusaric acid resistance protein</td>
</tr>
<tr>
<td>4493</td>
<td>fusaric acid resistance protein</td>
</tr>
<tr>
<td>4494</td>
<td>RND efflux system, outer membrane lipoprotein</td>
</tr>
<tr>
<td>4496</td>
<td>RND family multidrug transporter membrane fusion protein</td>
</tr>
<tr>
<td>4497</td>
<td>multidrug efflux system transmembrane protein</td>
</tr>
<tr>
<td>5156</td>
<td>multidrug resistance transporter, Bcr/CflA family</td>
</tr>
<tr>
<td>5306</td>
<td>multidrug resistance-transport membrane protein</td>
</tr>
<tr>
<td>5424</td>
<td>permease</td>
</tr>
<tr>
<td>5425</td>
<td>multidrug efflux system outer membrane protein</td>
</tr>
<tr>
<td>5426</td>
<td>multidrug resistance efflux pump</td>
</tr>
<tr>
<td>5427</td>
<td>membrane protein</td>
</tr>
</tbody>
</table>
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 β-ketothiolase (PhaA) to produce acetoacetyl-CoA. In the presence of NADH, acetoacetyl-CoA reductase (PhaB) catalyzes the reduction of acetoacetyl-CoA to β-hydroxybutyryl-CoA. Then, PHA synthase (PhaC) polymerizes β-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, β-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, β-hydroxyacyl-CoA.

The PHA degradation pathway in UW4 is catalyzed by PHA depolymerase (PhaZ). PhaZ first depolymerizes PHA to β-hydroxybutyrate monomers, which are subsequently converted to acetoacetate by β-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 β-ketoadipate pathway was identified (Fig. 4-13). In addition, the protocatechuate branch of the β-ketoadipate 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 *mhpRABCDFET* 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, β-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; β-hydroxybutyrate dehydrogenase: PputUW4_2378; Acetoacetyl-CoA synthetase: PputUW4_1999 and 2000.
Fig. 4-13. Benzoate degradation pathway in *P. putida* UW4.
benzoate
PputUW4_3260, 3261, 3262
\[\text{Benzoate 1,2-dioxygenase} \quad + \text{O}_2\]
1,2-cis-dihydroxybenzoate
PputUW4_3259
\[\text{1,6-dihydroxycyclohexa-2,4-diene-1-carboxylate dehydrogenase} \quad - \text{CO}_2\]
catechol
PputUW4_3264
\[\text{Catechol 1,2-dioxygenase} \quad + \text{O}_2\]
cis,cis-muconate
PputUW4_3266
\[\text{Muconate cycloisomerase}\]
muconolactone
PputUW4_3265
\[\text{Muconate delta-isomerase}\]
3-oxoadipate-enol-lactone
PputUW4_1210, 1495, 3201
\[\text{3-oxoadipate enol-lactonase}\]
3-oxoadipate
\[\text{3-oxoadipate:succinyl-CoA transferase}\]
PputUW4_1204, 1205
\[\text{succinyl-CoA}\]
succinate
\[\text{3-oxoadipyl-CoA}\]
\[\text{coenzyme A}\]
\[\text{acetyl-CoA}\]
\[\text{β-ketoadipyl CoA thiolase}\]
PputUW4_1206
\[\text{succinyl-CoA}\]
TCA cycle
Fig. 4-14. 4-hydroxybenzoate and L-quinate degradation pathways in *P. putida* UW4.
4-hydroxybenzoate 3-monooxygenase

4-hydroxybenzoate → protocatechuate

Protocatechuate 3,4-dioxygenase

Protocatechuate → 3-carboxy-cis,cis-muconate

3-carboxy-cis,cis-muconate cycloisomerase

3-carboxy-cis,cis-muconate → 2-carboxy-5-oxo-2,5-dihydrofuran-2-acetate

4-carboxymuconolactone decarboxylase

2-carboxy-5-oxo-2,5-dihydrofuran-2-acetate → 3-oxoadipate-enol-lactone

3-oxoadipate enol-lactonase

3-oxoadipate → 3-oxoadipyl-CoA

3-oxoadipyl-CoA transferase

3-oxoadipyl-CoA → succinyl-CoA

β-ketoadipyl CoA thiolase

β-ketoadipyl-CoA → acetyl-CoA

Sucinyl-CoA

TCA cycle
encoding genes, \textit{mhpA CDFE}, were found (Table 4-12), indicating the presence of an incomplete degradation pathway for 3-HPP.

\textbf{4.2.9 Heavy metal resistance}

Based on the genome sequence of \textit{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 (\textit{nikABCDE}) that was originally identified in \textit{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 \textit{E. coli}) or upstream of NikA (in \textit{P. putida} KT2440 and \textit{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 \textit{nikA}, there is a putative gene encoding an outer membrane porin, which is transcribed in the same orientation as \textit{nikA}. However, based on a sequence similarity search, NikR is not encoded in the UW4 genome.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Product</th>
<th>W619 ORF ID</th>
<th>UW4 ORF ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>mhpT</td>
<td>3-Hydroxyphenylpropionic acid transporter</td>
<td>1985</td>
<td></td>
</tr>
<tr>
<td>mhpE</td>
<td>4-Hydroxy-2-oxovalerate aldolase</td>
<td>1984</td>
<td>1665</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2007</td>
<td></td>
</tr>
<tr>
<td>mhpF</td>
<td>Acetaldehyde dehydrogenase</td>
<td>1983</td>
<td>1664</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2008</td>
<td></td>
</tr>
<tr>
<td>mhpD</td>
<td>2-Hydroxypenta-2,4-dienoate hydratase</td>
<td>1982</td>
<td>1663</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2011</td>
<td></td>
</tr>
<tr>
<td>mhpC</td>
<td>2-Hydroxy-6-ketonona-2,4-dienedioic acid hydrolase</td>
<td>1981</td>
<td>1662</td>
</tr>
<tr>
<td>mhpB</td>
<td>2,3-Dihydroxyphenylpropionate 1,2-dioxygenase</td>
<td>1980</td>
<td></td>
</tr>
<tr>
<td>mhpA</td>
<td>3-(3-Hydroxy-phenyl)propionate hydroxylase</td>
<td>1979</td>
<td>2149</td>
</tr>
<tr>
<td>mhpR</td>
<td>Mhp operon transcriptional activator</td>
<td>1978</td>
<td></td>
</tr>
</tbody>
</table>
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 multicopper 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.

<table>
<thead>
<tr>
<th>Metal</th>
<th>Gene</th>
<th>PputUW4</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nickel</td>
<td><em>nikA3</em></td>
<td>0749</td>
<td>nickel ABC transporter periplasmic nickel-binding protein</td>
</tr>
<tr>
<td></td>
<td><em>nikA2</em></td>
<td>0748</td>
<td>nickel ABC transporter periplasmic nickel-binding protein</td>
</tr>
<tr>
<td></td>
<td><em>nikA1</em></td>
<td>0746</td>
<td>nickel ABC transporter periplasmic nickel-binding protein</td>
</tr>
<tr>
<td></td>
<td><em>nikB</em></td>
<td>0745</td>
<td>nickel transporter permease</td>
</tr>
<tr>
<td></td>
<td><em>nikC</em></td>
<td>0744</td>
<td>nickel transporter permease</td>
</tr>
<tr>
<td></td>
<td><em>nikD</em></td>
<td>0743</td>
<td>nickel import ATP-binding protein</td>
</tr>
<tr>
<td></td>
<td><em>nikE</em></td>
<td>0742</td>
<td>nickel import ATP-binding protein</td>
</tr>
<tr>
<td>Copper</td>
<td><em>copP</em></td>
<td>0583</td>
<td>heavy metal transport/detoxification protein</td>
</tr>
<tr>
<td></td>
<td><em>cueA</em></td>
<td>0581</td>
<td>copper-translocating P-type ATPase</td>
</tr>
<tr>
<td></td>
<td><em>cueR</em></td>
<td>0580</td>
<td>MerR family transcriptional regulator</td>
</tr>
<tr>
<td></td>
<td><em>copA</em></td>
<td>3489</td>
<td>copper resistance protein</td>
</tr>
<tr>
<td></td>
<td><em>copB</em></td>
<td>3490</td>
<td>copper resistance protein</td>
</tr>
<tr>
<td></td>
<td><em>copC</em></td>
<td>3491</td>
<td>copper resistance protein</td>
</tr>
<tr>
<td></td>
<td><em>copD</em></td>
<td>3492</td>
<td>copper resistance protein</td>
</tr>
<tr>
<td></td>
<td><em>cinQ</em></td>
<td>3503</td>
<td>7-cyano-7-deazaguanine (pre-Q0) reductase</td>
</tr>
<tr>
<td></td>
<td><em>cinA</em></td>
<td>3504</td>
<td>copper-containing azurin-like protein</td>
</tr>
<tr>
<td></td>
<td><em>cinR</em></td>
<td>3505</td>
<td>two-component heavy metal response transcriptional regulator</td>
</tr>
<tr>
<td></td>
<td><em>cinS</em></td>
<td>3506</td>
<td>heavy metal sensor histidine kinase</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2454</td>
<td>involved in survival in the presence of high bioavailable Cu(II)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2050</td>
<td>heavy metal sensor signal transduction histidine kinase</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2051</td>
<td>two-component heavy metal response transcriptional regulator</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4498</td>
<td>two-component heavy metal response transcriptional regulator</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4499</td>
<td>heavy metal sensor signal transduction histidine kinase</td>
</tr>
<tr>
<td>Cadmium</td>
<td><em>cadA</em></td>
<td>5170</td>
<td>cadmium translocating P-type ATPase</td>
</tr>
<tr>
<td></td>
<td><em>cadR</em></td>
<td>5171</td>
<td>MerR family transcriptional regulator</td>
</tr>
<tr>
<td></td>
<td><em>cadA2</em></td>
<td>5411</td>
<td>cadmium-translocating P-type ATPase</td>
</tr>
<tr>
<td>Zinc</td>
<td><em>znuA</em></td>
<td>0067</td>
<td>zinc ABC transporter periplasmic protein</td>
</tr>
<tr>
<td></td>
<td><em>znuB</em></td>
<td>0064</td>
<td>zinc ABC transporter permease</td>
</tr>
<tr>
<td></td>
<td><em>znuC</em></td>
<td>0065</td>
<td>zinc import ATP-binding protein</td>
</tr>
<tr>
<td></td>
<td><em>zur</em></td>
<td>0066</td>
<td>ferric uptake regulator family protein</td>
</tr>
<tr>
<td>Molybdate</td>
<td><em>modA</em></td>
<td>2404</td>
<td>molybdate-binding periplasmic protein</td>
</tr>
<tr>
<td></td>
<td><em>modB</em></td>
<td>2403</td>
<td>molybdate ABC transporter permease</td>
</tr>
<tr>
<td></td>
<td><em>modC</em></td>
<td>2402</td>
<td>molybdate ABC transporter ATP-binding protein</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4989</td>
<td>ModE family transcriptional regulator</td>
</tr>
<tr>
<td></td>
<td>Gene</td>
<td>Location</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td>----------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>Cobalt</td>
<td>cbtA</td>
<td>2364</td>
<td>cobalt transporter subunit A</td>
</tr>
<tr>
<td></td>
<td>cbtB</td>
<td>2365</td>
<td>cobalt transporter subunit B</td>
</tr>
<tr>
<td>Arsenate</td>
<td>arsR</td>
<td>2256</td>
<td>arsenical resistance operon repressor</td>
</tr>
<tr>
<td></td>
<td>arsB</td>
<td>2255</td>
<td>arsenite efflux transporter</td>
</tr>
<tr>
<td></td>
<td>arsC1</td>
<td>2254</td>
<td>arsenate reductase</td>
</tr>
<tr>
<td></td>
<td>arsH</td>
<td>2253</td>
<td>NADPH-dependent FMN reductase</td>
</tr>
<tr>
<td></td>
<td>arsC2</td>
<td>1085</td>
<td>arsenate reductase</td>
</tr>
<tr>
<td></td>
<td>arsC3</td>
<td>4122</td>
<td>arsenate reductase</td>
</tr>
<tr>
<td>Chromate</td>
<td>chrA</td>
<td>3072</td>
<td>chromate transporter</td>
</tr>
<tr>
<td></td>
<td>oscA</td>
<td>0155</td>
<td>hypothetical protein</td>
</tr>
</tbody>
</table>
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.
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-Q0) 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.
### (A)

- **cureR**: 0580
- **cureA**: 0581
- **hypothetical**: 0582
- **copP**: 0583

### (B)

<table>
<thead>
<tr>
<th>PNL-MK25_CueA</th>
<th>UW4_CueA</th>
<th>PNL-MK25_CueA</th>
<th>UW4_CueA</th>
<th>PNL-MK25_CueA</th>
<th>UW4_CueA</th>
<th>PNL-MK25_CueA</th>
<th>UW4_CueA</th>
<th>PNL-MK25_CueA</th>
<th>UW4_CueA</th>
</tr>
</thead>
<tbody>
<tr>
<td>VYTHIPIGNTACAGVRERAIKVSAYASVNLATQARYV</td>
<td>VYTHIPIGNTACAGVRERAIKVSAYASVNLATQARYV</td>
<td>TPLWAVAGYSVPRAPVESIDGKTACACVRERAIKVS</td>
<td>TPLWAVAGYSVPRAPVESIDGKTACACVRERAIKVS</td>
<td>TPLWAVAGYSVPRAPVESIDGKTACACVRERAIKVS</td>
<td>TPLWAVAGYSVPRAPVESIDGKTACACVRERAIKVS</td>
<td>GSSSGPACPSLQFPLPHPLPCGGNANAFL</td>
<td>GSSSGPACPSLQFPLPHPLPCGGNANAFL</td>
<td>GSSSGPACPSLQFPLPHPLPCGGNANAFL</td>
<td>GSSSGPACPSLQFPLPHPLPCGGNANAFL</td>
</tr>
<tr>
<td>D 50</td>
<td>D 80</td>
<td>D 110</td>
<td>D 140</td>
<td>D 170</td>
<td>D 200</td>
<td>D 230</td>
<td>D 260</td>
<td>D 290</td>
<td>D 320</td>
</tr>
<tr>
<td>S 50</td>
<td>S 80</td>
<td>S 110</td>
<td>S 140</td>
<td>S 170</td>
<td>S 200</td>
<td>S 230</td>
<td>S 260</td>
<td>S 290</td>
<td>S 320</td>
</tr>
<tr>
<td>P 50</td>
<td>P 80</td>
<td>P 110</td>
<td>P 140</td>
<td>P 170</td>
<td>P 200</td>
<td>P 230</td>
<td>P 260</td>
<td>P 290</td>
<td>P 320</td>
</tr>
<tr>
<td>V 50</td>
<td>V 80</td>
<td>V 110</td>
<td>V 140</td>
<td>V 170</td>
<td>V 200</td>
<td>V 230</td>
<td>V 260</td>
<td>V 290</td>
<td>V 320</td>
</tr>
<tr>
<td>D 50</td>
<td>D 80</td>
<td>D 110</td>
<td>D 140</td>
<td>D 170</td>
<td>D 200</td>
<td>D 230</td>
<td>D 260</td>
<td>D 290</td>
<td>D 320</td>
</tr>
<tr>
<td>S 50</td>
<td>S 80</td>
<td>S 110</td>
<td>S 140</td>
<td>S 170</td>
<td>S 200</td>
<td>S 230</td>
<td>S 260</td>
<td>S 290</td>
<td>S 320</td>
</tr>
<tr>
<td>P 50</td>
<td>P 80</td>
<td>P 110</td>
<td>P 140</td>
<td>P 170</td>
<td>P 200</td>
<td>P 230</td>
<td>P 260</td>
<td>P 290</td>
<td>P 320</td>
</tr>
<tr>
<td>V 50</td>
<td>V 80</td>
<td>V 110</td>
<td>V 140</td>
<td>V 170</td>
<td>V 200</td>
<td>V 230</td>
<td>V 260</td>
<td>V 290</td>
<td>V 320</td>
</tr>
<tr>
<td>D 50</td>
<td>D 80</td>
<td>D 110</td>
<td>D 140</td>
<td>D 170</td>
<td>D 200</td>
<td>D 230</td>
<td>D 260</td>
<td>D 290</td>
<td>D 320</td>
</tr>
<tr>
<td>S 50</td>
<td>S 80</td>
<td>S 110</td>
<td>S 140</td>
<td>S 170</td>
<td>S 200</td>
<td>S 230</td>
<td>S 260</td>
<td>S 290</td>
<td>S 320</td>
</tr>
<tr>
<td>P 50</td>
<td>P 80</td>
<td>P 110</td>
<td>P 140</td>
<td>P 170</td>
<td>P 200</td>
<td>P 230</td>
<td>P 260</td>
<td>P 290</td>
<td>P 320</td>
</tr>
<tr>
<td>V 50</td>
<td>V 80</td>
<td>V 110</td>
<td>V 140</td>
<td>V 170</td>
<td>V 200</td>
<td>V 230</td>
<td>V 260</td>
<td>V 290</td>
<td>V 320</td>
</tr>
</tbody>
</table>

**Notes:**
- The table above represents a comparison of amino acid sequences for PNL-MK25_CueA and UW4_CueA, with specific attention to conserved regions.
- The sequences are aligned to highlight similarities and differences, aiding in the analysis of evolutionary relationships and functional comparisons.

**125**
(C)

<table>
<thead>
<tr>
<th>PNL-MK25_CueR</th>
<th>UW4_CueR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNIGQAAR</td>
<td>SGLSAKMRYYESIGLLKAA</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>RGYGDDTHLAFIKRSLDRGFSLEEVGKLTTLWQD</td>
<td>RGYGDDTHLAFIKRSLDRGFSLEEVGKLTTLWQD</td>
</tr>
</tbody>
</table>

(D)

<table>
<thead>
<tr>
<th>PNL-MK25_CopP</th>
<th>UW4_CopP</th>
</tr>
</thead>
<tbody>
<tr>
<td>MQVFNVEGSMSCGHVCVAITTLAVQALDPAAVSVRV</td>
<td>MQVFNVEGSMSCGHVCVAITQLAVQALKDPAAVSVRV</td>
</tr>
<tr>
<td></td>
<td>HMA domain</td>
</tr>
<tr>
<td>DLAKEVGEVAESLDQVIAIREEGYEVLV</td>
<td>DLAKEVGEVAESLDQVIEAISEEYGVL</td>
</tr>
</tbody>
</table>

(E)

**hypothesis**

GGCTGAATCTCCTGGTCTCGGTCTCATTTGATGG

ACGCAGCTTGCGCAGGTAGCGCCGAGTGGCCAGCCCGGGCAG

-35

CAATTTCA**GAATCTTGAGCCTTGCCATGTTGGCAG**

-10

TCGAAAC

TGGACTCAACCCTCGGAATAAGGAGTCATCCC **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).
(E)

**UW4_CopD**

<table>
<thead>
<tr>
<th>1</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>MADLSLRLQELVNLDDLFLGAFGLYSF</td>
<td>PFRLLRCHAVT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSDSANVVERFVLMLFLGVALFLGYSF</td>
<td>PVLPFRRLTCLSL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**PNL-MK25_CopD**

<table>
<thead>
<tr>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>CALISAGCMLPFAVSKTOLALPHEOMNLPTDVLAWALRNLALAI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CALISTVGQELPVAMSGEDVATLPHEMNLPTDVMAYAVRMALAVAGL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**UW4_CopD**

<table>
<thead>
<tr>
<th>110</th>
<th>120</th>
<th>130</th>
<th>140</th>
<th>150</th>
<th>160</th>
</tr>
</thead>
<tbody>
<tr>
<td>VVIFSSLLPASLACETALASLAWSGHCDDTVDRIKLFWLSDLHIIL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AVMLNFAAPGILSVLASPACETALASLAWSGHCAMDEGSRKWHETDILHIIL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**PNL-MK25_CopD**

<table>
<thead>
<tr>
<th>170</th>
<th>180</th>
<th>190</th>
<th>200</th>
<th>210</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGAWGCLWPLVLMASGRGD</td>
<td>RURLLADVKRFERNVGALIVQTSTVTGV</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAGAWGCAQKGFLMAQFTLQEBKIRLLAABVRFERNVGALIVIITVTGV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**UW4_CopD**

<table>
<thead>
<tr>
<th>220</th>
<th>230</th>
<th>240</th>
<th>250</th>
<th>260</th>
<th>270</th>
</tr>
</thead>
<tbody>
<tr>
<td>NYLFIVGPRLESVLPGTYGGLL</td>
<td>IKVFLAALNRPLGCFSTQGSLRD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VNYLFIVGPRLESVLSPYGGLL</td>
<td>IKVFLAALNRPLGCFSTQGSLRD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**PNL-MK25_CopD**

<table>
<thead>
<tr>
<th>280</th>
<th>290</th>
<th>300</th>
<th>309</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOYRFLARLSVAELALTVLVAWGLTSPG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GOYRVAFLARLSVAEMVNLTVLVAWGLTSPGVE</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(F)

![Graph showing predicted transmembrane helix position](image)

Violet boxes (below abscissa): Predicted transmembrane helix position. (DIG index)

Gray line: Modified hydrophobic moment index. (INDA index)

Blue line: Beta preference. (BET index)

Red line: Transmembrane helix preference. (THM index)

Click here for numeric data (the original output of the prediction program)!
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 (HXXMXMX/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*. 
(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$_{12}$ 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.
Fig. 4-21. Amino acid sequence alignment of PputUW4_2454 and CopG of KT2440.
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.
### (B)

<table>
<thead>
<tr>
<th>P. putida 06909_CadR</th>
<th>P. putida 06909_CadR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MKIGELAKTDCAVETIRYERELLPEPARSDGYRLETOH</td>
<td>MKIGELAKTDCAVETIRYERELLPEPARSDGYRLETOH</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>UW4_CadR</td>
<td>UW4_CadR</td>
</tr>
<tr>
<td>RALDMTLEIRSLALRDSFGDCSVNALIDEHHEVVOARIDGGLVALCOCL</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>P. putida 06909_CadR</td>
<td>P. putida 06909_CadR</td>
</tr>
<tr>
<td>DRRCNAGA---ECALOQFTNCSVPEEHSHVGRSHCH</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>UW4_CadR</td>
<td>UW4_CadR</td>
</tr>
<tr>
<td>DRRCGERPGIDGCSLQRLESVSATAPEHSHVGRSHCH</td>
<td></td>
</tr>
</tbody>
</table>
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.
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.
Fig. 4-25. Sequence comparison of ChrA between (A) *P. putida* UW4 and *P. aeruginosa* (B) *P. putida* UW4 and *P. putida* KT2440.
Fig. 4-26. Genetic organization of \textit{oscA} in the genome of \textit{P. putida} UW4. \textit{cysP}, encoding sulphate-binding protein; \textit{cysT}, encoding sulfate ABC transporter permease protein; \textit{cysW}, encoding sulfate ABC transporter permease protein; \textit{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+ symporters (Pilsyk 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-SO3H), sulfamates (C-NH-SO3H), and sulfate esters (C-O-SO3H) (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.
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 pqqABCDEFGH (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.

<table>
<thead>
<tr>
<th>System</th>
<th>Gene</th>
<th>PputUW4</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>(a) Sec secretory pathway</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chaperone</td>
<td>secB</td>
<td>0281</td>
</tr>
<tr>
<td>ATPase</td>
<td>secA</td>
<td>4321</td>
</tr>
<tr>
<td>TM complex</td>
<td>secY</td>
<td>4869</td>
</tr>
<tr>
<td></td>
<td>secE</td>
<td>4902</td>
</tr>
<tr>
<td></td>
<td>secG</td>
<td>0706</td>
</tr>
<tr>
<td>Auxiliary proteins</td>
<td>secD</td>
<td>0923</td>
</tr>
<tr>
<td></td>
<td>secF</td>
<td>0924</td>
</tr>
<tr>
<td></td>
<td>yajC</td>
<td>0922</td>
</tr>
<tr>
<td>Membrane insertion proteins</td>
<td>ftsY</td>
<td>5140</td>
</tr>
<tr>
<td></td>
<td>ffh</td>
<td>4439</td>
</tr>
<tr>
<td></td>
<td>yidC</td>
<td>5445</td>
</tr>
<tr>
<td>Spase I</td>
<td>lepB</td>
<td>0960</td>
</tr>
<tr>
<td>Spase II</td>
<td>lspA</td>
<td>4672</td>
</tr>
<tr>
<td><em>(b) Twin arginine targeting (Tat) secretory pathway</em></td>
<td>tatA</td>
<td>0323</td>
</tr>
<tr>
<td></td>
<td>tatB</td>
<td>0322</td>
</tr>
<tr>
<td></td>
<td>tatC</td>
<td>0321</td>
</tr>
<tr>
<td><em>(c) Large conductance mechanosensitive ion channel</em></td>
<td>mscL</td>
<td>4710</td>
</tr>
<tr>
<td><em>(d) Type I secretory system</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TolC family outer membrane protein</td>
<td></td>
<td>0117</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1722</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3956</td>
</tr>
<tr>
<td>HylD family membrane fusion protein</td>
<td></td>
<td>0119</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1723</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2637</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3958</td>
</tr>
<tr>
<td>ATPase</td>
<td></td>
<td>0118</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1724</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2638</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3957</td>
</tr>
<tr>
<td>Putative substrate: RTX toxins and related Ca²⁺ binding proteins</td>
<td></td>
<td>0114</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1725</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2636</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3955</td>
</tr>
</tbody>
</table>
Table 4-14. Continued.

(e) Type II secretory system

<table>
<thead>
<tr>
<th>Protein Type</th>
<th>Gene</th>
<th>Start Codon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secretin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inner membrane protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>APTase</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(f) Type III secretory system

<table>
<thead>
<tr>
<th>Protein Type</th>
<th>Gene</th>
<th>Start Codon</th>
</tr>
</thead>
<tbody>
<tr>
<td>HopJ type III effector protein</td>
<td></td>
<td>0810</td>
</tr>
<tr>
<td>Apparatus protein</td>
<td>orgAB</td>
<td>3618</td>
</tr>
<tr>
<td>Apparatus lipoprotein</td>
<td>prgK</td>
<td>3619</td>
</tr>
<tr>
<td>Needle complex protein</td>
<td>prgJ</td>
<td>3620</td>
</tr>
<tr>
<td>Cytoplasmic protein</td>
<td>prgI</td>
<td>3621</td>
</tr>
<tr>
<td>Needle complex protein</td>
<td>prgH</td>
<td>3622</td>
</tr>
<tr>
<td>Cell invasion protein</td>
<td>iagB</td>
<td>3623</td>
</tr>
<tr>
<td>Invasion protein regulator</td>
<td>hilA</td>
<td>3624</td>
</tr>
<tr>
<td>Acyl carrier protein</td>
<td>iacP</td>
<td>3625</td>
</tr>
<tr>
<td>Cell invasion protein</td>
<td>sipD</td>
<td>3626</td>
</tr>
<tr>
<td>Invasin</td>
<td>sipC</td>
<td>3627</td>
</tr>
<tr>
<td></td>
<td>sipB</td>
<td>3628</td>
</tr>
<tr>
<td>Chaperone</td>
<td>sicA</td>
<td>3629</td>
</tr>
<tr>
<td>Surface presentation of antigens protein</td>
<td>spaS</td>
<td>3630</td>
</tr>
<tr>
<td></td>
<td>spaR</td>
<td>3631</td>
</tr>
<tr>
<td></td>
<td>spaQ</td>
<td>3632</td>
</tr>
<tr>
<td></td>
<td>spaP</td>
<td>3633</td>
</tr>
<tr>
<td></td>
<td>spaO</td>
<td>3634</td>
</tr>
<tr>
<td>Invasion protein</td>
<td>invJ</td>
<td>3635</td>
</tr>
<tr>
<td>Secretory apparatus</td>
<td>invl</td>
<td>3636</td>
</tr>
<tr>
<td>Secretory apparatus ATP synthase</td>
<td>invC</td>
<td>3637</td>
</tr>
<tr>
<td>Invasion protein</td>
<td>invB</td>
<td>3638</td>
</tr>
<tr>
<td></td>
<td>invA</td>
<td>3639</td>
</tr>
<tr>
<td></td>
<td>invE</td>
<td>3640</td>
</tr>
<tr>
<td></td>
<td>invG</td>
<td>3641</td>
</tr>
<tr>
<td></td>
<td>invF</td>
<td>3642</td>
</tr>
</tbody>
</table>
Table 4-14. Continued.

(g) *Type V secretory system*

<table>
<thead>
<tr>
<th>Protein Type</th>
<th>Gene</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autotransporters</td>
<td>estA</td>
<td>4925</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2802</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0219</td>
</tr>
</tbody>
</table>

(h) *Type VI secretory system*

<table>
<thead>
<tr>
<th>Protein Type</th>
<th>Gene</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>OmpA/MotB domain-containing protein</td>
<td></td>
<td>3076</td>
</tr>
<tr>
<td>Lysozyme-like protein</td>
<td>tssE</td>
<td>3077</td>
</tr>
<tr>
<td>Lipoprotein</td>
<td>tssJ</td>
<td>3078</td>
</tr>
<tr>
<td>Secretion protein</td>
<td></td>
<td>3079</td>
</tr>
<tr>
<td>Secretion protein</td>
<td></td>
<td>3080</td>
</tr>
<tr>
<td>Secretion-associated protein</td>
<td></td>
<td>3081</td>
</tr>
<tr>
<td>ImcF domain-containing protein</td>
<td>tssM</td>
<td>3082</td>
</tr>
<tr>
<td>Hypothetical protein</td>
<td></td>
<td>3083</td>
</tr>
<tr>
<td>PAAR repeat-containing protein</td>
<td></td>
<td>3084</td>
</tr>
<tr>
<td>Hypothetical protein</td>
<td></td>
<td>3085</td>
</tr>
<tr>
<td>Hypothetical protein</td>
<td></td>
<td>3086</td>
</tr>
<tr>
<td>Hypothetical protein</td>
<td></td>
<td>3087</td>
</tr>
<tr>
<td>Vgr family protein</td>
<td>vgrG1</td>
<td>3088</td>
</tr>
<tr>
<td>ClpV1 family type VI secretion ATPase</td>
<td>clpV</td>
<td>3089</td>
</tr>
<tr>
<td>Hcp1 family effector</td>
<td>hcp</td>
<td>3090</td>
</tr>
<tr>
<td>OmpA/MotB domain-containing protein</td>
<td></td>
<td>3091</td>
</tr>
<tr>
<td>Hypothetical protein</td>
<td>tssL</td>
<td>3092</td>
</tr>
<tr>
<td>Secretion protein</td>
<td></td>
<td>3093</td>
</tr>
<tr>
<td>EvpB family secretion protein</td>
<td>tssC</td>
<td>3094</td>
</tr>
<tr>
<td>Secretion protein</td>
<td>tssB</td>
<td>3095</td>
</tr>
<tr>
<td>Vgr family protein</td>
<td>vgrG2</td>
<td>1753</td>
</tr>
<tr>
<td>Vgr family protein</td>
<td>vgrG3</td>
<td>3284</td>
</tr>
</tbody>
</table>
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 (Daefler 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 β 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 β 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, pilACD, 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 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>
<thead>
<tr>
<th>Gene</th>
<th>Gene Product</th>
<th>PputUW4</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>pilM</em></td>
<td>type IV pili biogenesis protein PilM</td>
<td>0346</td>
</tr>
<tr>
<td><em>pilN</em></td>
<td>type IV pili biogenesis protein PilN</td>
<td>0347</td>
</tr>
<tr>
<td><em>pilO</em></td>
<td>type IV pili biogenesis protein PilO</td>
<td>0348</td>
</tr>
<tr>
<td><em>pilP</em></td>
<td>type IV pili biogenesis protein PilP</td>
<td>0349</td>
</tr>
<tr>
<td><em>pilQ</em></td>
<td>type IV pili biogenesis protein PilQ</td>
<td>0350</td>
</tr>
<tr>
<td><em>pilF</em></td>
<td>type IV pili biogenesis protein PilF</td>
<td>0940</td>
</tr>
<tr>
<td><em>pilZ</em></td>
<td>type IV pilus assembly PilZ</td>
<td>3510</td>
</tr>
<tr>
<td><em>fimD</em></td>
<td>fimbral biogenesis outer membrane usher protein</td>
<td>4454</td>
</tr>
<tr>
<td><em>pilA</em></td>
<td>type IV pili prepilin PilA</td>
<td>4645</td>
</tr>
<tr>
<td><em>pilC</em></td>
<td>type IV fimbrial assembly protein</td>
<td>4646</td>
</tr>
<tr>
<td><em>pilD</em></td>
<td>type IV pili prepilin peptidase PilD/XcpA</td>
<td>4647</td>
</tr>
<tr>
<td><em>pilE</em></td>
<td>type IV pili biogenesis protein PilE</td>
<td>4663</td>
</tr>
<tr>
<td><em>pilX</em></td>
<td>type IV pilus assembly protein</td>
<td>4665</td>
</tr>
<tr>
<td><em>pilW</em></td>
<td>type IV pilus assembly protein</td>
<td>4666</td>
</tr>
<tr>
<td><em>pilV</em></td>
<td>type IV fimbrial biogenesis protein PilV</td>
<td>4667</td>
</tr>
<tr>
<td><em>fimT</em></td>
<td>type IV pilin</td>
<td>4668</td>
</tr>
<tr>
<td><em>traX</em></td>
<td>conjugal transfer pilus acetylation protein</td>
<td>5018</td>
</tr>
<tr>
<td><em>pilL/chpA</em></td>
<td>type IV pili sensor histidine kinase/response regulator</td>
<td>5106</td>
</tr>
<tr>
<td><em>pilJ</em></td>
<td>type IV pili methyl-accepting chemotaxis transducer</td>
<td>5107</td>
</tr>
<tr>
<td><em>pilH</em></td>
<td>type IV pili signal transduction protein PilI</td>
<td>5108</td>
</tr>
<tr>
<td><em>pilH</em></td>
<td>type IV pili response regulator PilH</td>
<td>5109</td>
</tr>
<tr>
<td><em>pilG</em></td>
<td>type IV pili response regulator PilG</td>
<td>5110</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Species</th>
<th>Putative orthologous shared with <em>P. putida</em> UW4, no. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>P. fluorescens</em> Pf0-1</td>
<td>4113 (76)</td>
</tr>
<tr>
<td><em>P. fluorescens</em> Pf-5</td>
<td>3789 (70)</td>
</tr>
<tr>
<td><em>P. fluorescens</em> SBW25</td>
<td>3595 (66)</td>
</tr>
<tr>
<td><em>P. putida</em> KT2440</td>
<td>3356 (62)</td>
</tr>
<tr>
<td><em>P. putida</em> GB-1</td>
<td>3434 (63)</td>
</tr>
<tr>
<td><em>P. putida</em> F1</td>
<td>3435 (63)</td>
</tr>
<tr>
<td><em>P. putida</em> BIRD-1</td>
<td>3394 (62)</td>
</tr>
<tr>
<td><em>P. putida</em> W619</td>
<td>3405 (63)</td>
</tr>
<tr>
<td><em>P. putida</em> S16</td>
<td>3466 (64)</td>
</tr>
<tr>
<td><em>P. brassicacearum</em> NFM-421</td>
<td>3815 (69)</td>
</tr>
<tr>
<td><em>P. entomophila</em> L48</td>
<td>3387 (62)</td>
</tr>
<tr>
<td><em>P. syringae</em> DC3000</td>
<td>3015 (55)</td>
</tr>
<tr>
<td><em>P. syringae</em> 1448a</td>
<td>2992 (55)</td>
</tr>
<tr>
<td><em>P. syringae</em> B728a</td>
<td>3061 (56)</td>
</tr>
<tr>
<td><em>P. mendocina</em> NK-01</td>
<td>2918 (54)</td>
</tr>
<tr>
<td><em>P. stutzeri</em> A1501</td>
<td>2414 (44)</td>
</tr>
<tr>
<td><em>P. aeruginosa</em> LESB58</td>
<td>3175 (58)</td>
</tr>
<tr>
<td><em>P. aeruginosa</em> PAO1</td>
<td>3182 (58)</td>
</tr>
<tr>
<td><em>P. aeruginosa</em> UCBPP-PA14</td>
<td>3185 (58)</td>
</tr>
<tr>
<td><em>P. aeruginosa</em> PA7</td>
<td>3182 (58)</td>
</tr>
</tbody>
</table>
Table 4-17. Predicted CDSs that share similarities with other *Pseudomonas* sp.

<table>
<thead>
<tr>
<th>locus tag</th>
<th>product</th>
</tr>
</thead>
<tbody>
<tr>
<td>PputUW4_0243</td>
<td>transcription repressor protein</td>
</tr>
<tr>
<td>PputUW4_0802</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0918</td>
<td>type III restriction protein res subunit</td>
</tr>
<tr>
<td>PputUW4_1235</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1655</td>
<td>short chain dehydrogenase</td>
</tr>
<tr>
<td>PputUW4_1656</td>
<td>dehydratase</td>
</tr>
<tr>
<td>PputUW4_1658</td>
<td>3-ketosteroid-5-isomerase</td>
</tr>
<tr>
<td>PputUW4_1662</td>
<td>2,6-dioxo-6-phenylhexa-3-enoate hydrolase</td>
</tr>
<tr>
<td>PputUW4_1670</td>
<td>short-chain dehydrogenase/reductase SDR</td>
</tr>
<tr>
<td>PputUW4_1671</td>
<td>lipase</td>
</tr>
<tr>
<td>PputUW4_1676</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1677</td>
<td>3-oxosteroid-delta-1-dehydrogenase</td>
</tr>
<tr>
<td>PputUW4_1678</td>
<td>short chain dehydrogenase</td>
</tr>
<tr>
<td>PputUW4_1690</td>
<td>2,3-dihydroxybiphenyl-1,2-dioxygenase</td>
</tr>
<tr>
<td>PputUW4_1691</td>
<td>CoA-transferase subunit alpha</td>
</tr>
<tr>
<td>PputUW4_1692</td>
<td>CoA-transferase, beta subunit</td>
</tr>
<tr>
<td>PputUW4_1693</td>
<td>enoyl-CoA hydratase</td>
</tr>
<tr>
<td>PputUW4_1694</td>
<td>2-nitropropane dioxygenase</td>
</tr>
<tr>
<td>PputUW4_1695</td>
<td>enoyl-CoA hydratase</td>
</tr>
<tr>
<td>PputUW4_1696</td>
<td>acyl-CoA dehydrogenase</td>
</tr>
<tr>
<td>PputUW4_1697</td>
<td>acyl-CoA dehydrogenase</td>
</tr>
<tr>
<td>PputUW4_1699</td>
<td>short chain dehydrogenase</td>
</tr>
<tr>
<td>PputUW4_1740</td>
<td>LysR family transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_1815</td>
<td>leucyl-tRNA synthetase</td>
</tr>
<tr>
<td>PputUW4_1893</td>
<td>glutamine synthetase</td>
</tr>
<tr>
<td>PputUW4_1993</td>
<td>AraC family transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_2075</td>
<td>AMP-dependent synthetase and ligase</td>
</tr>
<tr>
<td>PputUW4_2077</td>
<td>AMP-dependent synthetase and ligase</td>
</tr>
<tr>
<td>PputUW4_2080</td>
<td>acyl-CoA dehydrogenase</td>
</tr>
<tr>
<td>PputUW4_2082</td>
<td>enoyl-CoA hydratase</td>
</tr>
<tr>
<td>PputUW4_2197</td>
<td>diguanylate cyclase</td>
</tr>
<tr>
<td>PputUW4_2257</td>
<td>ribosomal-protein-alanine acetyltransferase</td>
</tr>
<tr>
<td>PputUW4_2269</td>
<td>exported protein</td>
</tr>
<tr>
<td>PputUW4_2270</td>
<td>alpha/beta fold family hydrolase</td>
</tr>
<tr>
<td>PputUW4_2280</td>
<td>phenylserine aldolase</td>
</tr>
<tr>
<td>PputUW4_2308</td>
<td>long-chain-fatty-acid--CoA ligase</td>
</tr>
<tr>
<td>PputUW4_2568</td>
<td>plasmid stabilization system protein</td>
</tr>
<tr>
<td>PputUW4_2610</td>
<td>phage integrase</td>
</tr>
<tr>
<td>PputUW4 2622</td>
<td>AraC family transcriptional regulator</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>PputUW4 2624</td>
<td>alpha/beta hydrolase fold family protein</td>
</tr>
<tr>
<td>PputUW4 2646</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 3047</td>
<td>betaine aldehyde dehydrogenase</td>
</tr>
<tr>
<td>PputUW4 3144</td>
<td>aldehyde dehydrogenase</td>
</tr>
<tr>
<td>PputUW4 3257</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 3279</td>
<td>MutT/nudix-family hydrolase</td>
</tr>
<tr>
<td>PputUW4 3554</td>
<td>dTDP-4-dehydrohamnose 3,5-epimerase</td>
</tr>
<tr>
<td>PputUW4 3593</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 3617</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 3618</td>
<td>oxygen-regulated gene required for bacterial internalization</td>
</tr>
<tr>
<td>PputUW4 3619</td>
<td>type III secretion apparatus lipoprotein</td>
</tr>
<tr>
<td>PputUW4 3620</td>
<td>type III secretion system needle complex protein</td>
</tr>
<tr>
<td>PputUW4 3621</td>
<td>type III secretion cytoplasmic protein</td>
</tr>
<tr>
<td>PputUW4 3622</td>
<td>type III secretion system needle complex protein</td>
</tr>
<tr>
<td>PputUW4 3624</td>
<td>invasion protein regulator</td>
</tr>
<tr>
<td>PputUW4 3625</td>
<td>acyl carrier protein</td>
</tr>
<tr>
<td>PputUW4 3626</td>
<td>type III effector protein cell invasion protein</td>
</tr>
<tr>
<td>PputUW4 3627</td>
<td>invasin</td>
</tr>
<tr>
<td>PputUW4 3628</td>
<td>invasin</td>
</tr>
<tr>
<td>PputUW4 3629</td>
<td>chaperone protein SicA</td>
</tr>
<tr>
<td>PputUW4 3630</td>
<td>surface presentation of antigens protein</td>
</tr>
<tr>
<td>PputUW4 3631</td>
<td>surface presentation of antigens protein</td>
</tr>
<tr>
<td>PputUW4 3632</td>
<td>surface presentation of antigens protein</td>
</tr>
<tr>
<td>PputUW4 3634</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 3635</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 3636</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 3640</td>
<td>invasion protein</td>
</tr>
<tr>
<td>PputUW4 3642</td>
<td>invasion protein</td>
</tr>
<tr>
<td>PputUW4 3921</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 4962</td>
<td>transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4 4963</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 5044</td>
<td>hypothetical protein</td>
</tr>
</tbody>
</table>
Table 4-18. Putative unique CDSs in *P. putida* UW4

<table>
<thead>
<tr>
<th>locus tag</th>
<th>product</th>
</tr>
</thead>
<tbody>
<tr>
<td>PputUW4_0044</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0101</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0115</td>
<td>conserved repeat domain protein</td>
</tr>
<tr>
<td>PputUW4_0143</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0144</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0145</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0146</td>
<td>peptidase</td>
</tr>
<tr>
<td>PputUW4_0151</td>
<td>IS3 family transposase</td>
</tr>
<tr>
<td>PputUW4_0204</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0219</td>
<td>extracellular serine protease</td>
</tr>
<tr>
<td>PputUW4_0231</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0239</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0359</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0488</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0499</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0500</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0539</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0544</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0545</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0573</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0594</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0635</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0639</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0640</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0645</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0646</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0647</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0651</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0658</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0677</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0726</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0727</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0737</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0740</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0741</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0861</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0866</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0894</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0895</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Gene Name</td>
<td>Description</td>
</tr>
<tr>
<td>-------------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>PputUW4_0903</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0912</td>
<td>ion channel family protein</td>
</tr>
<tr>
<td>PputUW4_0926</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0968</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1014</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1028</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1030</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1071</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1183</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1199</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1213</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1218</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1234</td>
<td>LuxR family transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_1236</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1238</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1266</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1267</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1314</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1375</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1393</td>
<td>K⁺-transporting ATPase subunit F</td>
</tr>
<tr>
<td>PputUW4_1414</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1487</td>
<td>transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_1488</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1489</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1490</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1499</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1547</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1556</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1557</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1559</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1567</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1619</td>
<td>metallothionein</td>
</tr>
<tr>
<td>PputUW4_1666</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1687</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1706</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1719</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1726</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1730</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1743</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1761</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1812</td>
<td>Class II Aldolase/Adducin Family Protein</td>
</tr>
</tbody>
</table>
Table 4-18. Continued

<table>
<thead>
<tr>
<th>Protein ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PputUW4 1877</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 1921</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 1932</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 1946</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 1959</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 1969</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 2009</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 2015</td>
<td>IS3 family transposase</td>
</tr>
<tr>
<td>PputUW4 2019</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 2021</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 2046</td>
<td>aldehyde dehydrogenase family protein</td>
</tr>
<tr>
<td>PputUW4 2047</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 2097</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 2098</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 2100</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 2106</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 2124</td>
<td>IS1182 family transposase</td>
</tr>
<tr>
<td>PputUW4 2125</td>
<td>ThiF family protein</td>
</tr>
<tr>
<td>PputUW4 2126</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 2128</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 2129</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 2130</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 2131</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 2133</td>
<td>IS1182 family transposase, truncated</td>
</tr>
<tr>
<td>PputUW4 2135</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 2136</td>
<td>IS1182 family transposase</td>
</tr>
<tr>
<td>PputUW4 2147</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 2164</td>
<td>IS1182 family transposase</td>
</tr>
<tr>
<td>PputUW4 2167</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 2169</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 2171</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 2179</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 2183</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 2187</td>
<td>heme peroxidase</td>
</tr>
<tr>
<td>PputUW4 2188</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 2192</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 2202</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 2203</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 2205</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 2216</td>
<td>LysR family transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4 2217</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Gene ID</td>
<td>Description</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>PputUW4_2223</td>
<td>IS1182 family transposase</td>
</tr>
<tr>
<td>PputUW4_2284</td>
<td>hemolysin-type calcium-binding region</td>
</tr>
<tr>
<td>PputUW4_2285</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2288</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2307</td>
<td>LuxR family ATP-dependent transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_2353</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2356</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2371</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2395</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2397</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2415</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2447</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2457</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2481</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2520</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2534</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2535</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2537</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2570</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2574</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2588</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2589</td>
<td>AraC family transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_2601</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2607</td>
<td>AraC family transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_2608</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2609</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2611</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2620</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2621</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2632</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2633</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2657</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2658</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2668</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2674</td>
<td>ion transport 2 domain-containing protein</td>
</tr>
<tr>
<td>PputUW4_2688</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2705</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2725</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2727</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2728</td>
<td>CadC family transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_2729</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>---------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>PputUW4_2730</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2731</td>
<td>AraC family transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_2732</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2733</td>
<td>OmpA-like transmembrane domain protein</td>
</tr>
<tr>
<td>PputUW4_2734</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2750</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2761</td>
<td>DNA-binding response regulator</td>
</tr>
<tr>
<td>PputUW4_2762</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2777</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2807</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2845</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2870</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2928</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2993</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2994</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3001</td>
<td>IS110 family transposase</td>
</tr>
<tr>
<td>PputUW4_3027</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3042</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3048</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3064</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3085</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3088</td>
<td>type VI secretion system Vgr family protein</td>
</tr>
<tr>
<td>PputUW4_3098</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3103</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3130</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3133</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3141</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3142</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3145</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3146</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3159</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3189</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3191</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3214</td>
<td>glutamine amidotransferase class-I</td>
</tr>
<tr>
<td>PputUW4_3226</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3246</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3255</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3270</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3280</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3286</td>
<td>hypothetical protein</td>
</tr>
</tbody>
</table>
Table 4.18. Continued

<table>
<thead>
<tr>
<th>Protein ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PputUW4_3297</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3304</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3311</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3312</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3324</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3329</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3349</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3359</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3416</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3417</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3418</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3421</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3478</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3482</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3486</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3547</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3548</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3551</td>
<td>IS630 family transposase, truncated</td>
</tr>
<tr>
<td>PputUW4_3560</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3587</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3588</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3590</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3591</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3594</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3595</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3597</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3616</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3677</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3678</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3679</td>
<td>surface antigen protein</td>
</tr>
<tr>
<td>PputUW4_3680</td>
<td>surface antigen protein</td>
</tr>
<tr>
<td>PputUW4_3681</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3683</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3763</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3768</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3812</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3954</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3967</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3968</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3996</td>
<td>aldehyde dehydrogenase family protein</td>
</tr>
<tr>
<td>PputUW4_4022</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 4069</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>PputUW4 4083</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 4113</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 4114</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 4154</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 4157</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 4163</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 4164</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 4166</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 4172</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 4194</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 4232</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 4234</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 4311</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 4336</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 4362</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 4363</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 4500</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 4518</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 4519</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 4625</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 4664</td>
<td>type IV pilus-associated protein</td>
</tr>
<tr>
<td>PputUW4 4696</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 4698</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 4828</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 4911</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 4964</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 5019</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 5020</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 5024</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 5088</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 5206</td>
<td>glutamine amidotransferase</td>
</tr>
<tr>
<td>PputUW4 5210</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 5218</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4 5350</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>locus_tag</td>
<td>product</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>PputUW4_0444</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0525</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0644</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0657</td>
<td>ADP-Ribosylglycohydrolase</td>
</tr>
<tr>
<td>PputUW4_0666</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0739</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0863</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0864</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0868</td>
<td>radical SAM family protein</td>
</tr>
<tr>
<td>PputUW4_0869</td>
<td>branched-chain amino acid aminotransferase</td>
</tr>
<tr>
<td>PputUW4_0870</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0871</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0872</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_0873</td>
<td>GntR family transcriptional regulator with aminotransferase domain</td>
</tr>
<tr>
<td>PputUW4_0874</td>
<td>histidine kinase</td>
</tr>
<tr>
<td>PputUW4_0875</td>
<td>two component transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_0886</td>
<td>methionine-R-sulfoxide reductase</td>
</tr>
<tr>
<td>PputUW4_0887</td>
<td>methionine sulfoxide reductase A</td>
</tr>
<tr>
<td>PputUW4_0892</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1333</td>
<td>Glucose/sorbose dehydrogenases</td>
</tr>
<tr>
<td>PputUW4_1334</td>
<td>tat (twin-arginine translocation) pathway signal sequence domain protein</td>
</tr>
<tr>
<td>PputUW4_1373</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1374</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1376</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1561</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1593</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1653</td>
<td>acyl-CoA dehydrogenase</td>
</tr>
<tr>
<td>PputUW4_1654</td>
<td>acyl-CoA dehydrogenase</td>
</tr>
<tr>
<td>PputUW4_1659</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1660</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1661</td>
<td>oxidoreductase, short chain dehydrogenase/reductase family</td>
</tr>
<tr>
<td>PputUW4_1668</td>
<td>LuxR family transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_1669</td>
<td>rieske [2Fe-2S] domain protein</td>
</tr>
<tr>
<td>PputUW4_1674</td>
<td>5'-nucleotidase SurE</td>
</tr>
<tr>
<td>PputUW4_1679</td>
<td>acyl-CoA synthetase</td>
</tr>
<tr>
<td>PputUW4_1680</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1681</td>
<td>FAD-binding oxidoreductase</td>
</tr>
<tr>
<td>PputUW4_1683</td>
<td>fumarate reductase/succinate dehydrogenase</td>
</tr>
<tr>
<td>Genbank Accession</td>
<td>Description</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>PputUW4_1685</td>
<td>acyl-CoA dehydrogenase</td>
</tr>
<tr>
<td>PputUW4_1686</td>
<td>flavin-dependent oxidoreductase</td>
</tr>
<tr>
<td>PputUW4_1729</td>
<td>17 kDa surface antigen</td>
</tr>
<tr>
<td>PputUW4_1735</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1736</td>
<td>arylsulfatase</td>
</tr>
<tr>
<td>PputUW4_1737</td>
<td>multiple antibiotic resistance (MarC)-like protein</td>
</tr>
<tr>
<td>PputUW4_1741</td>
<td>AMP-dependent synthetase and ligase</td>
</tr>
<tr>
<td>PputUW4_1744</td>
<td>enoyl-CoA hydratase</td>
</tr>
<tr>
<td>PputUW4_1808</td>
<td>Citrate lyase subunit beta</td>
</tr>
<tr>
<td>PputUW4_1809</td>
<td>MaoC Domain Protein Dehydratase</td>
</tr>
<tr>
<td>PputUW4_1810</td>
<td>ornithine cyclodeaminase</td>
</tr>
<tr>
<td>PputUW4_1811</td>
<td>D-3-phosphoglycerate dehydrogenase</td>
</tr>
<tr>
<td>PputUW4_1813</td>
<td>histidinol-phosphate aminotransferase</td>
</tr>
<tr>
<td>PputUW4_1858</td>
<td>LuxR family transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_1859</td>
<td>amidohydrolase family protein</td>
</tr>
<tr>
<td>PputUW4_1863</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1864</td>
<td>D-alanyl-D-alanine dipeptidase</td>
</tr>
<tr>
<td>PputUW4_1883</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_1890</td>
<td>amino acid permease family protein</td>
</tr>
<tr>
<td>PputUW4_2012</td>
<td>AraC family transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_2022</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2030</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2033</td>
<td>Arginine/lysine/ornithine decarboxylases</td>
</tr>
<tr>
<td>PputUW4_2034</td>
<td>putrescine transporter</td>
</tr>
<tr>
<td>PputUW4_2053</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2054</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2061</td>
<td>dihydroxy-acid dehydratase</td>
</tr>
<tr>
<td>PputUW4_2066</td>
<td>TetR family transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_2067</td>
<td>short-chain dehydrogenase/reductase SDR</td>
</tr>
<tr>
<td>PputUW4_2068</td>
<td>MaoC-like dehydratase</td>
</tr>
<tr>
<td>PputUW4_2069</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2070</td>
<td>thiolase</td>
</tr>
<tr>
<td>PputUW4_2072</td>
<td>LuxR family MalT-like ATP-dependent transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_2074</td>
<td>enoyl-CoA hydratase/carnitine racemase</td>
</tr>
<tr>
<td>PputUW4_2076</td>
<td>thioesterase superfamily protein</td>
</tr>
<tr>
<td>PputUW4_2079</td>
<td>thiolase</td>
</tr>
<tr>
<td>PputUW4_2081</td>
<td>acyl-CoA dehydrogenase</td>
</tr>
<tr>
<td>PputUW4_2087</td>
<td>alcohol dehydrogenase class III</td>
</tr>
<tr>
<td>PputUW4_2090</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2091</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Gene ID</td>
<td>Description</td>
</tr>
<tr>
<td>------------</td>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>PputUW4_2092</td>
<td>AraC family transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_2096</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2120</td>
<td>RND family efflux transporter, MFP subunit</td>
</tr>
<tr>
<td>PputUW4_2122</td>
<td>peptide ABC transporter permease</td>
</tr>
<tr>
<td>PputUW4_2123</td>
<td>peptide ABC transporter permease</td>
</tr>
<tr>
<td>PputUW4_2134</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2140</td>
<td>HxlR family transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_2142</td>
<td>glutathione-dependent formaldehyde-activating GFA</td>
</tr>
<tr>
<td>PputUW4_2143</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2145</td>
<td>RND family efflux transporter MFP subunit</td>
</tr>
<tr>
<td>PputUW4_2146</td>
<td>RND efflux system outer membrane lipoprotein</td>
</tr>
<tr>
<td>PputUW4_2148</td>
<td>transcriptional regulator ATPase, winged helix family</td>
</tr>
<tr>
<td>PputUW4_2150</td>
<td>TetR family transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_2152</td>
<td>TetR family transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_2154</td>
<td>fumarylacetooacetate (FAA) hydrolase</td>
</tr>
<tr>
<td>PputUW4_2157</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2174</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2185</td>
<td>LuxR family transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_2186</td>
<td>histidine kinase family protein</td>
</tr>
<tr>
<td>PputUW4_2189</td>
<td>sensor histidine kinase</td>
</tr>
<tr>
<td>PputUW4_2190</td>
<td>sensor histidine kinase</td>
</tr>
<tr>
<td>PputUW4_2191</td>
<td>two component LuxR family transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_2195</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2206</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2239</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2264</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2283</td>
<td>twin-arginine translocation pathway signal</td>
</tr>
<tr>
<td>PputUW4_2286</td>
<td>patatin-like phospholipase family protein</td>
</tr>
<tr>
<td>PputUW4_2298</td>
<td>thiolase</td>
</tr>
<tr>
<td>PputUW4_2299</td>
<td>MaoC-like dehydratase</td>
</tr>
<tr>
<td>PputUW4_2300</td>
<td>MaoC-like dehydratase</td>
</tr>
<tr>
<td>PputUW4_2301</td>
<td>3-oxoacyl-[acyl-carrier-protein] reductase</td>
</tr>
<tr>
<td>PputUW4_2331</td>
<td>major facilitator superfamily MFS_1</td>
</tr>
<tr>
<td>PputUW4_2554</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2571</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2573</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2585</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2586</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2587</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Genomic ID</td>
<td>Description</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>PputUW4_2591</td>
<td>LuxR family ATP-dependent transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_2592</td>
<td>acetyl-CoA acetyltransferase</td>
</tr>
<tr>
<td>PputUW4_2597</td>
<td>Radical SAM domain protein</td>
</tr>
<tr>
<td>PputUW4_2598</td>
<td>oxidoreductase</td>
</tr>
<tr>
<td>PputUW4_2599</td>
<td>arylsulfatase</td>
</tr>
<tr>
<td>PputUW4_2600</td>
<td>2,4-dihydroxyacetophenone dioxygenase</td>
</tr>
<tr>
<td>PputUW4_2612</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2625</td>
<td>major facilitator superfamily protein</td>
</tr>
<tr>
<td>PputUW4_2626</td>
<td>enoyl-CoA hydratase/isomerase</td>
</tr>
<tr>
<td>PputUW4_2627</td>
<td>AMP-dependent synthetase and ligase</td>
</tr>
<tr>
<td>PputUW4_2628</td>
<td>carbon monoxide dehydrogenase small subunit</td>
</tr>
<tr>
<td>PputUW4_2630</td>
<td>carbon-monoxide dehydrogenase, large subunit</td>
</tr>
<tr>
<td>PputUW4_2676</td>
<td>TetR family transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_2677</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2681</td>
<td>glutathione-dependent formaldehyde-activating GFA</td>
</tr>
<tr>
<td>PputUW4_2692</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2693</td>
<td>major facilitator family protein</td>
</tr>
<tr>
<td>PputUW4_2701</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2702</td>
<td>MarR family transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_2708</td>
<td>isochorismatase family protein</td>
</tr>
<tr>
<td>PputUW4_2713</td>
<td>cytochrome b561 family protein</td>
</tr>
<tr>
<td>PputUW4_2714</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2716</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2717</td>
<td>cytochrome c oxidase subunit III</td>
</tr>
<tr>
<td>PputUW4_2718</td>
<td>cytochrome c oxidase subunit III</td>
</tr>
<tr>
<td>PputUW4_2719</td>
<td>cytochrome c oxidase, subunit I</td>
</tr>
<tr>
<td>PputUW4_2720</td>
<td>cytochrome c oxidase subunit II</td>
</tr>
<tr>
<td>PputUW4_2723</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2724</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2745</td>
<td>ABC transporter permease protein</td>
</tr>
<tr>
<td>PputUW4_2746</td>
<td>ABC transporter permease protein</td>
</tr>
<tr>
<td>PputUW4_2747</td>
<td>ABC transporter family protein</td>
</tr>
<tr>
<td>PputUW4_2748</td>
<td>ABC transporter family protein</td>
</tr>
<tr>
<td>PputUW4_2749</td>
<td>ABC transporter substrate-binding protein</td>
</tr>
<tr>
<td>PputUW4_2847</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_2995</td>
<td>LuxR family two component transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_3012</td>
<td>IclR family transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_3024</td>
<td>nitrilase</td>
</tr>
<tr>
<td>PputUW4_3026</td>
<td>glutathione S-transferase</td>
</tr>
<tr>
<td>PputUW4_3043</td>
<td>spermidine/putrescine ABC transporter ATPase subunit</td>
</tr>
</tbody>
</table>
Table 4-19. Continued

<table>
<thead>
<tr>
<th>PputUW4_3044</th>
<th>spermidine/putrescine-binding periplasmic protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>PputUW4_3058</td>
<td>prolyl-tRNA synthetase</td>
</tr>
<tr>
<td>PputUW4_3086</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3097</td>
<td>sulfate transporter</td>
</tr>
<tr>
<td>PputUW4_3104</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3127</td>
<td>peroxidase</td>
</tr>
<tr>
<td>PputUW4_3148</td>
<td>phosphotransferase family protein</td>
</tr>
<tr>
<td>PputUW4_3149</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3150</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3151</td>
<td>macrolide glycosyltransferase</td>
</tr>
<tr>
<td>PputUW4_3152</td>
<td>spermidine/putrescine ABC transporter, ATP-binding protein</td>
</tr>
<tr>
<td>PputUW4_3155</td>
<td>spermidine/putrescine-binding periplasmic protein</td>
</tr>
<tr>
<td>PputUW4_3156</td>
<td>GntR family transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_3157</td>
<td>glycosyl transferase family protein</td>
</tr>
<tr>
<td>PputUW4_3158</td>
<td>FAD dependent oxidoreductase</td>
</tr>
<tr>
<td>PputUW4_3172</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3173</td>
<td>ATP-dependent DNA helicase</td>
</tr>
<tr>
<td>PputUW4_3176</td>
<td>succinylglutamate desuccinylase/aspartoacylase</td>
</tr>
<tr>
<td>PputUW4_3177</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3242</td>
<td>glutathione S-transferase</td>
</tr>
<tr>
<td>PputUW4_3325</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3334</td>
<td>alpha/beta hydrolase fold protein</td>
</tr>
<tr>
<td>PputUW4_3335</td>
<td>LysR family transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_3336</td>
<td>demethylmenaquinone methyltransferase</td>
</tr>
<tr>
<td>PputUW4_3338</td>
<td>dimethylmenaquinone methyltransferase</td>
</tr>
<tr>
<td>PputUW4_3345</td>
<td>glycerol-3-phosphate cytidyltransferase</td>
</tr>
<tr>
<td>PputUW4_3346</td>
<td>polysaccharide biosynthesis protein</td>
</tr>
<tr>
<td>PputUW4_3349</td>
<td>glucosyltransferase</td>
</tr>
<tr>
<td>PputUW4_3545</td>
<td>dihydrolipoamide acetyltransferase</td>
</tr>
<tr>
<td>PputUW4_3546</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3598</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_3638</td>
<td>surface presentation of antigens protein</td>
</tr>
<tr>
<td>PputUW4_4107</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_4153</td>
<td>two component LuxR family transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_4192</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>PputUW4_4195</td>
<td>histidine kinase</td>
</tr>
<tr>
<td>PputUW4_4695</td>
<td>GntR family transcriptional regulator</td>
</tr>
<tr>
<td>PputUW4_4697</td>
<td>hypothetical protein</td>
</tr>
</tbody>
</table>

182
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 $1 \times 10^{-5}$. 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^T^ (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\(^T\) (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.
|-----------|----------------|----------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|-----------------|-----------------|------------------|------------------|
| ACRYTCA   | G
| CATGCTGG  | G
| TCCCTGGA  | G
| CAGATTTG  | G
| GTGTTAAC  | G
| ACCAAGCG  | G
| ACCCACTC  | G
| TCTACGTC  | G
| CATGCGTT  | G
| GTGCAGAA  | G
| TGGATCAC  | G
| TCAAATC   | G

|-----------|----------------|----------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|-----------------|-----------------|------------------|------------------|
| CATTATCA  | C
| ACCCAACG  | C
| ACAATCAA  | C
| AACAGCTT  | C
| TTTGACCT  | C
| CTTACCA   | C
| ACCAAGaa  | C
| TCAACACG  | C
| GAGGAGGT  | C
| GTGTAACG  | C
| GCCGAGTC  | C
| TCCCTGGA  | C

|-----------|----------------|----------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|-----------------|-----------------|------------------|------------------|
| AGTGGCTT  | C
| CAGCCCGG  | C
| CCCCGCTT  | C
| CTTACCCA  | C
| ACTGGTAA  | C
| AGTCTTGC  | C
| GCCAGAGG  | C
| TGAACAGG  | C
| AGTCTTGC  | C
| CTTACCCA  | C
| ACTGGTAA  | C
| GCCAGAGG  | C

|-----------|----------------|----------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|-----------------|-----------------|------------------|------------------|
| ACTGGTAA  | A
| CTTACCCA  | A
| AGTCTTGC  | A
| GCCAGAGG  | A
| TGAACAGG  | A
| AGTCTTGC  | A
| GCCAGAGG  | A
| TGAACAGG  | A
| AGTCTTGC  | A
| GCCAGAGG  | A
| TGAACAGG  | A
| AGTCTTGC  | A

|-----------|----------------|----------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|-----------------|-----------------|------------------|------------------|
| RGGGTATT  | A
| GCCAAGCC  | A
| TATCGGTA  | A
| GTGACTGG  | A
| GAACTGCA  | A
| AATCAAAT  | A
| TCAACATG  | A
| TACCAAGC  | A
| GAGGAATG  | A
| TGGATTAC  | A
| CCGCTGAC  | A
| TCAACATG  | A
| TACCAAGC  | A

191
Table 4-20. 16S rRNA Genes Sequences Pairwise Identities.

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

ID: Identical
only one copy of 16S rRNA gene from each of the type strain, IAM12022 and IAM1236T, 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

E. coli K12 MG1655

P. aeruginosa UCBPP PA14

P. aeruginosa PAO1

P. aeruginosa LESB58

P. aeruginosa PA7

P. stutzeri A1501

P. mendocina ymp

P. syringae phaseolicola 1448A

P. syringae glycinea

P. syringae syringae B728a

P. syringae tomato DC3000

P. putida UW4

P. fluorescens Pf01

P. fluorescens Pf5

P. fluorescens SBW25

P. entomophila L48

P. putida GB1

P. putida W619

P. putida KT2440

P. putida F1

P. putida PCL1445
C. GrpE

E. coli K12 MG1655

P. putida F1

P. syringae tomato DC3000

P. stutzeri A1501

P. entomophila L48

P. aeruginosa PA7

P. mendocina ymp

P. syringae phaseolicola 1448A

P. fluorescens Pf01

P. putida UW4

0.98

P. fluorescens SBW25
E. GroES

- E. coli K12 MG1655
- P. syringae syringae B728a
  - P. syringae phaseolicola 1448A
  - P. syringae tomato DC3000
    - P. aeruginosa PAO1
    - P. aeruginosa PA7
      - P. stutzeri A1501
        - P. mendocina ymp
  - P. putida GB1
    - P. entomophila L48
      - P. putida KT2440
        - P. putida F1
          - P. putida W619
    - P. fluorescens Pf5
      - P. fluorescens Pf01
        - P. fluorescens SBW25
          - P. putida UW4
F. ClpB

- P. putida UW4
  - P. fluorescens Pf01
  - P. fluorescens SBW25
  - P. fluorescens Pf5
  - P. syringae syringae B728a
  - P. syringae phaseolicola 1448A
  - P. syringae tomato DC3000
  - P. putida KT2440
    - P. putida F1
    - P. putida GB1
    - P. putida W619
    - P. entomophila L48
      - P. stutzeri A1501
        - P. mendocina ymp
        - P. aeruginosa PAO1
          - P. aeruginosa LESB58
            - P. aeruginosa UCBPP PA14
              - P. aeruginosa PA7

- E. coli K12 MG1655

Distance-scale: 0.2
G. IbνA
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
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). LESP58 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.**

<table>
<thead>
<tr>
<th><em>Pseudomonas</em> sp.</th>
<th>Genome Size, bp</th>
<th>CDS no.</th>
<th>Pseudogenes</th>
<th>G+C content, %</th>
<th>Coding density, %</th>
<th>tRNA no.</th>
<th>rRNA genes (operon no.)</th>
<th>Plasmid no.</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>UW4</td>
<td>6,183,388</td>
<td>5,430</td>
<td>18</td>
<td>60.92</td>
<td>87.4</td>
<td>72</td>
<td>22 (7)</td>
<td>—</td>
<td>This study</td>
</tr>
<tr>
<td><em>aeruginosa</em> LESB58</td>
<td>6,601,757</td>
<td>5,925</td>
<td>34</td>
<td>66.3</td>
<td>88.4</td>
<td>67</td>
<td>13 (4)</td>
<td>—</td>
<td>Winstanley et al. 2009</td>
</tr>
<tr>
<td><em>aeruginosa</em> PA7</td>
<td>6,588,339</td>
<td>6,286</td>
<td>8</td>
<td>66.5</td>
<td>89.5</td>
<td>63</td>
<td>12 (4)</td>
<td>—</td>
<td>Roy et al. 2010</td>
</tr>
<tr>
<td><em>aeruginosa</em> PAO1</td>
<td>6,264,404</td>
<td>5,566</td>
<td>5</td>
<td>66.5</td>
<td>89.3</td>
<td>63</td>
<td>13 (4)</td>
<td>—</td>
<td>Stover et al. 2000</td>
</tr>
<tr>
<td><em>aeruginosa</em> UCBPP-PA14</td>
<td>6,537,648</td>
<td>5,892</td>
<td>none</td>
<td>66.3</td>
<td>89.4</td>
<td>59</td>
<td>13 (4)</td>
<td>—</td>
<td>Lee et al. 2006</td>
</tr>
<tr>
<td>brassicacearum subsp. brassicacearum NFM421</td>
<td>6,843,248</td>
<td>6,097</td>
<td>N/A</td>
<td>60.8</td>
<td>88.2</td>
<td>65</td>
<td>16 (5)</td>
<td>—</td>
<td>Ortet et al. 2011</td>
</tr>
<tr>
<td>entomophila L48</td>
<td>5,888,780</td>
<td>5,169</td>
<td>N/A</td>
<td>64.2</td>
<td>89.1</td>
<td>78</td>
<td>22 (7)</td>
<td>—</td>
<td>Vodovar et al. 2006</td>
</tr>
<tr>
<td>fluorescens Pf-5</td>
<td>7,074,893</td>
<td>6,144</td>
<td>N/A</td>
<td>63.3</td>
<td>88.8</td>
<td>71</td>
<td>16 (5)</td>
<td>—</td>
<td>Paulsen et al. 2005</td>
</tr>
<tr>
<td>fluorescens Pf0-1</td>
<td>6,438,405</td>
<td>5,741</td>
<td>9</td>
<td>60.6</td>
<td>89.8</td>
<td>73</td>
<td>19 (6)</td>
<td>—</td>
<td>Silby et al. 2009</td>
</tr>
<tr>
<td>fluorescens SBW25</td>
<td>6,722,539</td>
<td>6,009</td>
<td>88</td>
<td>60.5</td>
<td>88.3</td>
<td>66</td>
<td>16 (5)</td>
<td>—</td>
<td>Silby et al. 2009</td>
</tr>
<tr>
<td>mendocina NK-01</td>
<td>5,434,353</td>
<td>4,958</td>
<td>N/A</td>
<td>62.5</td>
<td>88.7</td>
<td>65</td>
<td>12 (4)</td>
<td>—</td>
<td>Guo et al. 2011</td>
</tr>
<tr>
<td>putida BIRD-1</td>
<td>5,731,541</td>
<td>5,124</td>
<td>N/A</td>
<td>61.7</td>
<td>86.9</td>
<td>64</td>
<td>22 (7)</td>
<td>—</td>
<td>Matilla et al. 2011</td>
</tr>
<tr>
<td>putida F1</td>
<td>5,959,964</td>
<td>5,300</td>
<td>49</td>
<td>61.9</td>
<td>88.7</td>
<td>76</td>
<td>19 (6)</td>
<td>—</td>
<td>Wu et al. 2010</td>
</tr>
<tr>
<td>putida GB-1</td>
<td>6,078,430</td>
<td>5,417</td>
<td>8</td>
<td>61.9</td>
<td>89.4</td>
<td>74</td>
<td>22 (7)</td>
<td>—</td>
<td>Wu et al. 2010</td>
</tr>
<tr>
<td>putida KT2440</td>
<td>6,181,863</td>
<td>5,420</td>
<td>N/A</td>
<td>61.6</td>
<td>86.7</td>
<td>73</td>
<td>22 (7)</td>
<td>—</td>
<td>Nelson et al. 2002</td>
</tr>
<tr>
<td>putida S16</td>
<td>5,984,790</td>
<td>5,218</td>
<td>N/A</td>
<td>62.3</td>
<td>84.9</td>
<td>70</td>
<td>19 (6)</td>
<td>—</td>
<td>Yu et al. 2011</td>
</tr>
<tr>
<td>putida W619</td>
<td>5,774,330</td>
<td>5,471</td>
<td>26</td>
<td>61.4</td>
<td>88.9</td>
<td>75</td>
<td>22 (7)</td>
<td>—</td>
<td>Wu et al. 2010</td>
</tr>
<tr>
<td>stutzeri A1501</td>
<td>4,567,418</td>
<td>4,146</td>
<td>N/A</td>
<td>63.8</td>
<td>89.8</td>
<td>61</td>
<td>12 (4)</td>
<td>—</td>
<td>Yan et al. 2008</td>
</tr>
<tr>
<td>syringae pv. phaseolicola 1448A</td>
<td>5,928,787</td>
<td>5,144</td>
<td>N/A</td>
<td>58</td>
<td>87</td>
<td>64</td>
<td>16 (5)</td>
<td>2</td>
<td>Joardar et al. 2005</td>
</tr>
<tr>
<td>syringae pv. syringae B728a</td>
<td>6,093,698</td>
<td>5,137</td>
<td>47</td>
<td>59.2</td>
<td>88.5</td>
<td>64</td>
<td>16 (5)</td>
<td>—</td>
<td>Feil et al. 2005</td>
</tr>
<tr>
<td>syringae pv. tomato str. DC3000</td>
<td>6,397,126</td>
<td>5,615</td>
<td>N/A</td>
<td>58.4</td>
<td>86.8</td>
<td>63</td>
<td>15 (5)</td>
<td>2</td>
<td>Buell et al. 2003</td>
</tr>
</tbody>
</table>
surface. Strain Pf-5 is a soil bacteria 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 alignate 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\(^{2+}\) 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 rhizopheric 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. syi*angae 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. syi*angae 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 (Pf01_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 \textit{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.

\subsection{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 \textit{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, \textit{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.

\subsection{5.2.5 Acetoin}

Bacterial volatile compounds such as acetoin can stimulate the growth of plants such as \textit{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 O₂, 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; Cvríková 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. brassicaearum* 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 *phal*. 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, *phal* 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$_2$ 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 β-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


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


Belimov AA, Safronova VI, Sergeyeva TA, Egorova TN, Matveyeva VA, Tsyganov VE, Borisov


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-

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


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 CHA0. Microbiology 146:2417-2424

Bnayahu BY. 1991. Root excretions and their environmental effects: influence on availability of


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


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


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


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


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


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 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


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


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


Duijff BJ, Gianinazzi-Pearson V, Lemanceau P. 1997. Involvement of the outer membrane lipopolysaccharides in the endophytic colonization of tomato roots by biocontrol


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


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


Ge L and Seah SY. 2006. Heterologous expression, purification, and characterization of an L-
ornithine N⁵-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

supercoiling catalyzed by DNA gyrase. Proceedings of the National Academy of Sciences
of the United States of America 73:4474-4478

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

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


Goldstein A. 1995. Recent progress in understanding the molecular genetics and biochemistry of calcium phosphate solubilization by Gram negative bacteria. Biological Agriculture and


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


of the assimilatory nitrite-nitrate reductase operon (nasAB) of *Azotobacter vinelandii*. Molecular Microbiology 18:579-591


Hahn HP. 1997. The type-4 pilus is the major virulence-associated adhesion of *Pseudomonas aeruginosa*-a review. Gene 192:99–108


Hass D and Défago G. 2005. Biological control of soil-borne pathogens by fluorescent


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


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


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-i


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


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


Lazdunski AM, Venture I, Sturgis J. 2004. Regulatory circuits and communication in Gram-
negative bacteria. Nature Reviews Microbiology 2:581-592


radish by *Pseudomonas fluorescens*. Phytopathology 86:149-155


Letham DS. 1963. Zeatin, a factor inducing cell division from *Zea mays*. Life Sciences 8:569-573


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


and conventional transformation-mediated mutagenesis. Applied and Environmental Microbiology 65:2558–2564


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


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


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


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


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


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


Symbiosis 21:263-274
Podile AR and Prakash AP. 1996. Lysis and biological control of \textit{Aspergillus niger} by \textit{Bacillus subtilis} AF1. Canadian Journal of Microbiology 42:533-538
Prayitno J, Rolfe BG, Mathesius U. 2006. The Ethylene-insensitive sickle mutant of \textit{Medicago}
truncatula shows altered auxin transport regulation during nodulation. Plant Physiology 142:168–180


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


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$_2$-fixer) and _Bacillus licheniformis_ (P-solubilizer), both from a semiarid mangrove rhizosphere. FEMS Microbiology Ecology 35:181-187


N₂-fixing and phosphate solubilizing bacteria. Plant and Soil 265:123-129


Sevilla M, Burris RH, Gunapala N, Kennedy C. 2001. Comparison of benefit to sugarcane plant growth and $^{15}$N$_2$ incorporation following inoculation of sterile plants with Acetobacter diazotrophicus wild-type and Nif-mutant strains. Molecular Plant-Microbe Interactions 14:358-366


Soil Biology & Biochemistry 35:1615-1623


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


polymyxa. Soil Biology & Biochemistry 31:1847-1852

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


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


Valverde A, Burgos A, Fiscella T, Rivas R, Velázquez E, Rodriguez-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


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


Molecular Plant-Microbe Interactions 10:716-724


Vasanthakumar A and McManus PS. 2004. Indole-3-acetic acid-producing bacteria are associated with cranberry stem gall. Phytopathology 94:1164-1171


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. Microbiology 155:95-105

2006. Complete genome sequence of the entomopathogenic and metabolically versatile soil bacterium *Pseudomonas entomophila*. Nature Biotechnology 24:673-679


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


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


Appendix A. Examples of primers used for gaps closure.

<table>
<thead>
<tr>
<th>Oligo name</th>
<th>Sequence</th>
<th>Estimated Gap Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-5</td>
<td>GACTATGCCGAGCGTTACTGG</td>
<td>2619</td>
</tr>
<tr>
<td>1-3</td>
<td>GCCTGCTTGGATGACAAACAC</td>
<td></td>
</tr>
<tr>
<td>2-5</td>
<td>CAAATGGCTCCACAACACTCGC</td>
<td>341</td>
</tr>
<tr>
<td>2-3</td>
<td>CATCCACTACACGCACCAAC</td>
<td></td>
</tr>
<tr>
<td>3-5</td>
<td>GCAGGTCTGGAGACGAAGGTG</td>
<td>1357</td>
</tr>
<tr>
<td>3-3</td>
<td>GTGGGCAAGTTTGAGGTTT</td>
<td></td>
</tr>
<tr>
<td>4-5</td>
<td>GGTCATCGGCAATCTTTTCG</td>
<td></td>
</tr>
<tr>
<td>4-3</td>
<td>CGCCAGGTCGATGCAAGT</td>
<td>68</td>
</tr>
<tr>
<td>5-5</td>
<td>CGCCTGGGATTTCTGTTT</td>
<td></td>
</tr>
<tr>
<td>5-3</td>
<td>AGCGGCATTGAACCCCTG</td>
<td></td>
</tr>
<tr>
<td>6-5</td>
<td>GGATGGCGTGGCGTTTCTA</td>
<td></td>
</tr>
<tr>
<td>6-3</td>
<td>CGTTGGAGGTGTATGGAGGC</td>
<td>4578</td>
</tr>
<tr>
<td>7-5</td>
<td>GGCTCGGCACAAATACTCCA</td>
<td></td>
</tr>
<tr>
<td>7-3</td>
<td>CAAACCGGATGCGAAGAA</td>
<td>1197</td>
</tr>
<tr>
<td>8-5</td>
<td>GGTCGTGTGCTGTGTTGTTT</td>
<td></td>
</tr>
<tr>
<td>8-3</td>
<td>GACCGCACCACCCAGCTAAA</td>
<td></td>
</tr>
<tr>
<td>9-5</td>
<td>GCCCAAAACGCTGCTCAGTCA</td>
<td></td>
</tr>
<tr>
<td>9-3</td>
<td>CGCTGTGAGAATGCTG</td>
<td>5220</td>
</tr>
<tr>
<td>10-5</td>
<td>GCCGTGGCGTTTTCCTATT</td>
<td></td>
</tr>
<tr>
<td>10-3</td>
<td>TGTTGGCTTGAGGCGATA</td>
<td></td>
</tr>
<tr>
<td>11-5</td>
<td>GAGTTTCGTGGCGGATT</td>
<td></td>
</tr>
<tr>
<td>11-3</td>
<td>CGCCGACTCCAGCAAGAT</td>
<td>20</td>
</tr>
<tr>
<td>12-5</td>
<td>GCCCTGTGCAATCTGCTGT</td>
<td></td>
</tr>
<tr>
<td>12-3</td>
<td>AGCACTGCTTGCTGCTG</td>
<td>1361</td>
</tr>
<tr>
<td>13-5</td>
<td>GCCGTTGTCCATCCACCA</td>
<td></td>
</tr>
<tr>
<td>13-3</td>
<td>CGTCAGCAGCAAGGCAAT</td>
<td>547</td>
</tr>
<tr>
<td>14-5</td>
<td>CAAATGCGCTCGGTGCTG</td>
<td></td>
</tr>
<tr>
<td>14-3</td>
<td>GGGGTTTTGATTCTTGGCTG</td>
<td>1245</td>
</tr>
<tr>
<td>15-5</td>
<td>ATGAGCAGCAGCAGCAACAA</td>
<td></td>
</tr>
<tr>
<td>15-3</td>
<td>CAGGTTAAATGGCGCAAAC</td>
<td>5349</td>
</tr>
<tr>
<td>16-5</td>
<td>GGAGTTTGCGACCGCCTGA</td>
<td></td>
</tr>
<tr>
<td>16-3</td>
<td>CAAATCGGAGGAAAGCCCA</td>
<td>1628</td>
</tr>
<tr>
<td>17-5</td>
<td>AGACTCTGCTGCTGCTG</td>
<td></td>
</tr>
<tr>
<td>17-3</td>
<td>GCCGCTGAGCAGCAGACCGATT</td>
<td>351</td>
</tr>
<tr>
<td>18-5</td>
<td>GCGGCGAATGTGGGATATCG</td>
<td></td>
</tr>
<tr>
<td>18-3</td>
<td>CGCTTTTCGGCGCAAACCTCT</td>
<td>590</td>
</tr>
<tr>
<td>19-5</td>
<td>CGCGGAGATTGCTGGATT</td>
<td></td>
</tr>
<tr>
<td>19-3</td>
<td>CGCCTTCTGCTGCTGTTG</td>
<td>20</td>
</tr>
<tr>
<td>20-5</td>
<td>CGGGGTTTACAGTTCTGCTC</td>
<td>826</td>
</tr>
<tr>
<td>20-3</td>
<td>CCCTTGAAATGCTTGCTGTTG</td>
<td></td>
</tr>
<tr>
<td>21-5</td>
<td>GCTGCCCAACGACTGAATAG</td>
<td></td>
</tr>
<tr>
<td>21-3</td>
<td>GCTGGTGACCAGGTATAG</td>
<td></td>
</tr>
<tr>
<td>22-5</td>
<td>CCAAAGGAGGTTGCGGGAGA</td>
<td></td>
</tr>
<tr>
<td>22-3</td>
<td>TGGTGATGCGGGGATGTGA</td>
<td></td>
</tr>
<tr>
<td>23-5</td>
<td>AACTGGGCCGGCGTGAGACT</td>
<td></td>
</tr>
<tr>
<td>23-3</td>
<td>GCCTGCCCCATCTCTGCGTGTA</td>
<td></td>
</tr>
<tr>
<td>24-5</td>
<td>CACTGACCAGGACCGAC</td>
<td></td>
</tr>
<tr>
<td>24-3</td>
<td>CCTGACCCGAGCCAAATCC</td>
<td></td>
</tr>
<tr>
<td>25-5</td>
<td>CGGACACCCACCAAATACAG</td>
<td></td>
</tr>
<tr>
<td>25-3</td>
<td>GCCCTGGGACCATCAAGACC</td>
<td></td>
</tr>
<tr>
<td>26-5</td>
<td>CGAGAAAGCCGCAAGTGAGG</td>
<td></td>
</tr>
<tr>
<td>26-3</td>
<td>AGCGAGCAACCGAAAACCC</td>
<td></td>
</tr>
<tr>
<td>27-5</td>
<td>GCTGGTGTGGTGCGGAGTGAGT</td>
<td></td>
</tr>
<tr>
<td>27-3</td>
<td>TAACCACCGCTGCCTCTTGG</td>
<td></td>
</tr>
<tr>
<td>28-5</td>
<td>GGATGATGAAGTGCAGCCAAGCC</td>
<td></td>
</tr>
<tr>
<td>28-3</td>
<td>GACCGACGGGCTTCGTGGGC</td>
<td></td>
</tr>
<tr>
<td>29-5</td>
<td>TCTTTCTCGGCCCTTGACGC</td>
<td></td>
</tr>
<tr>
<td>29-3</td>
<td>TGATTCCGTCGCCCTTGAGCC</td>
<td></td>
</tr>
<tr>
<td>30-5</td>
<td>GCACGCAAGACTCCTCAAACA</td>
<td></td>
</tr>
<tr>
<td>30-3</td>
<td>CGTTGAGCGAGCGGGAAAT</td>
<td></td>
</tr>
<tr>
<td>31-5</td>
<td>GCCGTCAGCAGCAATCCCA</td>
<td></td>
</tr>
<tr>
<td>31-3</td>
<td>GCCATTGCCTGCTACACCA</td>
<td></td>
</tr>
<tr>
<td>32-5</td>
<td>CTGGGTTTTGTGGTGTGATGAGG</td>
<td></td>
</tr>
<tr>
<td>32-3</td>
<td>AGTTACTGGGAAGGCTTGGA</td>
<td></td>
</tr>
<tr>
<td>33-5</td>
<td>CTGGTGAAGCCAACCGAGAT</td>
<td></td>
</tr>
<tr>
<td>33-3</td>
<td>GCTCCGCCCTTGAAACATCG</td>
<td></td>
</tr>
<tr>
<td>34-5</td>
<td>TGATGGCTGCCCCTTGGTG</td>
<td></td>
</tr>
<tr>
<td>34-3</td>
<td>CCAAAGGCTGCCGCTGCA</td>
<td></td>
</tr>
<tr>
<td>35-5</td>
<td>TGTTGCGGGTGATGCTGGC</td>
<td></td>
</tr>
<tr>
<td>35-3</td>
<td>GCCTGGAAATACTCGCTGGTG</td>
<td></td>
</tr>
<tr>
<td>36-5</td>
<td>CGAAGAGGGCAGGAGCAT</td>
<td></td>
</tr>
<tr>
<td>36-3</td>
<td>CAGCAGTTTGGAGCGTGTT</td>
<td></td>
</tr>
<tr>
<td>37-5</td>
<td>CTACAGTTTCAGCGTGGCA</td>
<td></td>
</tr>
<tr>
<td>37-3</td>
<td>GTTGGCTGCCATGCTGGTA</td>
<td></td>
</tr>
<tr>
<td>38-5</td>
<td>AAACCTGTGCCTGCGTATCA</td>
<td></td>
</tr>
<tr>
<td>38-3</td>
<td>GCTGGCGGACCGTGGTTT</td>
<td></td>
</tr>
<tr>
<td>39-5</td>
<td>CAAAGGAGGCGTGGCGTGGT</td>
<td></td>
</tr>
<tr>
<td>39-3</td>
<td>GCCTGGCGGACCCATTACCT</td>
<td></td>
</tr>
<tr>
<td>40-5</td>
<td>GCAATTACCGGACCGACA</td>
<td></td>
</tr>
<tr>
<td>40-3</td>
<td>CGCTTTGACTCATACCCCTACCT</td>
<td></td>
</tr>
<tr>
<td>296</td>
<td>41-5</td>
<td>ACCTCTGAAGCCTGGACG</td>
</tr>
<tr>
<td>296</td>
<td>41-3</td>
<td>TCGCATCTGTAGGAGCATC</td>
</tr>
<tr>
<td>296</td>
<td>42-5</td>
<td>TGGCGGTTTCGGCTATCAA</td>
</tr>
<tr>
<td>296</td>
<td>42-3</td>
<td>ACTGCTCTCGTGCTGGGTTCG</td>
</tr>
<tr>
<td>296</td>
<td>43-5</td>
<td>GTGAGCGGACAACATTTCG</td>
</tr>
<tr>
<td>296</td>
<td>43-3</td>
<td>GCGAATGTTTGCCCTGTT</td>
</tr>
<tr>
<td>296</td>
<td>44-5</td>
<td>GGCACAGGGATGGAGGAAGG</td>
</tr>
<tr>
<td>296</td>
<td>44-3</td>
<td>CCGGGCAAGGCATTGCTACTC</td>
</tr>
<tr>
<td>296</td>
<td>45-5</td>
<td>CGTCGGAAATCGCCACAGG</td>
</tr>
<tr>
<td>296</td>
<td>45-3</td>
<td>CCCAATGCGTGCTAAATCCC</td>
</tr>
<tr>
<td>296</td>
<td>46-5</td>
<td>GCCACCTTCTTCAGCGCAATACC</td>
</tr>
<tr>
<td>296</td>
<td>46-3</td>
<td>TTCTCGCCAGCCAGACC</td>
</tr>
<tr>
<td>296</td>
<td>47-5</td>
<td>AGAAGGCCCTCAACCTGTGCC</td>
</tr>
<tr>
<td>296</td>
<td>47-3</td>
<td>GCCCTTCTCGCAACCTCATCC</td>
</tr>
<tr>
<td>296</td>
<td>48-5</td>
<td>CCACTGCTGCTTTGTCTCCG</td>
</tr>
<tr>
<td>296</td>
<td>48-3</td>
<td>TGCTGCCAGGAACAAACCC</td>
</tr>
<tr>
<td>296</td>
<td>49-5</td>
<td>TCACGTTCAACCGCAACCA</td>
</tr>
<tr>
<td>296</td>
<td>49-3</td>
<td>GGTCAGCGCCAGGGACATCG</td>
</tr>
<tr>
<td>296</td>
<td>50-5</td>
<td>TCGGTGAGTGCTGCTGCC</td>
</tr>
<tr>
<td>296</td>
<td>50-3</td>
<td>GGGCGATGAACGAAAGGAAT</td>
</tr>
<tr>
<td>296</td>
<td>51-5</td>
<td>TTGGCGTCCAGGGAGTTCG</td>
</tr>
<tr>
<td>296</td>
<td>51-3</td>
<td>GCAGCGAGTTGTGTTGTA</td>
</tr>
<tr>
<td>296</td>
<td>52-5</td>
<td>TGGAGGAACCCCGCAAGA</td>
</tr>
<tr>
<td>296</td>
<td>52-3</td>
<td>GCAAAAGCCTATCGGTTCAGC</td>
</tr>
<tr>
<td>296</td>
<td>53-5</td>
<td>GCCGTCTGTTGTTGTCTTG</td>
</tr>
<tr>
<td>296</td>
<td>53-3</td>
<td>CTGGAATGGGCGCACACG</td>
</tr>
<tr>
<td>296</td>
<td>54-5</td>
<td>TTGCTTTTGCTGCGGTTC</td>
</tr>
<tr>
<td>296</td>
<td>54-3</td>
<td>GGGTCAGGGTGCCGGTGTAG</td>
</tr>
<tr>
<td>296</td>
<td>55-5</td>
<td>GCCGATGTAGGGTTGTCGG</td>
</tr>
<tr>
<td>296</td>
<td>55-3</td>
<td>GCCATTTTTCGTCTCAACCTC</td>
</tr>
<tr>
<td>296</td>
<td>56-5</td>
<td>GGATTTAGTCGTTGAAGCC</td>
</tr>
<tr>
<td>296</td>
<td>56-3</td>
<td>CGGTATGGGCGTGAGGAGA</td>
</tr>
<tr>
<td>296</td>
<td>57-5</td>
<td>GCATAGACCATACGCCCCGACAA</td>
</tr>
<tr>
<td>296</td>
<td>57-3</td>
<td>CTGCGGTCTGTTGCCAAAAG</td>
</tr>
<tr>
<td>296</td>
<td>58-5</td>
<td>GCCACCTCGGACAGTCCCTT</td>
</tr>
<tr>
<td>296</td>
<td>58-3</td>
<td>AGCCAGGCCCAGGACAGA</td>
</tr>
<tr>
<td>296</td>
<td>59-5</td>
<td>CCCTACCTCCTGCTCAAACCTG</td>
</tr>
<tr>
<td>296</td>
<td>59-3</td>
<td>GCCCACTGGAATGCTGTATAT</td>
</tr>
<tr>
<td>296</td>
<td>60-5</td>
<td>ACACGCAACCTGAGGAATAC</td>
</tr>
<tr>
<td>296</td>
<td>60-3</td>
<td>GCTACGGCGGAGTACGAGT</td>
</tr>
<tr>
<td>296</td>
<td>61-5</td>
<td>CCGAGCAACAAACCCACCC</td>
</tr>
</tbody>
</table>
297

61-3  CCAGTGAGGCATTGAAGACG
62-5  CGGCAGATTGTGGTTTGGC
62-3  GGATTGAGCATTGCAGTGGGC
63-5  AGCCCGCCGAGGTGTATTA
63-3  GTCTGAGCCTCCGCTCGT
64-5  CACCACGCAACGGAACCA
64-3  CGGTGAGCAGCGAAATGGAC
65-5  GTGGTGCTTCTGAGCCTGTT
65-3  CCGACGCCGAGGAGTGTG
66-5  CTTTGAGTGGCCTGAGATGG
66-3  ATGGGCTGCACTTGCAGTC
67-5  GGACGAAACAGGCAAGAACA
67-3  GAGCAATGACGCGGCGAGTG
68-5  GCCTGTGGTGGCTGCTTG
68-3  CTTCCGCAAGTGGCTGTCC
69-5  GCTGGAGAATCTTGACG
69-3  CACGGAACTCTCCTCAAAC
70-5  TCGGCTTGGGTGGCTGATG
70-3  GCACGCAACCCCGAGCCTAT
71-5  TGTAAGATTCTGCGGGAAGGC
71-3  GCGCTTCTCTGCTTCTCAC
72-5  CGCACCACGCTGAGCAGAAG
72-3  TGACGAGGACGGCGATAGA
73-5  CGGTCCAGTACGCTCGTGTT
73-3  GCTGACAGTGGACGAGACGC
74-5  TCGCAATCATCATCGCACTC
74-3  GATGACCCTGGCGAAGATGG
75-5  CAGGGAAGCCGCACTCATC
75-3  TGTCCGGCTGGCGAGATTTT
76-5  TCCCGCAACTGGTGCAAGG
76-3  GCAACTCAGCCAGGTGCTGTCT
77-5  CCAATCCACACACCTGACCC
77-3  CCTTGACGGGCATGGAAA
78-5  CAGGTTGCGATGTTATGCG
78-3  GGCGTGAGGGCCGATGAC
79-5  GCCTGGAAGATTCGGCAACTG
79-3  CGCATCAACCGACTCTCGCA
80-5  CACCTGTAGCCAGCAGCTC
80-3  AGCGCGAGGGATGGAGTG
81-5  CGAGGTCCTTGAGCGCGGTGAT
81-3  GGCGGCAACCGCATCTCAA

2272
344
6032
169
1357
676
249
385
430
3435
388
2537
582
81
707
1225
683
1245
1245
393
<table>
<thead>
<tr>
<th></th>
<th>DNA Sequence</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>82-5</td>
<td>TGTCACAGGACAGACAGAC</td>
<td>1361</td>
</tr>
<tr>
<td>82-3</td>
<td>GGGAAGGAAACTAAATGGCTC</td>
<td></td>
</tr>
<tr>
<td>83-5</td>
<td>AGGCCGAAGAAGATGCGAAGATG</td>
<td>20</td>
</tr>
<tr>
<td>83-3</td>
<td>TCAGAAGGCGACGACGAC</td>
<td></td>
</tr>
<tr>
<td>84-5</td>
<td>TTCCACCTGTGACACCA</td>
<td>1124</td>
</tr>
<tr>
<td>84-3</td>
<td>GTTTCGCGGCGGATGTTGT</td>
<td></td>
</tr>
<tr>
<td>85-5</td>
<td>GGGCGAGTTCTTGGCTACG</td>
<td>5670</td>
</tr>
<tr>
<td>85-3</td>
<td>TGTGCCTGCTCCAA</td>
<td></td>
</tr>
<tr>
<td>86-5</td>
<td>CGCAACAAACCGAACCTG</td>
<td>359</td>
</tr>
<tr>
<td>86-3</td>
<td>GACTCGCAAGCGGTTTC</td>
<td></td>
</tr>
<tr>
<td>87-5</td>
<td>CAGGTCGGATATGGATGTG</td>
<td>308</td>
</tr>
<tr>
<td>87-3</td>
<td>GTTACGGTGCAAGGGGT</td>
<td></td>
</tr>
<tr>
<td>88-5</td>
<td>CGCTCGGGGACTTATCTCG</td>
<td>5370</td>
</tr>
<tr>
<td>88-3</td>
<td>CTTCTCCGCTCGGTTTCG</td>
<td></td>
</tr>
<tr>
<td>89-5</td>
<td>CGGAACCTGAAGACGCTACTAC</td>
<td>459</td>
</tr>
<tr>
<td>89-3</td>
<td>CGATGGCTTTTGAAGTAGAGGG</td>
<td></td>
</tr>
<tr>
<td>90-5</td>
<td>GCTCAAGGCAAGCGACC</td>
<td>494</td>
</tr>
<tr>
<td>90-3</td>
<td>GGGCAAGCAAGGCGAG</td>
<td></td>
</tr>
<tr>
<td>91-5</td>
<td>AATGCTCGCTCGCTCTTCA</td>
<td>290</td>
</tr>
<tr>
<td>91-3</td>
<td>CTTCTGGCTCAACCTGCTGCT</td>
<td></td>
</tr>
<tr>
<td>92-5</td>
<td>ACCTCAGGGGCACGCGAATC</td>
<td>1361</td>
</tr>
<tr>
<td>92-3</td>
<td>GCTGGCTTGGCGAAGAA</td>
<td></td>
</tr>
<tr>
<td>93-5</td>
<td>GGAAGGATGCCGGGTT</td>
<td></td>
</tr>
<tr>
<td>93-3</td>
<td>GCACCCAGTCCAAAGGCTCA</td>
<td>1357</td>
</tr>
<tr>
<td>94-5</td>
<td>CGCCACGCTGTCAAGGGA</td>
<td>320</td>
</tr>
<tr>
<td>94-3</td>
<td>TTGCGGCACTGAAACCTG</td>
<td></td>
</tr>
<tr>
<td>95-5</td>
<td>ACGAACAGGGCGGAGATC</td>
<td>1361</td>
</tr>
<tr>
<td>95-3</td>
<td>TGCCITCGCTATCTCCTTC</td>
<td></td>
</tr>
<tr>
<td>96-5</td>
<td>GGCCTTGCGCTGGGTAGAA</td>
<td>1197</td>
</tr>
<tr>
<td>96-3</td>
<td>ACGAACGAAACGGCGACA</td>
<td></td>
</tr>
<tr>
<td>97-5</td>
<td>AATACGGGCACGGCTCC</td>
<td>470</td>
</tr>
<tr>
<td>97-3</td>
<td>CTGGAGTTGGCGGAAAGC</td>
<td></td>
</tr>
<tr>
<td>98-5</td>
<td>CTGGCTGCTTTGATTGC</td>
<td>4982</td>
</tr>
<tr>
<td>98-3</td>
<td>GGTGTTCCAAGCCTATG</td>
<td></td>
</tr>
<tr>
<td>99-5</td>
<td>GGCAGCCTTATGGAATTCCG</td>
<td>4097</td>
</tr>
<tr>
<td>99-3</td>
<td>CCCGAGCGGCTGAGCAGA</td>
<td></td>
</tr>
<tr>
<td>100-5</td>
<td>CGACAGAAGCGGTCAGAC</td>
<td>10000</td>
</tr>
<tr>
<td>100-3</td>
<td>GAACAGCGTGGGGAGAT</td>
<td></td>
</tr>
</tbody>
</table>
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.