Table 4.14. Putative restriction modification systems with predicted recognition sequences in *P. mendocina* S5.2. Restriction, specificity and methylation subunits are represented by R, M and S, respectively.

<table>
<thead>
<tr>
<th>Type</th>
<th>Subunit</th>
<th>Localization</th>
<th>Enzyme Name</th>
<th>Predicted Recognition Sequence</th>
<th>Genome Coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>I R</td>
<td>Chromosome</td>
<td>Pme5IP</td>
<td>CCCNNNNNNNTGCG</td>
<td>4807737-4810787</td>
<td></td>
</tr>
<tr>
<td>I S</td>
<td>Chromosome</td>
<td>S.Pme5I</td>
<td>CCCNNNNNNNTGCG</td>
<td>4810792-4811952</td>
<td></td>
</tr>
<tr>
<td>I M</td>
<td>Chromosome</td>
<td>M2.Pme5I</td>
<td>CCCNNNNNNNTGCG</td>
<td>4811949-4813463</td>
<td></td>
</tr>
<tr>
<td>I M</td>
<td>Chromosome</td>
<td>M1.Pme5I</td>
<td>CCCNNNNNNNTGCG</td>
<td>4814493-4815992</td>
<td></td>
</tr>
<tr>
<td>II RM</td>
<td>Plasmid</td>
<td>RM.Pme5II</td>
<td>GACGAG</td>
<td>115726-118473</td>
<td></td>
</tr>
<tr>
<td>II M</td>
<td>Plasmid</td>
<td>M.Pme5ORFCP</td>
<td>CAGCTG</td>
<td>40755-42116</td>
<td></td>
</tr>
<tr>
<td>II M</td>
<td>Plasmid</td>
<td>M.Pme5ORFBP</td>
<td>CCNGG</td>
<td>50220-52031</td>
<td></td>
</tr>
</tbody>
</table>

Legend: ➡: Other open reading frame (ORFs)

Figure 4.18 Components of Type I R-M system in *P. mendocina* S5.2 as predicted by REBASE. Methyltransferase catalytic site is represented by dppy whereas S-adenosylmethionine (SAM) binding site is depicted as fgg.
Legend:

**dppy**: Methyltransferase catalytic site

**fgg**: S-adenosylmethionine (SAM) binding site

**Figure 4.19**  RM.Pme5II gene consisting of fused restriction (R) and methylation (M) domains located in pPME5 plasmid of *P. mendocina* S5.2 as predicted by REBASE. Methyltransferase catalytic site is represented by dppy whereas S-adenosylmethionine (SAM) binding site is depicted as fgg.

**Figure 4.20.** Locations of RM.Pme5II and possible prophage sequences in pPME5 plasmid of *P. mendocina* S5.2.
4.9 Absence of GACG\textsuperscript{m6} AG DNA Methylation Sites in Cured Strain S5.2

As described previously, continuous culturing in non-selective LB medium has resulted in the occurrence of cured \textit{P. mendocina} strain S5.2 as denoted by the presence (PP) and absence (NP) of plasmid pPME5. Such event was also accompanied by loss of 3,560 GACG\textsuperscript{m6} AG methylation sites (Figure 4.21), mediated by the plasmid-borne RM.Pme5II (Table 4.14 and Figure 4.19) after methylome analysis of the plasmidless derivative of S5.2 (Table 4.13).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure421.png}
\caption{Distribution map of GACG\textsuperscript{m6} AG modification motif in \textit{P. mendocina} strain S5.2 complete chromosome and \textit{P. mendocina} strain S5.2 plasmid. Lost of these methylation sites were accompanied by curing of pPME5 plasmid from \textit{P. mendocina} S5.2.}
\end{figure}
4.9.1 Characterization of GACG\textsuperscript{m6}AG DNA Methylation Sites and Its Association with Metal Resistance Genes

To further investigate the possible influences of GACG\textsuperscript{m6}AG methylation particularly on the metal resistance traits of \textit{P. mendocina} S5.2, the methylation sites on both intragenic and intergenic regions across the genome was first disseminated using the intersection procedure from BEDTools (Chapter 3.8.2). Table 4.15 summarizes the number of methylated and unmethylated sites, segregated across the intragenic and intergenic regions of the genome. Among the 399 (11.2\%) methylated sites identified in the intergenic regions, 2 sites were detected upstream from start codon of a multicopper oxidase (\textit{cumA}) gene and putative chromate transport (\textit{srpC}) gene (Table 4.16). Prediction of the promoter regions of the genes were performed using PromoterHunter module from phiSITE database (Klucar \textit{et al.}, 2010).

\begin{table}[h]
\centering
\caption{Summary of GACG\textsuperscript{m6}AG methylation sites distributed across the intragenic and intergenic regions of \textit{P. mendocina} S5.2 genome.}
\begin{tabular}{llllll}
\hline
 & \multicolumn{2}{c}{Chromosome} & \multicolumn{2}{c}{pPME5 Plasmid} & \\
 & Methylated & Unmethylated & Methylated & Unmethylated & Total \\
\hline
Intragenic & 3,189 & 5 & 66 & 2 & 3,261 \\
Intergenic & 365 & 0 & 33 & 0 & 399 \\
\hline
Total & 3,454 & 5 & 99 & 2 & 3,560 \\
\hline
\end{tabular}
\end{table}
Table 4.16. Detection of GACG$^{m6}$AG methylation sites in the intergenic regions of metal resistance determinants in the *P. mendocina* S5.2 genome.

<table>
<thead>
<tr>
<th>Coding Sequence (Prokka)</th>
<th>ORF</th>
<th>Metal Resistance</th>
<th>Intergenic</th>
<th>Distance from Start Codon</th>
<th>Modified Genome Coordinate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prokka02271</td>
<td><em>cumA</em></td>
<td>Cu$^{2+}$</td>
<td>1 (Start)</td>
<td>70</td>
<td>2,413,083</td>
</tr>
<tr>
<td>Prokka03183</td>
<td><em>srpC</em></td>
<td>Cr$^{6+}$</td>
<td>1 (Start)</td>
<td>120</td>
<td>3,386,149</td>
</tr>
</tbody>
</table>

4.9.2 Unmethylated Sites of pPME5 Plasmid

Analysis on the unmethylated sites revealed that GACGAG motif within a type IV pilin biogenesis gene (Prokka00342) was found to be one of the two unmethylated intragenic site within the pPME plasmid. This gene was previously described to be associated with higher motility phenotype of strain S5.2 (Chapter 4.7.1). On the other hand, another unmethylated intragenic site was detected within a hypothetical protein.
CHAPTER 5: DISCUSSION

5.1. Isolation and Identification of Pseudomonas Strains

Selection of vineyard soil from Riquewihr, France, as isolation source was originally to assess the bacteria communities capable of degrading quorum sensing signals in previous studies. Based on the observation that mostly Pseudomonas strains were isolated, speculations on the vineyard soil chemistry that drove the prevalence of this genus among the soil inhabitants was raised. Preliminary assessment on the copper resistance traits (Chapter 4.2.1) has demonstrated the elevated capacity of these strains to grow in the presence of high copper content, suggesting the strong association of widely reported copper-based pesticide usage that could influence the microbial diversity of the vineyard soil. Of note, P. mendocina strain S5.2 and P. putida strain S13.1.2 displayed the most notable resistance hence they were selected for subsequent investigations. Systematic identification procedure involving 16S rRNA gene phylogenetic, peptide mass spectrometry fingerprinting and average nucleotide identity (ANI) analyses in the study have successfully validated the species circumscription on the selected strains as P. mendocina strain S5.2 and P. putida strain S13.1.2.

Pseudomonas mendocina is a non-fluorescent denitrifying Pseudomonas species that was first isolated in 1970 from soil and water samples that were collected in Mendoza, Argentina (Palleroni et al. 1970). Although P. mendocina is found in the environment, this species is also a rare human pathogen due to its association with opportunistic infections such as sepsis, endocarditis and spondylodiscitis (Chi et al., 2005; Mert et al., 2007; Nseir et al., 2011). From the biotechnological point of view, the potential use of P. mendocina in various industrial applications have been demonstrated such as in the production of green plastics (Guo et al., 2011; Li et al., 2013), biodegradation of polymer wastes (Mao et al., 2013; Z. Wang et al., 2013), bioreduction of toxic tellurite (Rajwade & Paknikar, 2003), wastewater treatment (Feng...
et al., 2012) and as a biocontrol agent of cyanobacterial blooms (Shi et al., 2009). *P. mendocina* was also implied as a plant growth promoting rhizobacterium (Kohler et al., 2006). Of note, this work represented the first comprehensive description on the heavy metal resistant traits and genetic determinants in *P. mendocina*.

*Pseudomonas putida* is a generally non-pathogenic bacterium renowned for its metabolic versatility and low nutritional demand (Timmis, 2002). Frequently isolated from polluted soils, the environmental adaptability of *P. putida* has lead to its prominence in biotechnological applications, ranging from biodegradation of toxic compounds to synthesis of chemicals and bio-based polymers (PHA) (Poblete-Castro et al., 2012). Elaboration on the complete genome of *P. putida* was first documented on *P. putida* strain KT2440, a plant growth-promoting and potential bioremediation agent with activity against organic pollutants in soil (Nelson et al., 2002; Regenhardt et al., 2002). The following *in silico* survey on the genome sequence revealed that strain KT2440 harbours an array of genetic determinants possibly involved in metal resistance or homeostasis. These include systems for arsenic (*arsRBCH*), divalent cations (*cadA* and *czc*), nickel (*nikABCDE*), copper (*cus* and *cop*) and more (Cánovas et al., 2003) to which were similarly observed in *P. putida* strain S13.1.2 later in the study.
Deciphering the Genomic Architecture of *Pseudomonas* Strains

As starting point for functional genomics, the availability of complete genome sequences offers advantages in accurate definition of gene coordinates and distances, or recognition of paralogous gene families during genome annotation procedures (Fraser *et al.*, 2002). The use of SMRT sequencing in this work has successfully generated the complete genome sequences and assemblies of *P. mendocina* strain S5.2, consisting of a circular chromosome and a linear plasmid pPME5 whereas *P. putida* strain S13.1.2 consisting of a circular chromosome.

One of the key advantages of SMRT sequencing approach is the unbiased and long read lengths that permit sequencing of extended repetitive and GC biased nucleotide sequences, hence elevating the capacity to complete bacterial genomes (Roberts *et al.*, 2013). In comparison, other technologies such as sequencing by synthesis (SBS) engage in a more cost-effective sequencing of usually smaller fragments (up to 600 bp) that involves amplification processes for enhancement of sequencing coverage. The amplification steps however, generally lead to compromised sequence quality with the occurrence of GC biased genome sequences due to poor amplification the GC-rich or poor sequences or repeats (Aird *et al.*, 2011).

In view of this, the requirement of SMRT sequencing to close both GC-rich *Pseudomonas* genomes was supported in this study. Thus, acquisition of completed *Pseudomonas* genomes has subsequently enabled further polishing processes, such as genome circularization and sequence rearrangement according to the origin of replication. Thus, downstream bioinformatics analyses involving precise definition and profiling of determinants associated with heavy metal resistance were greatly facilitated.

Another handiness of long-read sequencing in this work was the ability to clearly distinguish plasmid from chromosomal assemblies of *P. mendocina* strain S5.2.
Also, in the effort to overcome the possible shortcomings in sequencing coverage and accuracy of SMRT sequencing, employment of the latest P6/C4 sequencing chemistry in this study has yielded high sequencing coverage of more than 150× with accuracy exceeding 99.99% (Q40) (Table 4.5). Besides, the interpretation of polymerase kinetics during the sequencing run has additionally allowed direct detection and identification of DNA methylation profiles in both *Pseudomonas* strains (Flusberg et al., 2010).

### 5.3 Comprehensive Profiling of Copper Resistance Determinants

Presence of multiple chromosome-encoded genes associated with copper homeostasis, resistance and transport was identified in both *Pseudomonas* complete genomes. In relation, both strains exhibited noticeable resistance to copper sulphate *in vivo*. Notably, the most prominent resistance up to 3.5 mM of CuSO₄ by *P. putida* strain S13.1.2 could be explained with a series of chromosomal-encoded *cop* and *cus* genes.

Identification of *cop* genes strongly suggested the occurrence of a resistance mechanism based on protein-mediated sequestration and cellular accumulation of the copper ions in the cell (Cooksey, 1993). Notably, it is also highly possible that the transport of copper is further facilitated by *cus* operon. Activation of this transport mechanism is likely mediated by the *copR1* and *copS* genes located upstream of the *cus* gene cluster that encodes the transcriptional activator and sensor kinase. In *Pseudomonas syringae*, this pair forms a two-component regulatory system whereby phosphorylation of CopR by CopS induces the expression of the copper resistance operon (Mills et al., 1993). Another set of transcriptional activator and sensor kinase genes (*copR2* and *cusS*) was also found in the genome. However sequence alignments between both gene pairs showed a low identity (52.9 %) and similarity (66.1 %) at the
protein level that suggested their involvement in dissimilar copper resistance mechanisms.

5.4 Ecology of Copper Resistance in Vineyard Soil

The need for chemical pesticides in Alsatian vineyards involved the usage of copper sulphate in the form of Bordeaux mixture for decades and up to 400-500 mg/kg of copper was reported in soil (Brun et al., 1998). As such, anthropogenic accumulation of copper likely explains the predominance of soil inhabitants carrying the determinants involved in acquisition, efflux, sequestration or cellular distribution of copper (Andreasza et al., 2010).

Surface layer of soil poses the highest level of copper due to its limited mobility (Angelova et al., 1999; Coscione et al., 2009). Therefore, isolation of Pseudomonas strains at this depth (~5 cm) reflects the nature of their copper resistance traits. As Pseudomonas was widely reported as part of the grapevine-associated microbiota, the importance of these beneficial taxa in this specific biogeography of vineyard soil was supported (Salomon et al., 2014; West et al., 2010; Zarraonaindia et al., 2015). These features suggestively explained the prevalence of pseudomonads in the vineyard soil to upkeep the growth of plants treated with copper sulphate over the years in order to control fungal diseases. In the light of such observations, the adaptive behaviour of the selected P. mendocina and P. putida strains were further explored in this study.
5.5  Potential Adaptation to Other Heavy Metals

In addition to determinants and traits indicating the importance of copper resistance in vineyard soils, both strains also shared their resistance towards arsenic, nickel, cobalt and cadmium. Phenotypic display and the presence of nik, cad, czc and ars operons, together with other related genes also drove speculation on the tenacity of other vineyard soil chemistries in this study. Of note, vineyards have also been treated with sodium arsenate till the end of the last century, and vineyard posts in several parts of the world have been treated with a mixture of copper-chromium-arsenic salts. As a consequence, chromium and arsenic salts in vineyards soils and surrounding fields may remain at detectable concentrations (Robinson et al., 2006). Hence identification of ars operon and resistance in this strain suggested the persistence of arsenic in this vineyard soil environment.

The ars operons are quite diverse (Branco et al., 2008) in pseudomonads such as P. aeruginosa (Cai et al., 1998) and P. putida (Fernández et al., 2014). In response to the presence of arsenite, the transcriptional repressor ArsR bound to the cognate promoter is released, followed by the subsequent induction of the ars gene expression (Busenlehner et al., 2003; Murphy & Saltikov, 2009). These include the transmembrane efflux pump ArsB that extrudes arsenite and the arsenate reductase ArsC that converts As(V) to As(III), this later being readily transported out of the cell by ArsB (Cai et al., 1998; Jackson & Dugas, 2003). To date, no defined functions were assigned to the NADPH-dependent FMN reductase ArsH. It was suggested to respond to the oxidative stress caused by arsenite and recently, ArsH been demonstrated to oxidize trivalent organoarsenical herbicides to pentavalent species (Chen et al., 2015; Hervás et al., 2012). Intriguingly, since ars determinants are responsible for both arsenic and antimony (Sb) resistance (Branco et al., 2008; Cai et al., 1998; Carlin et al., 1995), strain S13.1.2 was tested for resistance to Sb(III) salt and found to be sensitive. This
may imply a single substrate specificity of *P. putida* strain S13.1.2 ArsR protein towards arsenic and it is also possible that presence of As(III) is required to confer resistance to Sb(III).

5.6 Carbon Metabolism in Relation to Root Exudation Profile and Heavy Metals

Tartaric, malic, oxalic, lactic, citric and succinic acids are among the main organic acids, together with glucose, sucrose and fructose as main sugars detected in grapevines and root exudates across various genotypes (Cançado *et al.*, 2015; Dharmadhikari, 1994; Kliewer, 1966; López-Rayo *et al.*, 2015; Mato *et al.*, 2007). The capability of both strains to utilize these organic acids and sugars were determined using Biolog Phenotype Microarray Analysis. Both *P. mendocina* strain S5.2 and *P. putida* strain S13.1.2 displayed distinctive metabolic profiles of carbon sources although only differ in utilization of a few organic acids substrates related to grapevine and root exudates. Such finding of distinctive specificity towards different carbon sources may indicate the differential localization in the native rhizosphere. This may also imply the preferences of each *Pseudomonas* strains in utilizing these carbon compounds for *in-situ* remediation of copper and other metal ions in vineyards. As various carbon sources serve differently as effective electron donor for a given metal resistance, such implication could be exemplified by the higher reduction rate of hexavalent chromium Cr(VI) and trivalent iron Fe(III) by *Cellulomonas* sp. ES6 in presence of molasses rather than pure sucrose, together with requirement of various carbon sources chromium reduction by *Klebsiella* sp. PB6 and even bacterial consortium from dichromate contaminated sediments (Field *et al.*, 2013; Smith *et al.*, 2002; Wani & Omozele, 2015).

Of note, changes in organic acid exudates concentrations and profiles were reportedly driven by metal contents in soil. For instance, under the influence of varying
Fe conditions, fluctuation in oxalic, tartaric and ascorbic acid exudations were observed in grapevine rootstocks (López-Rayo et al., 2015). In another similar study, changes in citric acid exudation levels were found to be correlated with Al resistance phenotype of ‘Kober 5BB’ grapevine genotype (Villa et al., 2009). In relation to elevated soil copper content, alteration of the root exudation profiles followed by microbial composition favouring those capable of metabolizing these exudates are highly probable. As such, it is crucial to correlate soil copper level with organic acid exudation in grapevine roots in the effort to elucidate the physiological responses of grapevine that drives the changes in microbial composition in soil.

5.7 Description on Plasmid pPME5 in *P. mendocina* strain S5.2

5.7.1 Linearity of Plasmid pPME5 and Its Features

As indicated in Chapter 4.6, collective observations from sequence assembly, PFGE and PCR amplification attempts have evidently verified the linearity of plasmid pPME5. To the best of my knowledge, this study entailed the first glance of linear megaplasmid-harbouring *P. mendocina*. The closest proximity of similar observation was demonstrated in *P. putida* strain AJ, containing linear megaplasmid with similar size of approximately 260 kb that enabled the metabolism of vinyl chloride (VC), ethene, and ethylene oxide as sole carbon source (Danko et al., 2004). Interestingly, prolonged exposure to these hydrocarbons as sole carbon source reportedly led to plasmid rearrangements, exemplified by appearance of two smaller plasmids (~80 kb and ~100 kb) and one larger plasmid (~390 kb) after 1 year cultivation on VC. Such observation hence postulated the required linearity of plasmid pPME5 to undergo rearrangements in response to vineyard soil environment though the factors were yet to be determined.
5.7.2 Replication and Copy Number of Plasmid pPME5

In reference to copy number of plasmid pPME5, *in silico* annotation of potential genes involved in replication of pPME5 has also provided indication on the copy number of plasmid pPME5. For instance, identification of genes encoding ParA (Prokka00082) and ParB (Prokka00083) chromosomal partitioning protein has led to the speculation of the low copy number of plasmid pPME5. This pairing of proteins were shown to be essential for proper positioning of low copy number plasmids in dividing progeny cells during cell replication (Bartosik *et al.*, 2009; Bignell & Thomas, 2001). Additionally, the nearly 1:1 ratio of sequencing coverage between the chromosome and plasmid (Table 4.5) has suggested the occurrence of pPME5 as low copy plasmid in *P. mendocina* strain S5.2 (Antipov *et al.*, 2016).

Still, much remained to be uncovered including the unknown features on protection against exonuclease cleavage of terminal nucleotides of pPME5 plasmid. Following the described linearity, presumption on the presence of terminal inverted repeats (TIR) that occurs in most linear plasmids was annulled through several investigations in this study. For start, analysis for possible occurrence of terminal inverted repeats (TIR) using DNA folding form of the Mfold web server showed low thermal stability of secondary DNA structure thus the unlikelihood of TIR regions at both ends of the plasmid (Zuker, 2003). Coverage plot for the plasmid assembly from SMRT sequencing was also known to provide clues on the presence of TIR, denoted by the spike in coverage at the extremity of the given contig (Gomez-Escribano *et al.*, 2015). Presence of coverage spike however, was not detected in the assembly of plasmid pPME5 (Appendix).
5.7.3 Curing of Plasmid pPME5 and the Potential Inferences on Vineyard Soil Ecology

Continuous culturing in non-selective LB medium has resulted in the occurrence of cured derivatives of strain S5.2 as denoted by the presence (PP) and absence (NP) of plasmid pPME5 in the various clones (Figure 4.16). Such event that yielded viable derivatives under laboratory conditions has nevertheless led to speculation on the essentiality of this plasmid for the survival in vineyard soil environment. However, factors contributing to the selectivity remain undetermined with majority of coding regions in the plasmid (82.32%) not assigned to any identification. A plausible hypothesis involves mercury resistance and bioconversion that is provided by presence of mer operon in pPME5. Mercury-containing pesticides have been banned for almost three decades from French vineyards but traces may remain in soils as observed in other locations worldwide (Reeder et al., 2006).

Aside from the in silico prediction on functions of pPME5 in vineyards soil environment, one of the direct physiological changes of P. mendocina following loss of plasmid pPME5 was the incapacitated twitching ability of the cured derivative. Of note, the identification of gene clusters associated with twitching motility in the plasmid (Chapter 4.7.1) suggested its role in early stage of biofilm formation through cell-to-surface interaction and microcolony development (Heydorn et al., 2002; O'Toole & Kolter, 1998). As stated in the literatures describing the communal mechanism of biofilm-mediated heavy metal resistance, it is thinkable that P. mendocina strain S5.2 required this replicon for the enhancement of resistance traits especially towards mercury.
5.8 Methylome of *P. mendocina* Strain S5.2

Apart from completion of genomes sequences, the advancement of SMRT sequencing technology has enabled genome-wide detection of methylation patterns at single base resolution (I. A. Murray *et al.*, 2012). Since then, prospering list of studies revolving around methylomes of bacteria was reported, particularly on their roles in bacterial pathogenesis. The addition of methylome study of *P. mendocina* strain S5.2 to the collection however, had our interests skewed towards the environmental aspects focusing on heavy metal resistance. This study served as a foundation to elucidate the possible functions of DNA methylation in regulating heavy metal resistance. As described in Chapter 4.8, the genome-wide detection of DNA methylation coupled with REBASE prediction have identified three 6mA and a 4mC methylation patterns governed by Type I and Type II RM systems in *P. mendocina* strain S5.2.

5.8.1 Uniqueness of Type I RM System

The Type I RM system of *P. mendocina* strain S5.2 is chromosomal encoded, consisting of two methyltransferase (M) subunits (M1.Pme5I and M2.Pme5I) that methylate bipartite sequences of 5’-CGC<sup>m6</sup>ANNNNNNGGG-3’ (motif 1) and the complementary 5’-C<sup>m4</sup>CCNNNNNTGCG-3’ (motif 2). Occurrence of motifs without modification will subsequently lead to cleavage by the restriction endonuclease (S) subunit (Pme5IP) fuelled by ATP, SAM and Mg<sup>2+</sup> as cofactor (Eskin & Linn, 1972). The restriction activity relies entirely on target recognition domain (TRD) of the specificity (S) subunit (S.Pme5I), for cognitive recognition of the given sequence motif. Conversely, combination of M and S subunits is sufficient for methylation activity (Janscak *et al.*, 1998).
Thus, the formation of the $R_1M_2S_1$ complex of Type I RM system in *P. mendocina* strain S5.2 has represented a unique stoichiometry of subunits, as opposed to the commonly observed $R_2M_2S_1$ complex (Janscak *et al.*, 1998). Comparative analysis in REBASE also indicated uniqueness of this system denoted by its absence in other known *P. mendocina* species. Interestingly, this Type I RM system of *P. mendocina* strain S5.2 showed resemblance in terms of both stoichiometry and sequence similarity, with uncultured *Desulfofustis* sp. PB-SRB1 derived from marine metagenome (Wilbanks *et al.*, 2014).

### 5.8.2 Plasmid-borne Type II RM System

*P. mendocina* strain S5.2 also codes for a Type II RM system, namely RM.Pme5II that recognized and methylated a total of 3,542 sites in the genome harbouring 5’-GACGAG-3’ motifs. REBASE analysis later designated RM.Pme5II as part of Type IIG subdivision of RM system and the corresponding gene was actually pPME5 plasmid-borne. The main attributes of Type IIG RM system involved the methylation and restriction activity that was carried out in a single polypeptide while the cleavage occurs 14-21 bp from the recognition site at a -3’ direction (Roberts *et al.*, 2003). The recognition sequence could be asymmetrical as shown in this study, as opposed to palindromic sequence recognition of Type II enzymes in general. Also, methylation only occurred in one strand of the recognized sequence (I. A. Murray *et al.*, 2012), which was consistent with the appearance of single stranded methylation in all 5’-GACG$^m_6$AG-3’ motifs as displayed by in this study (Appendix 10).
5.8.2.1 Assessment on Loss of DNA Methylation Sites in Cured S5.2

In search of direct effects on the methylome of *P. mendocina* strain S5.2 following the curing of pPME5 plasmid, another SMRT sequencing and methylome analysis was conducted on the cured derivative. Indeed, the absence of 3,542 5’-GACG<sup>5m6</sup>AG-3’ motifs (Table 4.13) was apparent, thus confirming the function of the plasmid-borne RM.Pme5II. Moreover, as both native and cured strains shared similar growth rates in rich medium, it is unlikely that RM.Pme5II mediates segregational stability or postsegregational killing of plasmid as other RM systems do (Kulakauskas *et al.*, 1995). In addition, PHAST (Phage Search Tool) showed that RM.Pme5II was not localized within any intact prophage sequences in pPME5, hence annulling the hypothesis that it was a prophage-derived RM system (Figure 4.21). As such, the significance of RM.Pme5II was hypothetically related to the survival of this strain in vineyard soil environment. One of the speculations revolves around the transcription regulation of essential genes by RM.Pme5II.

5.8.2.2 Correlation Between GACGAG Methylation and Gene Expression

Variation in DNA methylation states were known to affect gene expression levels by altering the binding affinities of regulatory proteins (Bendall *et al.*, 2013; Low *et al.*, 2001). In this case, presence of methylated sites located in the intergenic regions, upstream from transcription start site of a multicopper oxidase (*mco*) gene and putative chromate transport (*srpC*) gene indicates possible transcriptional regulation of these genes, hence reflecting additional implication on chromate and copper resistance by plasmid pPME5.
On the other hand, the 5'-GACGAG-3’ motif within the intragenic region of type IV biogenesis gene (Prokka00342) was amongst the two unmethylated sites detected in pPME5 plasmid. The expression of motility phenotype was observed in LB medium (Figure 4.18) and it was used for culture of P. mendocina strain S5.2 prior to harvest for SMRT sequencing. Hence, it was suggestive of a possible connection between occurrence of unmethylated DNA and twitching motility gene expression although the motif was observed in the intragenic region.

5.8.3 Unknown RM enzymes for BCTGCAGV motifs

No RM enzymes were found to modify 5'-BCTGCG<sup>m6</sup>AGV-3’ motifs in the genome of strain S5.2. In fact, specificity towards this motif reportedly resembled the 5'-CTGCAG-3’ sequence recognition of commercially available PstI restriction endonuclease isolated from Providencia staurtii (Chater, 1977). Following the notion that candidate genes responsible for the motif could be present in any missed out small plasmids during sequencing procedures, another SMRT sequencing attempt was conducted without the library size selection step (Chapter 3.6.2). Based on previous observations in our laboratory, such procedure bypass has allowed the retention of previously eliminated small plasmids. However the resulting HGAP assembly showed no contigs corresponded to any small plasmids, thus indicating that the unidentified gene candidate was indeed located in the chromosome or plasmid pPME5. In addition, the lower methylated motif percentage of less than 90% (Table 4.13) could be resulted from competitive binding of the unidentified methyltranferase with another DNA-binding protein. Similar results of low methylated motif and absence of any corresponding RM enzymes was also reported on 5'-GTTC<sup>m4</sup>CG-3’ motifs of P. putida strain S13.1.2 (Appendix 14).
5.9 Future Work

Further investigation will be conducted with accurate quantification of resistance toward each metal salts due the limited information provided in Biolog Microplates, to which only approximate range of concentrations were disclosed as displayed in Appendix 2. In addition, changes in metabolic profiles of grapevine root exudates by *Pseudomonas* strains will also be elucidated in presence of metal ions in the inoculum. Also with *in silico* assignment of functions, the current listing of metal resistance genes has provided a basis for experimental gene isolation, which is required for proper attribution on the exact functions of the genes.

Additional methylome analysis will be required to monitor the changes in methylome profile following exposure to heavy metals, including the assessment on methylation states of essential genes involved in heavy metal resistance and twitching motility. In relation to gene expression levels, comparison between the transcriptomes of the native and derivatives in absence of a given RM systems will allow further insights on roles RM system for survival fitness of the strains in vineyard soil environments. This could be achieved via site directed mutagenesis followed by RNA-seq analysis. However, it is worth noting that removal of certain RM system may alter biology and viability of cells.
CHAPTER 6: CONCLUSION

Investigation on microbiota inhabiting the vineyard soil in Riquewihr, France has led to the isolation and identification of copper resistant *P. mendocina* strain S5.2 and *P. putida* strain S13.1.2. Phenotypic microarray analysis combined with complete genome sequencing have demonstrated the versatility and adaptation of both strains towards other heavy metals including nickel, cobalt, cadmium, zinc and arsenic. Thus the persistent effect of the vineyard soil contamination by these heavy metals was hypothesized in this work. Comparatively, both strains also displayed distinctive metabolic profiles towards carbon sources related to grapevine root exudates. Subsequently, the correlation between grapevine root exudation profile and metal resistant abilities of *Pseudomonas* strains has been established in this work.

This work has also reported the complete sequence and characterization of a linear megaplasmid pPME5 carried by *P. mendocina* strain S5.2, to which plasmid curing was observed to pose direct impact on the motility phenotype and methylome profile of the strain. Furthermore, the possible involvement of the plasmid-borne RM systems in regulating the expressions of motility and metal resistance genes was described in this study, suggesting the essentiality of pPME5 for fitness of the strain in native vineyard soil environments. In a nutshell, this work has demonstrated the adaptation of both *Pseudomonas* strains to the grapevine soils enriched with copper and other metals.
REFERENCES


nitrogen-contaminated water samples when cultivated with *Cyperus alternifolius* L. *Journal of Bioscience and Bioengineering, 114*(2), 182-187.


APPENDIX A

Growth of *P. mendocina* strain S5.2 and *P. putida* strain S13.1.2 in M9 medium supplemented with various concentrations of CuSO₄.
## APPENDIX B

Approximate concentration range of heavy metal salts in Biolog Microplates.

<table>
<thead>
<tr>
<th>Biolog Microplates</th>
<th>Heavy Metal Salt</th>
<th>Wells</th>
<th>Approximate concentration range (µg/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM13B</td>
<td>NiCl$_2$</td>
<td>A9-A12</td>
<td>10-1000</td>
</tr>
<tr>
<td></td>
<td>K$_2$CrO$_4$</td>
<td>C9-C12</td>
<td>4-400</td>
</tr>
<tr>
<td></td>
<td>CsCl</td>
<td>F1-F4</td>
<td>20-2000</td>
</tr>
<tr>
<td></td>
<td>CoCl$_2$</td>
<td>G1-G4</td>
<td>8-800</td>
</tr>
<tr>
<td></td>
<td>MnCl$_2$</td>
<td>G5-G8</td>
<td>80-8000</td>
</tr>
<tr>
<td></td>
<td>CuCl$_2$</td>
<td>H1-H4</td>
<td>5-500</td>
</tr>
<tr>
<td>PM14A</td>
<td>CdCl$_2$</td>
<td>D1-D4</td>
<td>4-400</td>
</tr>
<tr>
<td></td>
<td>Na$_3$AsO$_4$</td>
<td>B9-B12</td>
<td>20-2000</td>
</tr>
<tr>
<td>PM15B</td>
<td>ZnCl$_2$</td>
<td>H9-H12</td>
<td>2-200</td>
</tr>
<tr>
<td>PM16A</td>
<td>NaSeO$_3$</td>
<td>F5-F8</td>
<td>40-4000</td>
</tr>
<tr>
<td></td>
<td>CrCl$_3$</td>
<td>G1-G4</td>
<td>8-800</td>
</tr>
<tr>
<td>PM18C</td>
<td>NaAsO$_2$</td>
<td>D1-D4</td>
<td>10-1000</td>
</tr>
<tr>
<td></td>
<td>SbCl$_3$</td>
<td>E9-E12</td>
<td>5-500</td>
</tr>
</tbody>
</table>
APPENDIX C

Biolog Microplate™ maps for carbon sources
APPENDIX D

Biolog Microplate™ maps for chemicals and individual growths of strain S5.2 and S13.1.2 shown in area under curve (AUC) on each substrate.

*P. mendocina* strain S5.2

*P. putida* strain S13.1.2

---

**PM13B MicroPlate™**
APPENDIX D (Continued)

P. mendocina strain S5.2

P. putida strain S13.1.2
APPENDIX D (Continued)

P. mendocina strain S5.2

P. putida strain S13.1.2

PM15B MicroPlate™
APPENDIX D (Continued)

P. mendocina strain S5.2

P. putida strain S13.1.2
APPENDIX D (Continued)

P. mendocina strain S5.2

![Graph showing P. mendocina strain S5.2]

P. putida strain S13.1.2

![Graph showing P. putida strain S13.1.2]

PM18C MicroPlate™

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX E

HGAP assembly report for *P. mendocina* strain S5.2
APPENDIX E (continued)

HGAP assembly report for *P. mendocina* strain S5.2
APPENDIX E (continued)

HGAP assembly report for *P. mendocina* strain S5.2
APPENDIX E (continued)

HGAP assembly report for *P. mendocina* strain S5.2
APPENDIX E (continued)

HGAP assembly report for *P. mendocina* strain S5.2.

**Pre-Assembler Report**
- Polymerase Read Bases: 1,471,272,427
- Length Cutoff: 17,310
- Seed Bases: 156,043,637
- Pre-Assembled bases: 67,835,716
- Pre-Assembled Yield: 652
- Pre-Assembled Reads: 11,353
- Pre-Assembled Reads Length: 5,975
- Pre-Assembled N50: 8,988

**Polished Assembly**
- Polished Contigs: 2
- Max Contig Length: 5,119,935
- ND0 Contig Length: 5,119,935
- Sum of Contig Lengths: 5,372,264

**Contig Coverage Vs Confidence**

**Top Corrections**

**Consensus Calling Results**

<table>
<thead>
<tr>
<th>Reference</th>
<th>Reference Length</th>
<th>Bases Called</th>
<th>Consensus Concordance</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>scf1310000000008</td>
<td>5119819</td>
<td>100%</td>
<td>99.9970%</td>
<td>181.75</td>
</tr>
<tr>
<td>scf1800000000009</td>
<td>252318</td>
<td>100%</td>
<td>99.9968%</td>
<td>206.96</td>
</tr>
</tbody>
</table>

**Corrections Across Reference**

For Research Use Only. Not for use in diagnostic procedures.
APPENDIX F

HGAP assembly report for *P. putida* strain S13.1.2
APPENDIX F (continued)

HGAP assembly report for *P. putida* strain S13.1.2
APPENDIX F (Continued)

HGAP assembly report for *P. putida* strain S13.1.2

![Observed Insert Length Distribution](image)

### Loading

<table>
<thead>
<tr>
<th>SMRT Cell ID</th>
<th>Productive ZMWs</th>
<th>ZMW Loading For Productivity 0</th>
<th>ZMW Loading For Productivity 1</th>
<th>ZMW Loading For Productivity 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>n33001_065912_4216</td>
<td>150,298</td>
<td>65.30%</td>
<td>27.75%</td>
<td>6.95%</td>
</tr>
<tr>
<td>n33001_065914_4716</td>
<td>150,298</td>
<td>64.47%</td>
<td>30.73%</td>
<td>5.80%</td>
</tr>
<tr>
<td>n31118_04466_4216</td>
<td>150,292</td>
<td>66.94%</td>
<td>24.14%</td>
<td>8.52%</td>
</tr>
<tr>
<td>n31118_04466_4212</td>
<td>150,292</td>
<td>31.15%</td>
<td>32.27%</td>
<td>36.62%</td>
</tr>
<tr>
<td>n31118_04466_4216</td>
<td>150,292</td>
<td>35.05%</td>
<td>30.07%</td>
<td>34.33%</td>
</tr>
<tr>
<td>n33001_093822_4216</td>
<td>150,292</td>
<td>67.89%</td>
<td>23.51%</td>
<td>8.56%</td>
</tr>
<tr>
<td>n31118_04466_4216</td>
<td>150,292</td>
<td>71.89%</td>
<td>20.70%</td>
<td>7.39%</td>
</tr>
</tbody>
</table>

### Mapping

- Mapped Subread Length N50 (bp): 3,432
- Mapped Polymerase Read Length 95% (bp): 15,610
- Mapped Subread Length Mean (bp): 2,591
- Mapped Polymerase Read Length Max (bp): 44,817
APPENDIX F (Continued)

HGAP assembly report for *P. putida* strain S13.1.2
APPENDIX F (Continued)

HGAP assembly report for *P. putida* strain S13.1.2
APPENDIX G

List of putative open reading frames (ORFs) with annotable functions assigned to pPME5 plasmid of *P. mendocina* strain 5.2.

<table>
<thead>
<tr>
<th>Coding Sequence</th>
<th>Product</th>
<th>Start</th>
<th>Stop</th>
<th>Orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prokka00001</td>
<td>Site-specific tyrosine recombinase XerC</td>
<td>61</td>
<td>1179</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00002</td>
<td>ATP-dependent DNA helicase PcrA</td>
<td>8386</td>
<td>8877</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00006</td>
<td>Transglycosylase SLT domain protein</td>
<td>1683</td>
<td>3827</td>
<td>+</td>
</tr>
<tr>
<td>Prokka00011</td>
<td>DNA primase DnaG</td>
<td>11659</td>
<td>13005</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00018</td>
<td>Tyrosine recombinase XerC</td>
<td>17754</td>
<td>18740</td>
<td>+</td>
</tr>
<tr>
<td>Prokka00019</td>
<td>Uracil DNA glycosylase superfamily protein</td>
<td>18750</td>
<td>19379</td>
<td>+</td>
</tr>
<tr>
<td>Prokka00029</td>
<td>Bifunctional (p)ppGpp synthase/hydrolase relA</td>
<td>36469</td>
<td>37107</td>
<td>+</td>
</tr>
<tr>
<td>Prokka00037</td>
<td>Modification methylase PvuII</td>
<td>40848</td>
<td>42116</td>
<td>+</td>
</tr>
<tr>
<td>Prokka00040</td>
<td>Chromosome partition protein Smc</td>
<td>43889</td>
<td>44953</td>
<td>+</td>
</tr>
<tr>
<td>Prokka00042</td>
<td>RNA polymerase-binding transcription factor DksA</td>
<td>45291</td>
<td>45698</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00043</td>
<td>Bacterial membrane flanked domain protein</td>
<td>45695</td>
<td>46207</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00049</td>
<td>C-5 cytosine-specific DNA methylase</td>
<td>50220</td>
<td>52031</td>
<td>+</td>
</tr>
<tr>
<td>Prokka00050</td>
<td>DNA-binding protein HU-beta</td>
<td>52085</td>
<td>52357</td>
<td>+</td>
</tr>
<tr>
<td>Prokka00060</td>
<td>Sel1 repeat protein</td>
<td>56011</td>
<td>57108</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00062</td>
<td>Endonuclease/Exonuclease/phosphatase family protein</td>
<td>57758</td>
<td>58819</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00063</td>
<td>ATPase family associated with various cellular activities (AAA)</td>
<td>58873</td>
<td>60525</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00065</td>
<td>Antirestriction protein KlcA</td>
<td>61055</td>
<td>61489</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00067</td>
<td>LemA family protein</td>
<td>62750</td>
<td>63319</td>
<td>+</td>
</tr>
<tr>
<td>Prokka00074</td>
<td>nucleoside triphosphate pyrophosphohydrolase</td>
<td>65836</td>
<td>66555</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00079</td>
<td>RES domain protein</td>
<td>68293</td>
<td>68733</td>
<td>-</td>
</tr>
<tr>
<td>Accession</td>
<td>Description</td>
<td>Start</td>
<td>End</td>
<td>Type</td>
</tr>
<tr>
<td>------------</td>
<td>------------------------------------------------------------------------------</td>
<td>-------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>Prokka00082</td>
<td>CobQ/CobB/MinD/ParA nucleotide binding domain protein</td>
<td>69991</td>
<td>70917</td>
<td>+</td>
</tr>
<tr>
<td>Prokka00083</td>
<td>putative chromosome-partitioning protein</td>
<td>70914</td>
<td>72029</td>
<td>+</td>
</tr>
<tr>
<td>Prokka00087</td>
<td>Restriction endonuclease</td>
<td>73269</td>
<td>74102</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00095</td>
<td>Abi-like protein</td>
<td>77039</td>
<td>77770</td>
<td>+</td>
</tr>
<tr>
<td>Prokka00098</td>
<td>Flagellin N-methylase</td>
<td>78323</td>
<td>78934</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00105</td>
<td>BtpA family protein</td>
<td>84729</td>
<td>85397</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00125</td>
<td>putative DNA-binding transcriptional regulator</td>
<td>95210</td>
<td>96478</td>
<td>+</td>
</tr>
<tr>
<td>Prokka00135</td>
<td>SNF2 family N-terminal domain; product=hypothetical protein</td>
<td>101529</td>
<td>103307</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00145</td>
<td>tRNA(tet)-specific endonuclease VapC</td>
<td>107451</td>
<td>107867</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00146</td>
<td>Antitoxin VapB</td>
<td>107867</td>
<td>108097</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00157</td>
<td>Putative molybdenum carrier</td>
<td>112928</td>
<td>113506</td>
<td>+</td>
</tr>
<tr>
<td>Prokka00165</td>
<td>GIY-YIG catalytic domain protein</td>
<td>119801</td>
<td>120601</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00166</td>
<td>Nucleoid-associated protein YejK</td>
<td>120689</td>
<td>121696</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00196</td>
<td>DNA polymerase III PolC-type</td>
<td>134086</td>
<td>134871</td>
<td>+</td>
</tr>
<tr>
<td>Prokka00217</td>
<td>Arc-like DNA binding domain protein</td>
<td>143766</td>
<td>144323</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00236</td>
<td>TrfA protein</td>
<td>151125</td>
<td>152252</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00238</td>
<td>DNA polymerase V subunit UmuC</td>
<td>154052</td>
<td>155164</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00240</td>
<td>Mercuric reductase MerA</td>
<td>156013</td>
<td>157707</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00241</td>
<td>Mercuric resistance protein MerC</td>
<td>157728</td>
<td>158162</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00242</td>
<td>Mercuric transport protein periplasmic component precursor MerP</td>
<td>158175</td>
<td>158450</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00243</td>
<td>MerT mercuric transport protein</td>
<td>158464</td>
<td>158814</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00244</td>
<td>Mercuric resistance operon regulatory protein</td>
<td>158886</td>
<td>159296</td>
<td>+</td>
</tr>
<tr>
<td>Prokka00247</td>
<td>Phage integrase family protein</td>
<td>160285</td>
<td>162138</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00249</td>
<td>Site-specific tyrosine recombinase XerC</td>
<td>163746</td>
<td>164903</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00250</td>
<td>LexA repressor</td>
<td>165127</td>
<td>165561</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00285</td>
<td>Carbon storage regulator CsrA</td>
<td>184082</td>
<td>184273</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00286</td>
<td>NYN domain protein</td>
<td>184285</td>
<td>185034</td>
<td>-</td>
</tr>
<tr>
<td>Gene Name</td>
<td>Description</td>
<td>Start</td>
<td>End</td>
<td>Type</td>
</tr>
<tr>
<td>------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>----------</td>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>Prokka00287</td>
<td>SprT-like family protein</td>
<td>185031</td>
<td>186038</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00291</td>
<td>Adenosine monophosphate-protein transferase VbhT</td>
<td>188435</td>
<td>189343</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00298</td>
<td>Twitching mobility protein PilT</td>
<td>195657</td>
<td>196817</td>
<td>+</td>
</tr>
<tr>
<td>Prokka00300</td>
<td>Plasmid segregation protein ParM</td>
<td>197117</td>
<td>198196</td>
<td>+</td>
</tr>
<tr>
<td>Prokka00311</td>
<td>Disulfide isomerase/thiol-disulfide oxidase</td>
<td>211063</td>
<td>212133</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00313</td>
<td>Macrophage killing protein with similarity to conjugation protein</td>
<td>212481</td>
<td>213197</td>
<td>+</td>
</tr>
<tr>
<td>Prokka00322</td>
<td>ATP-dependent DNA helicase PerA</td>
<td>220033</td>
<td>222114</td>
<td>+</td>
</tr>
<tr>
<td>Prokka00323</td>
<td>AAA-like domain protein</td>
<td>222114</td>
<td>225242</td>
<td>+</td>
</tr>
<tr>
<td>Prokka00326</td>
<td>DNA replication terminus site-binding protein (Ter protein); product=hypothetical protein</td>
<td>225918</td>
<td>226739</td>
<td>-</td>
</tr>
<tr>
<td>Prokka00329</td>
<td>YqaJ-like viral recombinase domain protein</td>
<td>228113</td>
<td>229627</td>
<td>+</td>
</tr>
<tr>
<td>Prokka00331</td>
<td>N-6 DNA Methylase</td>
<td>231105</td>
<td>238388</td>
<td>+</td>
</tr>
<tr>
<td>Prokka00336</td>
<td>Bacterial type II and III secretion system protein</td>
<td>240298</td>
<td>242142</td>
<td>+</td>
</tr>
<tr>
<td>Prokka00339</td>
<td>putative major pilin subunit</td>
<td>243423</td>
<td>244025</td>
<td>+</td>
</tr>
<tr>
<td>Prokka00341</td>
<td>Type II secretion system protein E EpsE</td>
<td>244832</td>
<td>246754</td>
<td>+</td>
</tr>
<tr>
<td>Prokka00342</td>
<td>Type IV pilin biogenesis protein</td>
<td>246754</td>
<td>247878</td>
<td>+</td>
</tr>
<tr>
<td>Prokka00347</td>
<td>Transglycosylase SLT domain protein</td>
<td>250974</td>
<td>251432</td>
<td>+</td>
</tr>
</tbody>
</table>
APPENDIX H

Complete genome assembly of strain S5.2 revealed the average G+C content of plasmid pPME5 at 55.6%, which is lower a value as compared to the host chromosome DNA at 62.40%. This implies a possible event of horizontal gene transfer from phylogenetically distant bacteria in vineyard soil. However, comparison of codon usages between the chromosome and plasmid did not show significant differences as illustrated below.

Codon usage fraction compared between chromosome and plasmid pPME5 of P. mendocina S5.2.

Codon usage fraction compared between chromosome and plasmid pPME5 of P. mendocina S5.2.
APPENDIX I

Upper: Demonstration on increased coverage of TIR region on the left extremity of coverage plot generated from SMRT sequencing assembly (Gomez-Escribano et al., 2015). Lower: Absence of increased coverage in coverage plot of plasmid pPME5.

Coverage plot of pPME5 assembly
APPENDIX J

Base modification analysis report for *P. mendocina* strain S5.2 (PP).
APPENDIX K

Base modification analysis report for *P. mendocina* strain S5.2 after curing of pPME5 plasmid (NP).
APPENDIX L

Base modification analysis report for *P. putida* strain S13.1.2
APPENDIX M

Graphical representation of single-stranded methylation of GACG\textsuperscript{m6}AG motif and its proximity with multicopper oxidase \textit{cumA} gene.

Proximity of GACG\textsuperscript{m6}AG methylation sites with the predicted promoter region of multicopper oxidase \textit{cumA} gene.
APPENDIX N

Methylome and Restriction Modification (RM) Systems in *P. putida* S13.1.2

Methylome analysis conducted on strain S13.1.2 indicated the presence of GTTC\(m^4\)CG methylation motif recognized by m4C MTases (Table 4.15). On the other hand, REBASE has identified a total of six MTases (Table 4.16) however none was associated with the recognition motif detected from SMRT sequencing. In addition, the recognition sequence of all predicted components of RM system in this strain remained undetermined.

Summary on genome-wide DNA methylation motifs in *P. putida* strain S13.1.2.

<table>
<thead>
<tr>
<th>Recognition Motif</th>
<th>Modified Position</th>
<th>Modification Type</th>
<th>No. of Motifs in Genome</th>
<th>No. of Modified Motifs Detected</th>
</tr>
</thead>
<tbody>
<tr>
<td>GTTC(\tilde{C})G</td>
<td>5</td>
<td>4mC</td>
<td>2,137</td>
<td>1,553</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>% Modified Motifs Detected</th>
<th>Mean Modification QV</th>
<th>Mean Motif Coverage</th>
<th>Partner Motif</th>
</tr>
</thead>
<tbody>
<tr>
<td>72.67%</td>
<td>72.7</td>
<td>71.59</td>
<td>-</td>
</tr>
</tbody>
</table>

- Modified position was denoted by underlined base.
APPENDIX N (Continued)

Methylome and Restriction Modification (RM) Systems in *P. putida* S13.1.2

**Putative restriction modification systems in *P. putida* strain S13.1.2.** None of the predicted RM systems were found to modify the GTTC\textsuperscript{m4}CG motifs detected from SMRT sequencing.

<table>
<thead>
<tr>
<th>Type</th>
<th>Gene</th>
<th>Name</th>
<th>Predicted Recognition Sequence</th>
<th>Genome Coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>II</td>
<td>M</td>
<td>M.PpuS1312ORF3785P</td>
<td>unknown</td>
<td>864451-865374</td>
</tr>
<tr>
<td>II</td>
<td>M</td>
<td>M.PpuS1312ORF6215P</td>
<td>unknown</td>
<td>1377321-1379066</td>
</tr>
<tr>
<td>II</td>
<td>M</td>
<td>M.PpuS1312ORF13915P</td>
<td>unknown</td>
<td>3144817-3145767</td>
</tr>
<tr>
<td>II</td>
<td>R</td>
<td>PpuS1312ORF20340P</td>
<td>unknown</td>
<td>4579637-4581832</td>
</tr>
<tr>
<td>II</td>
<td>M</td>
<td>M.PpuS1312ORF20340P</td>
<td>unknown</td>
<td>4581881-4582870</td>
</tr>
<tr>
<td>II</td>
<td>V</td>
<td>V.PpuS1312ORF20340P</td>
<td>unknown</td>
<td>4584135-4584545</td>
</tr>
<tr>
<td>II</td>
<td>R</td>
<td>PpuS1312ORF20355P</td>
<td>unknown</td>
<td>4582943-4584133</td>
</tr>
<tr>
<td>II</td>
<td>RM</td>
<td>PpuS1312ORF25705P</td>
<td>unknown</td>
<td>5813180-5817190</td>
</tr>
</tbody>
</table>

**Legend:**
- **R:** Restriction subunit
- **M:** Modification subunit
- **S:** Specificity subunit
- **V:** Nicking enzyme

Distribution map of GTTC\textsuperscript{m4}CG modification motifs in *P. putida* strain S13.1.2 complete chromosome.
LIST OF PUBLICATIONS AND PAPERS PRESENTED


