Comparative genome analysis reveals genetic adaptation to versatile environmental conditions and importance of biofilm lifestyle in *Comamonas testosteroni*

Wu, Yichao; Arumugam, Krithika; Tay, Martin Qi Xiang; Seshan, Hari; Mohanty, Anee; Cao, Bin

2015

Wu, Y., Arumugam, K., Tay, M. Q. X., Seshan, H., Mohanty, A., & Cao, B. (2015). Comparative genome analysis reveals genetic adaptation to versatile environmental conditions and importance of biofilm lifestyle in *Comamonas testosteroni*. *Applied Microbiology and Biotechnology*, 99(8), 3519-3532.

<https://hdl.handle.net/10356/106190>

<https://doi.org/10.1007/s00253-015-6519-z>

© 2015 Springer-Verlag Berlin Heidelberg. This is the author created version of a work that has been peer reviewed and accepted for publication by *Applied Microbiology and Biotechnology*, Springer-Verlag Berlin Heidelberg. It incorporates referee's comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: [<http://dx.doi.org/10.1007/s00253-015-6519-z>].

**Comparative genome analysis reveals genetic adaptation to versatile environmental
conditions and importance of biofilm lifestyle in *Comamonas testosteroni***

Yichao Wu^{a,b}, Krithika Arumugam^b, Martin Qi Xiang Tay^b, Hari Seshan^{b,c}, Anee Mohanty^{a,b},
Bin Cao^{a,b*}

^aSchool of Civil and Environmental Engineering, Nanyang Technological University,
Singapore 639798, e-mail: bincao@ntu.edu.sg

^bSingapore Centre on Environmental Life Sciences Engineering, Nanyang Technological
University, Singapore

^cDepartment of Civil and Environmental engineering, University of California, Davis, USA

*Corresponding author.

Abstract

Comamonas testosteroni is an important environmental bacterium capable of degrading a variety of toxic aromatic pollutants and has been demonstrated to be a promising biocatalyst for environmental decontamination. This organism is often found to be among the primary surface-colonizers in various natural and engineered ecosystems, suggesting an extraordinary capability of this organism in environmental adaptation and biofilm formation. The goal of this study was to gain genetic insights into the adaption of *C. testosteroni* to versatile environments and the importance of a biofilm lifestyle. Specifically, a draft genome of *C. testosteroni* I2 was obtained. The draft genome is 5,778,710 bp in length and comprises 110 contigs. The average G+C content was 61.88%. A total of 5365 genes with 5263 protein-coding genes were predicted, whereas 4324 (80.60% of total genes) protein-encoding genes were associated with predicted functions. The catabolic genes responsible for biodegradation of steroid and other aromatic compounds on draft genome were identified. Plasmid pI2 was found to encode a complete pathway for aniline degradation and a partial catabolic pathway for chloroaniline. This organism was found to be equipped with a sophisticated signalling system which help it find ideal niches and switch between planktonic and biofilm lifestyles. A large number of putative multidrug resistant genes coding for abundant outer membrane transporters, chaperones and heat shock proteins for the protection of cellular function were identified in the genome of strain I2. In addition, the genome of strain I2 was predicted to encode several proteins involved in producing, secreting and uptaking siderophores under iron-limiting conditions. The genome of strain I2 contains a number of genes responsible for the synthesis and secretion of exopolysaccharides, an extracellular component essential for biofilm formation. Overall, our results reveal the genomic features underlying the adaption of *C. testosteroni* to versatile environments and highlighting the importance of its biofilm lifestyle.

Keywords: *Comamonas*, genome sequence, biodegradation, biofilm

Introduction

Comamonas spp. are Gram-negative, aerobic, non-pigmented, and rod-shaped bacteria which are widely diversified in different environments, including soil, sediment, and water (Willems and De Vos 2006). They are of particular research interest because they have been implicated to play an important role in wastewater treatment processes (Bramucci and Nagarajan 2000; Gumaelius et al. 2001; Shokrollahzadeh et al. 2008), where they are often found to be the primary surface-colonizers (Andersson et al. 2008; Juang et al. 2010; Kim et al. 2014; Luo et al. 2012). Among them, *Comamonas testosteroni* is capable of mineralizing complex xenobiotic compounds and has been reported to be a promising organism for bioremediation of recalcitrant organics-contaminated environments (Harayama 1997; Sylvestre 1995; Wu et al. 2015).

Genome sequences of several *Comamonas* strains have provided novel insights into their metabolic capabilities. For example, the chromosomes of *C. testosteroni* strains ATCC11996, S44, and CNB-2 encode catabolic enzymes for biodegradation of aromatic compounds while lack key genes responsible for the assimilation of carbohydrates (Gong et al. 2012; Ma et al. 2009; Xiong et al. 2011). In addition, a high proportion of genes for transport (22% of total genes) and signalling regulation (6% of total genes) were found in the complete genomes of *C. testosteroni* (Ma et al. 2009), ensuring survival in a changing environment. *C. testosteroni* strains were also associated with heavy-metal resistance, as several operons have been found to encode heavy metal transporters (Ma et al. 2009; Xiong et al. 2011). Moreover, besides diverse genes localized on chromosome, catabolic plasmids such as IncP-1 **plasmids**, often play an important role in biodegradation of organic compounds (Boon et al. 2001; Król et al. 2012; Schlüter et al. 2007). For example, plasmid pCNB1 of *C. testosteroni* **CNB-1** encodes a metabolic pathway to convert 4-chloronitrobenzene (4-CNB) to 5-chloro-4-hydroxy-2-oxoaleate, without which this strain could not grow on 4-CNB (Ma et al. 2007; Wu et al. 2006). Another IncP-1 plasmid from *C. testosteroni* strain WDL7 capable of using 3-chloroaniline (3-CA) as sole carbon and nitrogen source carries a putative 3-CA oxidation gene cluster, *dcaQTA1A2BR* (Król et al. 2012).

C. testosteroni strain I2 was isolated from activated sludge as the first reported *C. testosteroni* strain capable of using both 3-CA and aniline as the sole carbon and nitrogen source (Boon et al. 2001). *C. testosteroni* I2 has been used in bioaugmentation and achieved complete 3-chloroaniline removal (Boon et al. 2000). The bioreactors augmented with I2 exhibited a higher community stability and a faster recovery of nitrification against 3-chloroaniline shock (Boon et al. 2003). *C. testosteroni* I2 also harbors one large IncP-1 plasmid pI2 of ~100 kb. The plasmid pI2 has been shown to contain a *tdnQ*-like oxidative deamination gene for aniline oxidation (Boon et al. 2001; Król et al. 2012). Further study revealed that plasmid pI2 was responsible for complete degradation of aniline but not 3-CA (Boon et al. 2001). However, in bioaugmentation, after introducing this plasmid into the reactors through a pI2-transconjugant, continuous 3-CA removal was achieved and a number of 3-CA degrading *C. testosteroni* strains harbouring pI2 were isolated afterwards (Bathe et al. 2005). Therefore, the genomic information of the genome and plasmid pI2 and their roles in degradation still needed to be determined.

Meanwhile, the colonization of substratum and formation of aggregates in sludge by *Comamonas* cells implies the importance of the biofilm lifestyle. In contrast to planktonic, freely suspended cells, cells in a biofilm are embedded in a matrix of self-produced extracellular polymeric substances (EPS) (Ding et al. 2014; Flemming and Wingender 2010). As compared to planktonic cells, cells in a biofilm exhibit distinct properties such as a reduced growth rate and an increased tolerance to harsh conditions (Booth et al. 2011). Hence, biofilms are very promising whole-cell catalysts in effectively decontaminating toxic pollutants. However, genomic features relevant to biofilm lifestyle for *C. testosteroni* have not elucidated.

Hence, the goal of this study was to sequence and comparatively analyse the genome of *C. testosteroni* I2 with reference to that of other *Comamonas* strains to obtain insights into its metabolic potentials, adaptive capabilities and regulatory mechanisms. Genomic features related to biofilm lifestyle were analysed.

Material and methods

Genome sequencing and assembly

C. testosteroni strain I2 (LMG19554 from BCCM/LMG) was grown in lysogeny broth (LB) medium at 30°C. Genomic DNA was subsequently isolated by using Qiagen Genomic-Kit (Qiagen K.K., Tokyo, Japan). Sequencing was performed using the Illumina MiSeq technology (Illumina, CA, USA) with 250-bp paired-end sequences (7,171,708 reads). The resulting sequences were quality trimmed and filtered to remove adapter sequences by using cutadapt 1.2.1 (Martin 2011). Filtered reads were assembled using Velvet 1.2.08 (k = 153) (Zerbino and Birney 2008).

Sequence annotation and analysis

The contigs were annotated at the Joint Genome Institute (JGI) Integrated Microbial Genomes Expert Review (IMG ER) website (<https://img.jgi.doe.gov/cgi-bin/mer/main.cgi>) (Hyatt et al. 2010; Markowitz et al. 2009). The genome sequence and additional information, including protein families (COGs, Pfams, TIGRfams) and functional categories (COG KEGG pathway) were available on IMG database (analysis project ID: Ga0022151) (Markowitz et al. 2014). Regions containing prophages and transposases were predicted using Phast and ISSaga (Varani et al. 2011; Zhou et al. 2011). Circular map of the chromosome was drawn using the Circos software (version 0.64) (Krzywinski et al. 2009).

Comparative sequence analysis

The draft genome sequence of strain I2 was compared with genome of *C. testosteroni* ATCC 11996 (accession no. NZ_AHIL000000000), CNB-2 (accession no. NC_013446.2), S44 (accession no. NZ_ADVQ000000000), KF-1 (accession no. NZ_AAUI000000000.2), NBRC 100989 (accession no. NZ_BAEC000000000), *Comamonas* sp. B-9 (accession no. NZ_APCR000000000), *C. composti* DSM 21721 (accession no. AUCQ000000000), and *C. badia* DSM 17552 (accession no. AXVM000000000). *In silico* subtractive hybridization analysis was done for strain I2 against eight other *Comamonas* strains to identify strain specific and conserved genes by using mGenomeSubtractor (Shao et al. 2010). Genes with a BLASTP-derived homology value > 0.81 were considered as conserved, while those

with a homology value < 0.42 were nominated as strain-specific. The rest were designated as unclassified. Plasmid sequence comparison was performed with BRIG (BLAST Ring Image Generator) (Alikhan et al. 2011). The plasmid sequence of pI2 (accession no. NC_016978) was used as the reference sequence and compared with pWDL7::rfp (accession no. GQ495894), pTB30 (accession no. NC_016968) and pCNB (accession no. NC_010935). All genetic sequences were retrieved from the NCBI database (<https://www.ncbi.nlm.nih.gov>). The 16S rRNA phylogenetic tree was constructed by MEGA 6.0 using neighbor joining method (Tamura et al. 2013).

Detection of siderophore production by chrome azurol S assay (CAS)

Siderophore production by *C. testosteroni* strains I2 and WDL7 was assessed by liquid CAS assay (Schwyn and Neilands 1987). Culture was growing in minimal M9 medium supplemented with 30 μ M FeCl₃ for iron-rich conditions or deferrated with 30 μ M 2, 2'-bipyridine for iron-deficient conditions. Siderophore-producing *Pseudomonas aeruginosa* PAO1 (ATCC 15692) was used as a positive control (Mohanty et al. 2015; Wang et al. 2013). Filtered supernatant was mixed with equal volume of CAS assay solution. Siderophore production was detected by a visual change in color from blue to orange.

Biofilm assay

Biofilm assay was conducted based on a 96-well plate method (Wakimoto et al. 2004; Wu et al. 2015). Two μ L of overnight LB culture (OD₆₀₀ ~1.4) was inoculated into 200 μ L of minimal M9 medium containing 10 mM pyruvate as carbon source. After 12 h incubation at 30°C, planktonic cells were removed and 1% crystal violet was used to stain surface-associated biomass for 15 min. Then 200 μ L of 95% ethanol was added to disperse the surface-associated biomass and the amount of biofilm biomass was quantified by measuring optical density at 590 nm (OD₅₉₀). Medium without inoculation was used as negative control.

Results

Genome features of *C. testosteroni*

The draft genome of strain I2 was 5,778,710 bp in length and comprised 110 contigs, with a N50 of 325,482 bp and a maximum length of 705,473 bp. The average G+C content was 61.88% (Table 1 and Fig. 1). A total of 5365 genes with 5263 protein-coding genes were predicted, whereas 4324 (80.60% of total genes) protein-encoding genes were associated with predicted functions. Approximately, 1405 (26.19% of total genes) and 1214 (22.63% of total genes) protein-coding genes were connected to KEGG and MetaCyc pathways, respectively. Comparing with other *Comamonas* strains which have been sequenced, all *C. testosteroni* strains shared a similar genome size and GC content (Table 1).

According to homology analysis, more than 70% of genes in strain I2 were identified as conserved when compared with other 5 *C. testosteroni* strains (Fig. 2). *C. testosteroni* KF-1 and NBRC 100989 showed highest similarity with strain I2, with 88.4% and 86.9% of genes in strain I2 having a homology with H-value > 0.81 in KF-1 and NBRC 100989, respectively. It was in a good agreement with the phylogenetic analysis which showed that strain I2 was closely related to *C. testosteroni* NBRC 100989 from soil and shared about 97% of homology with four other *C. testosteroni* strains (Fig. S1 in the Supplementary Material). Meanwhile, more genes in strain I2 were revealed to be strain-specific in the comparison with other three *Comamonas* spp. For the rest of the three distantly related strains, approximately 50% of genes were identified as strain-specific which revealed a divergent genomic profile. In addition, there were only 3.6 % of genes in the genome of I2 which were conserved among all the 9 *Comamonas* spp., while 4.2 % of genes were strain specific. The conserved genes are mainly involved in steroid and aromatic degradation as well as heavy-metal resistance (Gong et al. 2012; Ma et al. 2009; Xiong et al. 2011).

Basic nutritional metabolism

Members of the *Comamonas* genus are considered as aerobes which are able to grow on organic acid, amino acid and various complex aromatics, but with poor ability to assimilate sugar (Ma et al. 2009; Willems and De Vos 2006). Based on COG functional categories (Fig. 3), all *Comamonas* strains had

a similar distribution of genes classified into different functional groups. Even three distantly related species, strains DSM 17552, DSM 21721 and *Comamonas* sp. B-9 shared similar general features with only slight differences (Table 1 and Fig. 3). The majority of genes of *Comamonas* strains were associated with amino acid metabolism (a), energy production and conversion (i), transcription (u), inorganic ion metabolism (l), lipid metabolism (n) and cell envelope biogenesis (e) of which the proportions were higher than 5%. The high number of proteins in these classes indicated that I2 possessed a delicate regulation system as well as a requirement for sufficient organic and inorganic nutrient in its lifestyle.

In the wastewater treatment process, *Comamonas* strains are usually found to play a role in denitrification (Lu et al. 2014). In the genome of *C. testosteroni*, there are two types of denitrification clusters *narGHIJ* and *napAB* which are able to reduce nitrate to nitrite. *C. testosteroni* I2 could reduce nitrate to nitrite under microaerobic and anaerobic conditions (data not shown). Comparison of different functional categories with other model bacteria (*Escherichia coli* K12 (Blattner et al. 1997), *Pseudomonas putida* KT2440 (Nelson et al. 2002), *Shewanella oneidensis* MR-1 (Heidelberg et al. 2002; Ng et al. 2013)) revealed remarkable differences in the distribution of functional categories of predicted proteins (Fig. 3). *Comamonas* had the highest proportion of genes devoted to lipid metabolism (n), which was even higher than that of *P. putida* KT2440, an important environmental bacterium involved in biodegradation. Strain I2 also had an increased proportion of genes classified as inorganic metabolism (l) and secondary metabolites biosynthesis, transport and catabolism (s). Meanwhile, for amino acid (a) and carbohydrate (b) metabolism, a smaller ratio of genes in *Comamonas* was found compared with *E. coli* K12 and *P. putida* KT2440. The proportion of genes in *Comamonas* for these two COGs families was comparable with that of *S. oneidensis* MR-1, which lacks key enzymes for glucose metabolism (Serres and Riley 2006). The distinctive percentage of genes for various metabolisms indicated that *Comamonas* spp. had sophisticated systems to uptake and metabolize lipid, inorganic nutrient and secondary metabolites while having relative poor ability to assimilate carbohydrates and amino acids. This provides clues to different roles of *C. testosteroni* strains in various natural environments.

Mobile genetic elements

Obtaining mobile genetic elements through horizontal gene transfer is essential for microbes to gain an advantage and avoid danger in their environments (Frost et al. 2005). A large number of mobile genetic elements were identified in the genome sequence of strain I2. A total of 126 genes (2.3% of total genes) were predicted to be associated with 6 putative prophages and 76 genes (1.4% of total genes) were predicted as transposases (Table 1 and S1 in the Supplementary Material). Most of the transposases were from the IS3 and IS5 families. As compared to other *Comamonas* strains (Table 1), all *C. testosteroni* strains contained more prophage genes than other distantly related species. Comparing with the mobile genetic elements (MGEs) in the genomic sequence of strain CNB-1 which contained 221 genes (4.6% of total genes) as prophages and 67 genes (1.4% of total genes) as transposases, most of the transposases belong to IS21 and IS4 instead of IS3 and IS5 (Ma et al. 2009). The difference in mobile genetic elements between strain I2 and strain CNB-1 indicates a different evolutionary result caused by different niches (Frost et al. 2005). Remarkably, there are 7 genes identified as transposases which are also found in prophages (I2contig_1463, I2contig_4900, I2contig_4901, I2contig_4907, I2contig_4908, I2contig_4911 and I2contig_4925), which suggests that these transposons may be acquired through transduction. This type of hybrid horizontal gene transfer was also found in the genome of strain CNB-1 (Ma et al. 2009).

Biodegradation of aromatic compounds

C. testosteroni strains were named for their ability to degrade testosterone. Previous reports have suggested that these strains play an important role in removing aromatic compounds from the environment (Möbus et al. 1997; Watanabe et al. 2002). In previous studies, *C. testosteroni* I2 was found to be able to use aniline and 3-chloroaniline as the sole carbon and nitrogen source (Boon et al. 2000). According to sequence analysis, there were 39, 13 and 52 genes coding for putative dioxygenase, hydroxylase and oxidoreductases, respectively. For polycyclic hydrocarbons, 9 key genes responsible for the conversion of testosterone to DOHNAA, pyruvate and propanal were identified (I2contig_0646, I2contig_2299 to I2contig_2307 and I2contig_2350). These putative

catabolic genes could also participate in the catabolic transformation of steroid hormones, such as epiandrosterone and dehydroepiandrosterone (Fig. 4).

For single phenyl ring compounds including aniline and 3-CA, two different gene clusters were identified in the genome for the degradation of phenol, phenylboronic acid, aminobenzoate and other aromatic compounds. Six genes (I2contig_153 to I2contig_158) code for protocatechuate 4, 5-dioxygenase, hydroxymuconate semialdehyde dehydrogenase and other key enzymes which convert hydroxybenzoate to oxaloacetate and pyruvate (Fig. 4). Another metabolic pathway was found responsible for conversion of phenylboronic acid, phenol, aniline and chloroaniline to pyruvate, acetaldehyde, acetyl-CoA and 3-oxoadipate. However, the ring-cleavage enzyme in the pathway was not identified. A recent study characterized the protocatechuate 4, 5-dioxygenase from *C. testosteroni* CNB-1 which shared 96-99% identities with the one encoded by strain I2 (Ni et al. 2013b). This enzyme was demonstrated to be specific for 3-hydroxybenzoate, 4-hydroxybenzoate and vanillate, instead of phenol, benzoate and gentisate.

In addition to the catabolic genes encoded by the chromosomes, the large plasmids in *C. testosteroni* strains also play a significant role in biodegradation. Strain CNB-1 could not grow on 4-chloronitrobenzene in the absence of the plasmid pCNB1 (Wu et al. 2006). Furthermore, the IncP-1 β plasmid pI2 harbored by strain I2 was found to contain *tdnQ*-like catabolic genes for aniline degradation (Boon et al. 2001). Through sequence analysis, one gene cluster containing 14 catabolic genes for complete aniline degradation was identified. The first 6 genes (*dcaQTA1A2BR*) encode a L-glutamate based oxidative deamination pathway which transforms aniline to catechol (Takeo et al. 2013). The rest of the eight genes in the cluster encode the lower pathway featured by a 2, 3-dioxygenase and other enzymes capable of converting 2-hydroxy-muconatesemialdehyde into TCA intermediates (Fig. 5).

Comparing with plasmids from other strains such as pCNB-1 from CNB-1, pTB30 from TB30, and pWDL7::*rfp* from WDL7, high similarities were found between the six upper pathway catabolic genes in pI2 and the corresponding catabolic genes in pTB30 and pWDL7::*rfp* (Fig. 5). The

catabolic pathway encoded in pCNB-1 showed relative lower similarities with that of pI2. In a previous study, catabolic genes in pI2 could not be induced by 3-CA (Boon et al. 2001). However, in a later bioaugmentation experiment, pI2 endowed reactors with the capability to degrade 3-CA completely (Bathe et al. 2005). At the genomic scale, the upper pathway enzymes of pWDL7::rfp were found to be able to convert 3-CA to 4-chlorocatechol (4-CC) (Król et al. 2012). As the upper pathway enzymes from these two plasmids share a high similarity, pI2 may also be able to convert 3-CA to 4-CC. Moreover, during the degradation of aniline and chloroaniline, a yellow intermediate was observed and then gradually disappeared in the culture, implying the presence of a typical meta-ring cleavage product 2-hydroxymuconate semialdehyde and/or its chlorinated derivative (Boon et al. 2000; Cao and Loh 2008). Based on the current draft genome sequence and complete plasmid sequence, the only ring-cleavage enzyme for catechol in *C. testosteroni* I2 was 2, 3-dioxygenase on the plasmid pI2. Therefore, plasmid pI2 most likely encodes for an enzyme for partial chloroaniline metabolism which may not be directly induced by 3-CA.

Signaling transduction and regulation

According to COG analysis, 197 ORFs were identified as signal transduction genes. Further analysis found that strain I2 was rich in response regulators with a CheY-like receiver domain (28 ORFs, 0.52% of total genes), signal transduction histidine kinase (25 ORFs, 0.47% of total genes) and methyl-accepting chemotaxis protein (23 ORFs, 0.43% of total genes). The number of signaling genes was higher than those in *E. coli* K12, which contained 14 ORFs for CheY-like receiver domains, 17 ORFs for signal transduction histidine kinases, and 5 methyl-accepting chemotaxis proteins. From the genes assigned to signaling transduction, 21 paired two-component signal transduction systems were identified based on KEGG database (Table S2 in the Supplementary Material). Two chemotaxis family systems were identified to control twitching motility, flagella assembly and motor switch adaptation. Furthermore, four heavy metal sensing systems were also found which was in agreement with the predicted heavy-metal resistance.

Chemotaxis is vital for microorganisms to sense and move towards nutrients in their natural environments. In a previous study on the chemotactic responses of *C. testosteroni* CNB-1, a methyl-accepting chemotaxis sensory transducer (CtCNB1_2201) was found to be involved in chemotaxis towards 10 different aromatic compounds by sensing their relative TCA intermediates (Ni et al. 2013a). A similar chemotaxis sensory transducer (I2contig_05077, 96% identify) was identified in strain I2. This type of metabolite-dependent chemotaxis sensory transducer increases the competence in evolution because the strain gains the chemotactic responses towards specific compounds only when it obtains the corresponding degradation capability (Pandey and Jain 2002).

C-di-GMP controlling system

In addition to the motile lifestyle which bacteria adopt to search for nutrients and optimal environments, biofilm is a predominant lifestyle for microorganisms in most natural and engineered environments (Singh et al. 2006). The formation of biofilms provides a range of advantages to the cells, especially under unfavorable conditions such as exposure to toxic metals, high salinity and antimicrobial agents (Cao et al. 2012; Hall-Stoodley et al. 2004; Singh et al. 2006). One of the most significant molecular determinants in the biofilm regulation mechanism is a signaling molecule bis-(3' -5')-cyclic dimeric guanosine monophosphate (c-di-GMP) (Hengge 2009). C-di-GMP is synthesized by GGDEF-domain containing diguanylate cyclases (DGCs) and degraded by EAL-domain containing phosphodiesterases (PDEs) (Simm et al. 2004). Strain I2 contains 55 putative proteins with roles in c-di-GMP regulation, including 30 DGCs, 3 PDEs and 3 hybrid EAL-HD-GYP domain proteins and 15 hybrid DGC-PDE proteins (Fig. 6). Five of the *C. testosteroni* strains also contain a comparable number and similar distribution of GGDEF, HD-GYP and EAL domain-containing proteins, while the other three *Comamonas* species only had about half the number of regulatory enzymes. Among the 9 *Comamonas* strains analyzed in this study, *Comamonas* spp. strain B-9 was the only strain without genes encoding HD-GYP domain proteins. When compared with other model organisms, only 27 and 32 putative proteins with GGDEF, HD-GYP and EAL domains were predicted in the genome of *E. coli* K-12 and *P. putida* KT2440, respectively. The number of c-

di-GMP regulatory enzymes predicted in *C. testosteroni* strains was comparable with that reported in *S. oneidensis* MR-1, an environmental organism with the highest number of c-di-GMP-controlling proteins after normalization (Fredrickson et al. 2008). The genome of MR-1 encodes more EAL and HD-GYP domain proteins and fewer hybrid EAL-HD-GYP domain proteins than the genomes of *C. testosteroni* strains. The presence of multiple genes involved in c-di-GMP signaling suggests that *C. testosteroni* has a sophisticated regulatory system to switch between planktonic and biofilm lifestyles.

Interestingly, in strain I2, a total of 26 c-di-GMP controlling proteins were predicted to be multi-domain fusion enzymes with a sensor domain instead of CheY-like receiving domain. Among these multi-domain proteins, 19 of them contain a cytoplasmic PAS or GAF sensor domain, while 4 of them were found in a combination with integral membrane sensory domain MHYT and MASE with three coupling to the extracellular domain Cache (Table 2). Strains of *C. testosteroni* have approximately 25 genes containing a sensor domain, which was higher than the other *Comamonas* species and other model environmental organisms (Table 2 and Fig. 6).

Transporter systems - heavy metal resistance

A previous study has shown that strain I2 exhibited resistance to mercury and rifampin (Boon et al. 2001). According to the phylogenetic tree and genome comparison (Fig. 2, 3 and S1 in the Supplementary Material), strain I2 showed a high similarity with *C. testosteroni* S44, which was reported to be resistant to multiple heavy metals (Xiong et al. 2011). In the COG database, strain I2 was predicted to have 26 outer membrane proteins (Table S3) while the number for other environmental microorganisms such as *P. putida* KT2440 and *S. oneidensis* MR-1 was 19 and 6, respectively. Most of these putative outer membrane proteins were predicted to involve in multidrug resistance, suggesting that strain I2 is potentially resistant to biocides and heavy metals.

In the draft genome sequence of strain I2, a number of gene clusters were predicted to encode the transport system for drug and heavy metal efflux. For example, 3 gene clusters encode an arsenate efflux system (I2contig_2189 to I2contig_2191, I2contig_3556 to I2contig_3558 and I2contig_4694

to I2contig_4696). In addition to the arsenate specific efflux system, there are 34 putative gene clusters coding multidrug transport system dispersed in the *C. testosteroni* I2 genome (Table S4 in the Supplementary Material). Seventeen of them were predicted to be ABC-type multidrug transport systems, including 2 clusters for the export of organic solvents. The rest 17 clusters belong to the resistance-nodulation-cell division (RND) family transport system. Furthermore, strain I2 was also found to have 39 genes encoding chaperones and heat shock proteins (Table S5 in the Supplementary Material) for protection of cellular function when exposed to toxins and heavy metals.

The IncP-1 plasmids contribute to the bacterial adaptation capability because of the conserved backbone and adaptive genetic structure (Schlüter et al. 2007). Plasmid pCNB was reported to contain arsenate and chromate resistance genes (Ma et al. 2007). Transconjugant harboring plasmid pI2 was found to acquire exogenous mercury resistance (Boon et al. 2001). Here in this study, we identified one mercury resistance operon including 6 genes located on plasmid pI2 (*merTPADE*). The presence of various heavy-metal resistance mechanisms predicted by genomic features suggests *C. testosteroni* strain I2 being a multi-heavy-metal resistant bacterium.

Siderophore production

Iron is an indispensable nutrient for almost all living organisms because of its essentiality in metabolic and signaling functions (Wandersman and Delepelaire 2004). At low iron levels, siderophores are secreted by bacteria to scavenge iron ions from natural environments (Neilands 1995). Previous studies showed that several *Comamonas* strains isolated from agricultural soil and root surfaces were capable of producing siderophores (Das et al. 2014; Elo et al. 2000). However, no significant siderophore production could be observed for *C. testosteroni* (Mampel et al. 2004; Patel et al. 2007) although genetic analysis identified one inner membrane transport system for siderophores in the genome of *C. testosteroni* CNB-1 (Ma et al. 2009).

In this study, our genomic analysis revealed that the draft genome sequence of strain I2 contains 24 genes encoding siderophore and ferrienterochelin receptors, 14 genes encoding

siderophore transport systems and 3 genes for TonB proteins (Table S6 in the Supplementary Material). Among the receptor genes, 21 of them were TonB-dependent receptors. The CAS assay indicated the secretion of siderophore by strain I2 (Fig. S2 in the Supplementary Material), confirming the function of the single siderophore biosynthesis gene identified in the genome.

Biosynthesis of extracellular polysaccharides

In a previous bioaugmentation study, strain I2 was found to form aggregates in sludge bioreactors (Boon et al. 2000). It also formed mucoid colonies on agar plate, indicating exopolysaccharide secretion (Bossier and Verstraete 1996). In the draft genome sequence of strain I2, 16 genes related to extracellular polysaccharides synthesis were identified: 3 capsular regulation genes, 2 capsular polysaccharide export clusters, 1 capsular polysaccharide biosynthesis gene, 3 exopolysaccharide biosynthesis genes, 2 lipopolysaccharide biosynthesis and 1 lipopolysaccharide transport gene (Table S7 in the Supplementary Material).

Discussion

C. testosteroni strain I2 is capable of degrading a variety of toxic aromatic compounds and has been demonstrated to be promising in environmental biotechnological applications (Willems and De Vos 2006). Genomic analysis of this strain revealed the presence of a number of mobile genetic elements (MGEs) and multidrug efflux pumps genes in its genome. Lateral gene transfer has a significant impact on the bacterial adaption to the environment (Ochman et al. 2000). Strain I2 acquires better adaption abilities through MGEs. For example, in one of prophages, one heavy metal response regulator (I2contig_4904) and one heavy metal sensor kinase gene (I2contig_4905) were identified. This constitutes a putative two component system to render this strain more competent in an environment containing heavy metals. A total of 34 putative multidrug transport systems were found in the genome of strain I2 for detoxifying environmental pollutants. In the previous genomic analysis, 4 ABC-type and 5 RND-driven transport systems were found in *C. testosteroni* S44, while 11 RND-type efflux transporters were identified in *C. testosteroni* CNB-2 (Ma et al. 2009; Xiong et al. 2011).

A higher number of putative multidrug transport systems identified in strain I2 is likely to confer this strain a resistance to multiple heavy metals.

Through comparative genomic analysis, sophisticated signaling systems were identified in *C. testosteroni*. Strains of *C. testosteroni* have been isolated from many different environmental niches including soils, water and activated sludge. In some clinical cases, *C. testosteroni* was even isolated from abdominal cavity and blood (Abraham and Simon 2007). In the microbial signal transduction (MiST) database, *C. testosteroni* strains ATCC 11996, CNB-2, KF-1 and S44 comprise about 380 one-component systems and 130 putative two-component proteins. A high proportion of signaling genes suggests complex and efficient signal-sensing regulation systems which may explain the wide distribution of *C. testosteroni* in various environments.

TeiR found in *C. testosteroni* ATCC 11966 was able to regulate steroid degradation (Linares et al. 2008). In *C. testosteroni* strain TA441, a **TesR-encoding gene** that shared a high identity with the *teiR* gene, was also found to encode a positive regulator for steroid degradation (Horinouchi et al. 2004). Blast analysis revealed that one gene (I2contig_02380) in strain I2 showed 97% identity with the *teiR* gene, which may regulate testosterone biodegradation in I2. In a recent study, TeiR was characterized as a polar-localized membrane protein, which consisted of membrane association domain, steroid binding domain and catabolic gene inducing domain (Göhler et al. 2008). It implied that this gene could be involved in both regulation of biodegradation and chemotaxis of *C. testosteroni* towards steroids. The helix-turn-helix (HTH) motif (LuxR structure) in TeiR belongs to LuxR-type DNA-binding HTH domain, which is likely to be a transcriptional activator for quorum-sensing according to the classification in Pfam database. In *P. aeruginosa* CGMCC1.860, aromatic degradation capability was found to be regulated by the *rhl* quorum sensing (QS) system (Yong and Zhong 2013). It is likely that the steroid degradation in *C. testosteroni* is also regulated by QS.

In microplate biofilm assay, a considerable amount of biofilm biomass was found to attach on the inner surface of the wells, suggesting a good biofilm formation capability of *C. testosteroni* I2 (Fig. S3 in the Supplementary Material). Corresponding genomic analysis also suggests an

importance of biofilm lifestyle of *C. testosteroni*. About 50 genes were found in different *C. testosteroni* strains involved in controlling the level of a universal biofilm signaling molecule - c-di-GMP. Interestingly, about a half of these putative c-di-GMP controlling enzymes were predicted to be multi-domain proteins containing a sensor domain. In such polydomain enzymes, sensors at the N-termini were involved in detecting environmental signals and triggering the activation of C-termini to control c-di-GMP synthesis and hydrolysis (Cotter and Stibitz 2007; Tuckerman et al. 2009). The presence of large number of multi-domain c-di-GMP controlling proteins in *C. testosteroni* suggests a good competence of this organism in natural environments by integrating various environmental factors into the c-di-GMP regulation system (Cruz et al. 2012; Paul et al. 2004; Simm et al. 2004). In addition to c-di-GMP signalling system, quorum sensing is another important regulatory network for biofilms (Joo and Otto 2012). Previous studies have shown that certain *Comamonas* species are capable of quenching AHL signal molecules, suggesting a role of QS regulatory system in *Comomonas* (Christiaen et al. 2011; Uroz et al. 2009; Uroz et al. 2007). However, the genes responsible for QS signal synthesis and degradation are not well annotated in *Comomonas* species. Hence, we could not identify QS systems in the genome of *C. testosteroni* I2.

The secretion of siderophore and exopolysaccharides was also predicted based on the draft genome of strain I2. Siderophore is a group of compounds produced by certain environmental bacteria to facilitate their iron uptake and these compounds are often considered as public goods because their production is beneficial to all bacteria within the locality who are having cognate receptors (Buckling et al. 2007). Siderophore has been demonstrated to be an important factor influencing biofilm development, especially under iron-limiting conditions (Visca et al. 2007). It has also been shown to play an important role in biodegradation of aromatic compounds by *P. putida* (Joshi et al. 2014). An increased secretion of siderophore was observed when cells were exposed to benzyl alcohol, which would facilitate degradation process and boost bacterial growth because oxygenases for degradation are iron-dependent (Joshi et al. 2014). The capability of producing and uptaking siderophores by *C. testosteroni* strain I2 renders this strain more competitive, especially under iron-limiting conditions and this capability may also benefit its catabolic activities. In biofilms, exopolysaccharide was always

considered as a key component in extracellular polymeric substances (Wingender et al. 1999). A number of genes responsible for exopolysaccharide synthesis and secretion in the draft genome of *C. testosteroni* strain I2 were identified. Moreover, one two-component system (RcsCB) that was considered as genes regulating capsular polysaccharide synthesis and biofilm formation, was found in the genome of strain I2 (Majdalani and Gottesman 2005; Mikkelsen et al. 2013). Remarkably, similar tandem arrangement of these regulatory genes and one fimbriae assembly cluster in *Pseudomonas* species and *C. testosteroni* KF-1 was also found in strain I2 (I2contig_03789 to 03801). Previous study found that RcsCB two-component system could control the assembly of fimbriae and biofilm formation in *P. aeruginosa* (Mikkelsen et al. 2009; Nicastro et al. 2009). The presence of similar regulatory system in *C. testosteroni* I2 highlights the importance of biofilm lifestyle of this organism.

In summary, we sequenced and comparatively analyzed the genome of *C. testosteroni* strain I2 with reference to 8 other *Comamonas* genomes. Our results suggest that (i) strain I2 is equipped with a sophisticated signalling system to help it find ideal niches and to switch between planktonic and biofilm lifestyles; (ii) a number of putative multidrug resistant genes coding for abundant outer membrane transporters, chaperones and heat shock proteins for the protection of cellular function allow strain I2 to adapt to various environments; (iii) strain I2 exhibits a capability of producing and secreting siderophores under iron-limiting conditions, rendering it more competitive in natural environments; and (iv) the genome of strain I2 contains a number of genes responsible for the synthesis and secretion of exopolysaccharides, an extracellular component essential for biofilm formation. Taken together, this study reveals the genomic features underlying the adaption of *C. testosteroni* to versatile environments and highlighting the importance of its biofilm lifestyle.

Acknowledgements

The authors thank Daphne Ng for her helpful suggestions. This research was supported by the National Research Foundation and Ministry of Education Singapore under its Research Centre of Excellence Programme, Singapore Centre on Environmental Life Sciences Engineering (SCELSE) (M4330005.C70) and a Start-up Grant (M4080847.030) from College of Engineering, Nanyang

453 Technological University, Singapore. The authors thank the Singapore Ministry of Education
454 (MOE2011-T2-2-035, ACR 3/12) for the research scholarship to Yichao Wu.

455 The authors do not have any conflict of interest.

456 **References**

- 457 Abraham JM, Simon GL (2007) *Comamonas testosteroni* bacteremia: A case report and review of the
458 literature. *Infect Dis Clin Pract* 15(4):272-273
- 459 Alikhan NF, Petty NK, Zakour NLB, Beatson SA (2011) BLAST Ring Image Generator (BRIG):
460 simple prokaryote genome comparisons. *BMC Genomics* 12(1):402
- 461 Andersson S, Kuttuva Rajarao G, Land CJ, Dalhammar G (2008) Biofilm formation and interactions
462 of bacterial strains found in wastewater treatment systems. *FEMS Microbiol Lett* 283(1):83-
463 90
- 464 Bathe S, Schwarzenbeck N, Hausner M (2005) Plasmid-mediated bioaugmentation of activated sludge
465 bacteria in a sequencing batch moving bed reactor using pNB2. *Lett Appl Microbiol*
466 41(3):242-247
- 467 Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD,
468 Rode CK, Mayhew GF (1997) The complete genome sequence of *Escherichia coli* K-12.
469 *Science* 277(5331):1453-1462
- 470 Boon N, Goris J, De Vos P, Verstraete W, Top EM (2000) Bioaugmentation of activated sludge by an
471 indigenous 3-chloroaniline-degrading *Comamonas testosteroni* strain, I2gfp. *Appl Environ*
472 *Microbiol* 66(7):2906-2913
- 473 Boon N, Goris J, De Vos P, Verstraete W, Top EM (2001) Genetic diversity among 3-chloroaniline-
474 and aniline-degrading strains of the *Comamonadaceae*. *Appl Environ Microbiol* 67(3):1107-
475 1115
- 476 Boon N, Top EM, Verstraete W, Siciliano SD (2003) Bioaugmentation as a tool to protect the
477 structure and function of an activated-sludge microbial community against a 3-chloroaniline
478 shock load. *Appl Environ Microbiol* 69(3):1511-1520
- 479 Booth SC, Workentine ML, Wen J, Shaykhutdinov R, Vogel HJ, Ceri H, Turner RJ, Weljie AM (2011)
480 Differences in metabolism between the biofilm and planktonic response to metal stress. *J*
481 *Proteome Res* 10(7):3190-3199
- 482 Bossier P, Verstraete W (1996) *Comamonas testosteroni* colony phenotype influences
483 exopolysaccharide production and coaggregation with yeast cells. *Appl Environ Microbiol*
484 62(8):2687-2691
- 485 Bramucci MG, Nagarajan V (2000) Industrial wastewater bioreactors: sources of novel
486 microorganisms for biotechnology. *Trends Biotechnol* 18(12):501-505
- 487 Buckling A, Harrison F, Vos M, Brockhurst MA, Gardner A, West SA, Griffin A (2007)
488 Siderophore-mediated cooperation and virulence in *Pseudomonas aeruginosa*. *FEMS*
489 *Microbiol Ecol* 62(2):135-141
- 490 Cao B, Loh KC (2008) Catabolic pathways and cellular responses of *Pseudomonas putida* P8 during
491 growth on benzoate with a proteomics approach. *Biotechnol Bioeng* 101(6):1297-1312

492 Cao B, Majors PD, Ahmed B, Renslow RS, Silvia CP, Shi L, Kjelleberg S, Fredrickson JK, Beyenal
493 H (2012) Biofilm shows spatially stratified metabolic responses to contaminant exposure.
494 Environ Microbiol 14(11):2901-2910

495 Christiaen SE, Brackman G, Nelis HJ, Coenye T (2011) Isolation and identification of quorum
496 quenching bacteria from environmental samples. J Microbiol Methods 87(2):213-219

497 Cotter PA, Stibitz S (2007) C-di-GMP-mediated regulation of virulence and biofilm formation. Curr
498 Opin Microbiol 10(1):17-23

499 Cruz DP, Huertas MG, Lozano M, Zárate L, Zambrano MM (2012) Comparative analysis of
500 diguanylate cyclase and phosphodiesterase genes in *Klebsiella pneumoniae*. BMC Microbiol
501 12(1):139

502 Das S, Jean J-S, Kar S, Chou ML, Chen CY (2014) Screening of plant growth-promoting traits in
503 arsenic-resistant bacteria isolated from agricultural soil and their potential implication for
504 arsenic bioremediation. J Hazard Mater 272:112-120

505 Ding Y, Peng N, Du Y, Ji L, Cao B (2014) Disruption of putrescine biosynthesis in *Shewanella*
506 *oneidensis* enhances biofilm cohesiveness and performance in Cr (VI) immobilization. Appl
507 Environ Microbiol 80(4):1498-1506

508 Elo S, Maunuksela L, Salkinoja-Salonen M, Smolander A, Haahtela K (2000) Humus bacteria of
509 Norway spruce stands: plant growth promoting properties and birch, red fescue and alder
510 colonizing capacity. FEMS Microbiol Ecol 31(2):143-152

511 Flemming H-C, Wingender J (2010) The biofilm matrix. Nat Rev Microbiol 8(9):623-633

512 Fredrickson JK, Romine MF, Beliaev AS, Auchtung JM, Driscoll ME, Gardner TS, Nealson KH,
513 Osterman AL, Pinchuk G, Reed JL (2008) Towards environmental systems biology of
514 *Shewanella*. Nat Rev Microbiol 6(8):592-603

515 Frost LS, Leplae R, Summers AO, Toussaint A (2005) Mobile genetic elements: the agents of open
516 source evolution. Nat Rev Microbiol 3(9):722-732

517 Göhler A, Xiong G, Paulsen S, Trentmann G, Maser E (2008) Testosterone-inducible regulator is a
518 kinase that drives steroid sensing and metabolism in *Comamonas testosteroni*. J Biol Chem
519 283(25):17380-17390

520 Gong W, Kisiela M, Schilhabel MB, Xiong G, Maser E (2012) Genome sequence of *Comamonas*
521 *testosteroni* ATCC 11996, a representative strain involved in steroid degradation. J Bacteriol
522 194(6):1633-1634

523 Gumaelius L, Magnusson G, Pettersson B, Dalhammar G (2001) *Comamonas denitrificans* sp. nov.,
524 an efficient denitrifying bacterium isolated from activated sludge. Int J Syst Evol Microbiol
525 51(3):999-1006

526 Hall-Stoodley L, Costerton JW, Stoodley P (2004) Bacterial biofilms: From the natural environment
527 to infectious diseases. Nat Rev Microbiol 2(2):95-108

528 Harayama S (1997) Polycyclic aromatic hydrocarbon bioremediation design. Curr Opin Biotechnol
529 8(3):268-273

530 Heidelberg JF, Paulsen IT, Nelson KE, Gaidos EJ, Nelson WC, Read TD, Eisen JA, Seshadri R, Ward
531 N, Methe B (2002) Genome sequence of the dissimilatory metal ion-reducing bacterium
532 *Shewanella oneidensis*. Nat Biotechnol 20(11):1118-1123

533 Hengge R (2009) Principles of c-di-GMP signalling in bacteria. Nat Rev Microbiol 7(4):263-273

534 Horinouchi M, Kurita T, Yamamoto T, Hatori E, Hayashi T, Kudo T (2004) Steroid degradation gene
535 cluster of *Comamonas testosteroni* consisting of 18 putative genes from meta-cleavage
536 enzyme gene *tesB* to regulator gene *tesR*. Biochem Biophys Res Commun 324(2):597-604

537 Hyatt D, Chen GL, LoCascio P, Land M, Larimer F, Hauser L (2010) Prodigal: prokaryotic gene
538 recognition and translation initiation site identification. *BMC Bioinformatics* 11(1):119

539 Joo H-S, Otto M (2012) Molecular basis of in vivo biofilm formation by bacterial pathogens. *Chem*
540 *Biol* 19(12):1503-1513

541 Joshi H, Dave R, Venugopalan VP (2014) Pumping iron to keep fit: modulation of siderophore
542 secretion helps efficient aromatic utilization in *Pseudomonas putida* KT2440. *Microbiology*
543 160(Pt 7):1393-1400

544 Juang YC, Adav SS, Lee DJ, Lai JY (2010) Influence of internal biofilm growth on residual
545 permeability loss in aerobic granular membrane bioreactors. *Environ Sci Technol* 44(4):1267-
546 1273

547 Kim L, Pagaling E, Zuo YY, Yan T (2014) Impact of substratum surface on microbial community
548 structure and treatment performance in biological aerated filters. *Appl Environ Microbiol*
549 80(1):177-183

550 Król J, Penrod J, McCaslin H, Rogers L, Yano H, Stancik A, Dejonghe W, Brown C, Paraes R,
551 Wuertz S (2012) Role of IncP-1 β plasmids pWDL7:: *rfp* and pNB8c in chloroaniline
552 catabolism as determined by genomic and functional analyses. *Appl Environ Microbiol*
553 78(3):828-838

554 Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009)
555 Circos: an information aesthetic for comparative genomics. *Genome Res* 19(9):1639-1645

556 Linares M, Pruneda-Paz JL, Reyna L, Genti-Raimondi S (2008) Regulation of testosterone
557 degradation in *Comamonas testosteroni*. *J Steroid Biochem Mol Biol* 112(1):145-150

558 Lu H, Chandran K, Stensel D (2014) Microbial ecology of denitrification in biological wastewater
559 treatment. *Water Res* 64:237-254

560 Luo H, Xu P, Ren Z (2012) Long-term performance and characterization of microbial desalination
561 cells in treating domestic wastewater. *Bioresour Technol* 120:187-193

562 Möbus E, Jahn M, Schmid R, Jahn D, Maser E (1997) Testosterone-regulated expression of enzymes
563 involved in steroid and aromatic hydrocarbon catabolism in *Comamonas testosteroni*. *J*
564 *Bacteriol* 179(18):5951-5955

565 Ma YF, Wu JF, Wang SY, Jiang CY, Zhang Y, Qi SW, Liu L, Zhao GP, Liu SJ (2007) Nucleotide
566 sequence of plasmid pCNB1 from *Comamonas* strain CNB-1 reveals novel genetic
567 organization and evolution for 4-chloronitrobenzene degradation. *Appl Environ Microbiol*
568 73(14):4477-4483

569 Ma YF, Zhang Y, Zhang JY, Chen DW, Zhu Y, Zheng H, Wang SY, Jiang CY, Zhao GP, Liu SJ
570 (2009) The complete genome of *Comamonas testosteroni* reveals its genetic adaptations to
571 changing environments. *Appl Environ Microbiol* 75(21):6812-6819

572 Majdalani N, Gottesman S (2005) The Rcs phosphorelay: A complex signal transduction system
573 *Annu Rev Microbiol. Annu Rev Microbiol*, 59:379-405

574 Mampel J, Maier E, Tralau T, Ruff J, Benz R, Cook A (2004) A novel outer-membrane anion channel
575 (porin) as part of a putatively two-component transport system for 4-toluenesulphonate in
576 *Comamonas testosteroni* T-2. *Biochem J* 383:91-99

577 Markowitz VM, Chen I-MA, Chu K, Szeto E, Palaniappan K, Pillay M, Ratner A, Huang J, Pagani I,
578 Tringe S (2014) IMG/M 4 version of the integrated metagenome comparative analysis system.
579 *Nucleic Acids Res* 42(D1):D568-D573

580 Markowitz VM, Mavromatis K, Ivanova NN, Chen I-MA, Chu K, Kyrpides NC (2009) IMG ER: a
581 system for microbial genome annotation expert review and curation. *Bioinformatics*
582 25(17):2271-2278

583 Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads.
584 EMBnet J 17(1):10-12

585 Mikkelsen H, Ball G, Giraud C, Filloux A (2009) Expression of *Pseudomonas aeruginosa* *CupD*
586 fimbrial genes is antagonistically controlled by RcsB and the EAL-containing PvrR response
587 regulators. PLoS ONE 4(6):e6018

588 Mikkelsen H, Hui KL, Barraud N, Filloux A (2013) The pathogenicity island encoded PvrSR/RcsCB
589 regulatory network controls biofilm formation and dispersal in *Pseudomonas aeruginosa*
590 PA14. Mol Microbiol 89(3):450-463

591 Mohanty A, Liu Y, Yang L, Cao B (2015) Extracellular biogenic nanomaterials inhibit pyoverdine
592 production in *Pseudomonas aeruginosa*: a novel insight into impacts of metal(loid)s on
593 environmental bacteria. Appl Microbiol Biotechnol 99(4):1957-1966

594 Neilands J (1995) Siderophores: structure and function of microbial iron transport compounds. J Biol
595 Chem 270(45):26723-26726

596 Nelson K, Weinel C, Paulsen I, Dodson R, Hilbert H, Martins dos Santos V, Fouts D, Gill S, Pop M,
597 Holmes M (2002) Complete genome sequence and comparative analysis of the metabolically
598 versatile *Pseudomonas putida* KT2440. Environ Microbiol 4(12):799-808

599 Ng CK, Sivakumar K, Liu X, Madhaiyan M, Ji L, Yang L, Tang C, Song H, Kjelleberg S, Cao B
600 (2013) Influence of outer membrane c-type cytochromes on particle size and activity of
601 extracellular nanoparticles produced by *Shewanella oneidensis*. Biotechnol Bioeng
602 110(7):1831-1837

603 Ni B, Huang Z, Fan Z, Jiang CY, Liu SJ (2013a) *Comamonas testosteroni* uses a chemoreceptor for
604 tricarboxylic acid cycle intermediates to trigger chemotactic responses towards aromatic
605 compounds. Mol Microbiol 90(4):813-823

606 Ni B, Zhang Y, Chen DW, Wang BJ, Liu SJ (2013b) Assimilation of aromatic compounds by
607 *Comamonas testosteroni*: characterization and spreadability of protocatechuate 4, 5-cleavage
608 pathway in bacteria. Appl Microbiol Biotechnol 97(13):6031-6041

609 Nicastro GG, Boechat AL, Abe CM, Kaihama GH, Baldini RL (2009) *Pseudomonas aeruginosa* PA14
610 *cupD* transcription is activated by the RcsB response regulator, but repressed by its putative
611 cognate sensor RcsC. FEMS Microbiol Lett 301(1):115-123

612 Ochman H, Lawrence JG, Groisman EA (2000) Lateral gene transfer and the nature of bacterial
613 innovation. Nature 405(6784):299-304

614 Pandey G, Jain RK (2002) Bacterial chemotaxis toward environmental pollutants: role in
615 bioremediation. Appl Environ Microbiol 68(12):5789-5795

616 Patel K, Patel M, Chauhan K, Anto H, Trivedi U (2007) Production of an antioxidant naphthoquinone
617 pigment by *Comamonas testosteroni* during growth on naphthalene. J Sci Ind Res 66(8):605

618 Paul R, Weiser S, Amiot NC, Chan C, Schirmer T, Giese B, Jenal U (2004) Cell cycle-dependent
619 dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase
620 output domain. Genes Dev 18(6):715-727

621 Schlüter A, Szczepanowski R, Pühler A, Top EM (2007) Genomics of IncP-1 antibiotic resistance
622 plasmids isolated from wastewater treatment plants provides evidence for a widely accessible
623 drug resistance gene pool. FEMS Microbiol Rev 31(4):449-477

624 Schwyn B, Neilands J (1987) Universal chemical assay for the detection and determination of
625 siderophores. Anal Biochem 160(1):47-56

626 Serres MH, Riley M (2006) Genomic analysis of carbon source metabolism of *Shewanella oneidensis*
627 MR-1: predictions versus experiments. J Bacteriol 188(13):4601-4609

628 Shao Y, He X, Harrison EM, Tai C, Ou HY, Rajakumar K, Deng Z (2010) mGenomeSubtractor: a
629 web-based tool for parallel *in silico* subtractive hybridization analysis of multiple bacterial
630 genomes. *Nucleic Acids Res* 38(suppl 2):W194-W200

631 Shokrollahzadeh S, Azizmohseni F, Golmohammad F, Shokouhi H, Khademhaghighat F (2008)
632 Biodegradation potential and bacterial diversity of a petrochemical wastewater treatment
633 plant in Iran. *Bioresour Technol* 99(14):6127-6133

634 Simm R, Morr M, Kader A, Nimtz M, Römling U (2004) GGDEF and EAL domains inversely
635 regulate cyclic di-GMP levels and transition from sessility to motility. *Mol Microbiol*
636 53(4):1123-1134

637 Singh R, Paul D, Jain RK (2006) Biofilms: implications in bioremediation. *Trends Microbiol*
638 14(9):389-397

639 Sylvestre M (1995) Biphenyl/chlorobiphenyls catabolic pathway of *Comamonas testosteroni* B-356:
640 Prospect for use in bioremediation. *Int Biodeterior Biodegradation* 35(1):189-211

641 Takeo M, Ohara A, Sakae S, Okamoto Y, Kitamura C, Kato D-i, Negoro S (2013) Function of a
642 glutamine synthetase-like protein in bacterial aniline oxidation via γ -glutamylanilide. *J*
643 *Bacteriol* 195(19):4406-4414

644 Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary
645 Genetics Analysis Version 6.0. *Mol Biol Evol* 30(12):2725-2729

646 Tuckerman JR, Gonzalez G, Sousa EH, Wan X, Saito JA, Alam M, Gilles-Gonzalez M-A (2009) An
647 oxygen-sensing diguanylate cyclase and phosphodiesterase couple for c-di-GMP control.
648 *Biochemistry* 48(41):9764-9774

649 Uroz S, Dessaux Y, Oger P (2009) Quorum sensing and quorum quenching: the yin and yang of
650 bacterial communication. *ChemBioChem* 10(2):205-16

651 Uroz S, Oger P, Chhabra SR, Camara M, Williams P, Dessaux Y (2007) N-acyl homoserine lactones
652 are degraded via an amidolytic activity in *Comamonas* sp. strain D1. *Arch Microbiol*
653 187(3):249-56

654 Varani AM, Siguier P, Gourbeyre E, Charneau V, Chandler M (2011) ISSaga is an ensemble of web-
655 based methods for high throughput identification and semi-automatic annotation of insertion
656 sequences in prokaryotic genomes. *Genome Biol* 12(3):R30

657 Visca P, Imperi F, Lamont IL (2007) Pyoverdine siderophores: from biogenesis to biosignificance.
658 *Trends Microbiol* 15(1):22-30

659 Wakimoto N, Nishi J, Sheikh J, Nataro JP, Sarantuya J, Iwashita M, Manago K, Tokuda K,
660 Yoshinaga M, Kawano Y (2004) Quantitative biofilm assay using a microtiter plate to screen
661 for enteroaggregative *Escherichia coli*. *Am J Trop Med Hyg* 71(5):687-690

662 Wandersman C, Delepelaire P (2004) Bacterial iron sources: from siderophores to hemophores. *Annu*
663 *Rev Microbiol* 58:611-647

664 Wang VB, Chua S-L, Cao B, Seviour T, Nesatyy VJ, Marsili E, Kjelleberg S, Givskov M, Tolker-
665 Nielsen T, Song H (2013) Engineering PQS biosynthesis pathway for enhancement of
666 bioelectricity production in *Pseudomonas aeruginosa* microbial fuel cells. *PLoS ONE*
667 8(5):e63129

668 Watanabe K, Futamata H, Harayama S (2002) Understanding the diversity in catabolic potential of
669 microorganisms for the development of bioremediation strategies. *Antonie van Leeuwenhoek*
670 81(1-4):655-663

671 Willems A, De Vos P (2006) *Comamonas*. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H,
672 Stackebrandt E (eds) *The Prokaryotes*. Springer New York, pp 723-736

673 Wingender J, Neu T, Flemming H-C (1999) What are Bacterial Extracellular Polymeric Substances?
674 In: Wingender J, Neu T, Flemming H-C (eds) Microbial Extracellular Polymeric Substances.
675 Springer Berlin Heidelberg, pp 1-19

676 Wu J, Jiang C, Wang B, Ma Y, Liu Z, Liu S (2006) Novel partial reductive pathway for 4-
677 chloronitrobenzene and nitrobenzene degradation in *Comamonas* sp. strain CNB-1. Appl
678 Environ Microbiol 72(3):1759-1765

679 Wu Y, Ding Y, Cohen Y, Cao B (2015) Elevated level of the second messenger c-di-GMP in
680 *Comamonas testosteroni* enhances biofilm formation and biofilm-based biodegradation of 3-
681 chloroaniline. Appl Microbiol Biotechnol 99(4):1967-1976

682 Xiong J, Li D, Li H, He M, Miller SJ, Yu L, Rensing C, Wang G (2011) Genome analysis and
683 characterization of zinc efflux systems of a highly zinc-resistant bacterium, *Comamonas*
684 *testosteroni* S44. Res Microbiol 162(7):671-679

685 Yong YC, Zhong JJ (2013) Regulation of aromatics biodegradation by *rhl* quorum sensing system
686 through induction of catechol meta-cleavage pathway. Bioresour Technol 136:761-765

687 Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn
688 graphs. Genome Res 18(5):821-829

689 Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS (2011) PHAST: a fast phage search tool.
690 Nucleic Acids Res 39(suppl 2):W347-W352

691

692 **Table 1** General features of 9 *Comamonas* strains.

	I2	ATCC 11996	CNB-2	KF-1	NBRC 100989	S44	B-9	DSM 17552	DSM 21721
Size (bp)	5778710	5410155	5373643	6026527	5589358	5534155	5217447	3681020	4639907
GC content (%)	61.88	61.48	61.40	61.79	61.90	61.42	63.6	65.96	63.24
Predicted CDS	5365	5056	4893	5606	5169	5218	5249	3550	4165
Coding density (%)	87.47	86.82	86.84	87.54	87.28	86.82	88.43	92.18	87.95
Protein coding genes	5263 (98.10%)	4984 (98.58%)	4805 (98.20%)	5492 (97.97%)	5079 (98.26%)	5154 (98.77%)	5165 (98.40%)	3489 (98.28%)	4075 (97.84%)
rRNA genes	6 (0.11%)	3 (0.06%)	9 (0.18%)	18 (0.32%)	4 (0.08%)	3 (0.06%)	5 (0.10%)	7 (0.20%)	9 (0.22%)
tRNA genes	79 (1.47%)	69 (1.36%)	79 (1.61%)	96 (1.71%)	68 (1.32%)	61 (1.17%)	67 (1.28%)	41 (1.15%)	64 (1.54%)
Protein coding genes with function prediction	4324 (80.60%)	3776 (74.68%)	3762 (76.88%)	4009 (71.51%)	4275 (82.70%)	3831 (73.42%)	4103 (78.17%)	3082 (86.82%)	3551 (85.26%)
Prophages	126 (2.35%)	96 (1.90%)	129 (2.64%)	213 (3.80%)	16 (0.31%)	224 (4.29%)	-	0	25 (0.60%)
IS elements	76 (1.42%)	43 (0.85%)	70 (1.43%)	121 (2.16%)	47 (0.91%)	29 (0.56%)	-	30 (0.84%)	60 (1.44%)
Protein coding genes with COGs	3692 (68.82%)	3564 (70.49%)	3116 (63.68%)	3746 (66.82%)	3636 (70.34%)	3520 (67.46%)	3293 (62.74%)	2651 (74.68%)	3060 (73.47%)
Protein coding genes connected to KEGG pathways	1405 (26.19%)	1400 (27.69%)	1387 (28.35%)	1440 (25.69%)	1378 (26.66%)	1415 (27.12%)	1249 (23.80%)	1202 (33.86%)	1304 (31.31%)
Protein coding genes connected to transporter	891 (16.61%)	838 (16.57%)	816 (16.68%)	905 (16.14%)	878 (16.99%)	857 (16.42%)	904 (17.22%)	556 (15.66%)	759 (18.22%)

693 **Table 2.** Putative genes involved in cellular c-di-GMP concentration regulation

Locus tag	Domain organization
I2contig_00589, I2contig_00707, I2contig_00762, I2contig_01440, I2contig_01530, I2contig_01917, I2contig_01952, I2contig_03459, I2contig_03480, I2contig_03725, I2contig_03774, I2contig_04045, I2contig_04217, I2contig_04219, I2contig_04562, I2contig_04723, I2contig_04913, I2contig_05141	GGDEF domain
I2contig_01930	GAF-GGDEF domain
I2contig_00532, I2contig_02501	MASE-GGDEF domain
I2contig_00808, I2contig_00961, I2contig_01929, I2contig_01968, I2contig_02408, I2contig_04221	PAS-GGDEF domain
I2contig_01762, I2contig_02364, I2contig_04092	Cache-GGDEF domain
I2contig_02208, I2contig_03799	EAL domain
I2contig_05198	GAF-EAL domain
I2contig_00704, I2contig_02058, I2contig_03401, I2contig_05064	HD-GYP domain
I2contig_00763, I2contig_05344	GGDEF-EAL domain
I2contig_00089, I2contig_00705, I2contig_01182, I2contig_01291, I2contig_01841, I2contig_03538, I2contig_03688, I2contig_04043, I2contig_04110, I2contig_04384, I2contig_04596	PAS-GGDEF-EAL domain
I2contig_02275, I2contig_03397	MHYT-GGDEF-EAL
I2contig_01506, I2contig_01507, I2contig_04843	EAL-HD-GYP domain

694

695

Figure captions

Fig. 1. Circular diagram of *C. testosteroni* I2. The outer scale is marked in kb. Circles from the outer to the inner (1 to 6) provide the following information: (1) position of contigs; (2) forward strand CDSs; (3) reverse strand CDSs; (4) predicated transposases (green indicates IS256, blue indicates IS3, red indicates IS5, purple indicates ISL3); (5) predicated prophages; (6) rRNAs (green), tRNAs (red). Colors in circles (2) and (3) represent respective assigned COG categories: (a) amino acid transport and metabolism, (b) carbohydrate transport and metabolism, (c) cell cycle control, cell division, chromosome partitioning, (d) cell motility, (e) cell wall, membrane, envelope biogenesis, (f) chromatin structure and dynamics, (g) coenzyme transport and metabolism, (h) defense mechanisms, (i) energy production and conversion, (j) function unknown, (k) general function prediction only, (l) inorganic transport and metabolism, (m) intracellular trafficking, secretion and vesicular transport, (n) lipid transport and metabolism, (o) nucleotide transport and metabolism, (p) posttranslational modification, protein turnover, chaperones, (q) replication, recombination and repair, (r) RNA processing and modification, (s) secondary metabolites biosynthesis, transport and catabolism, (t) signal transduction mechanisms, (u) transcription, (v) translation, ribosomal structure and biogenesis.

Fig. 2. The percentage of conserved and strain-specific genes in *C. testosteroni* I2 compared with other 8 *Comamonas* strains. Genes with a BLASTP-derived homology value higher than 0.81 were classified as conserved (blue), while those with a homology value lower than 0.42 were considered strain-specific (red). The remaining genes were designated as unclassified (grey).

Fig. 3. Heat map comparison of proportions of COG categories between 9 *Comamonas* strains, *E. coli* K12, *P. putida* KT2440, *S. oneidensis* MR-1. The categories were denoted as: (a) amino acid transport and metabolism, (b) carbohydrate transport and metabolism, (c) cell cycle control, cell division, chromosome partitioning, (d) cell motility, (e) cell wall, membrane, envelope biogenesis, (f) chromatin structure and dynamics, (g) coenzyme transport and metabolism, (h) defense mechanisms, (i) energy production and conversion, (j) function unknown, (k) general function prediction only, (l) inorganic transport and metabolism, (m) intracellular trafficking, secretion and vesicular transport, (n)

lipid transport and metabolism, (o) nucleotide transport and metabolism, (p) posttranslational modification, protein turnover, chaperones, (q) replication, recombination and repair, (r) RNA processing and modification, (s) secondary metabolites biosynthesis, transport and catabolism, (t) signal transduction mechanisms, (u) transcription, (v) translation, ribosomal structure and biogenesis.

Fig. 4. Gene clusters involved in aromatic compounds metabolism in draft genome sequence of *C. testosteroni* I2. The black arrows indicate the degradation pathway for aromatic compounds, while dash arrows indicated the lack of specific genes. The filled arrows denote the genes involved in catabolism while the empty ones denote the genes in the same cluster without clear catabolic function.

p2mo, phenol 2-monooxygenase; *praC*, 4-oxalocrotonate tautomerase; *hsaB*, 3-hydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione monooxygenase reductase; *hsaD*, 4,5:9,10-diseco-3-hydroxy-5,9,17-trioxoandrosta-1(10),2-diene-4-oate hydrolase; *tesE*, 2-oxopent-4-enoate/cis-2-oxohex-4-enoate hydratase; *tesF*, acetaldehyde/propanal dehydrogenase; *tesG*, 4-hydroxy-2-oxovalerate/4-hydroxy-2-oxohexanoate aldolase; *catB*, muconate cycloisomerase; *catC*, muconolactone D-isomerase; *atoE*, short-chain fatty acids transporter; *catC*, muconolactone D-isomerase; *pcaD*, 3-oxoadipate enol-lactonase; *paaJ*, 3-oxoadipyl-CoA thiolase; *pcaJ*, 3-oxoadipate CoA-transferase, β -subunit; *pcaI*, 3-oxoadipate CoA-transferase, α -subunit; *vanA*, vanillate O-demethylase monooxygenase subunit; *vanB*, vanillate O-demethylase ferredoxin subunit; *ligC*, 2-hydroxy-4-carboxymuconate semialdehyde hemiacetal dehydrogenase; *ligB*, protocatechuate 4,5-dioxygenase, β -chain; *ligA*, protocatechuate 4,5-dioxygenase, α -chain; *ligI*, 2-pyrone-4,6-dicarboxylate lactonase; *ligK*, 4-hydroxy-4-methyl-2-oxoglutarate aldolase; *ligJ*, 4-oxalmesaconate hydratase; *pcaK*, 4-hydroxybenzoate transporter; β -HSD, β -hydroxysteroid dehydrogenase; *SDI*, steroid δ -isomerase; *kshB*, 3-ketosteroid 9 α -monooxygenase subunit B; *tesI*, 3-oxo-5 α -steroid 4-dehydrogenase; *kstD*, 3-oxosteroid 1-dehydrogenase; *hsaA*, 3-hydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione monooxygenase; *kshA*, 3-ketosteroid 9 α -monooxygenase subunit A; DOHNAA, 9,17-dioxo-1,2,3,4,10,19-hexanorandrostan-5-oic acid.

Fig. 5. (a) Circular map of plasmid pI2 by comparative BLASTN analysis against three other IncP-1 β plasmids. The outermost rings represent plasmid pI2 and all identified genes on the plasmid. The next three colored concentric rings depict shared identity with other IncP-1 β plasmids based on BLAST analysis, pCNCB-1 (green), pWDL7::*rfp* (blue) and pTB30 (red) (from out to insider). (b) The aniline degradation pathway encoded by plasmid pI2.

Fig. 6. Specific genes involved in c-di-GMP regulation. (a) Numbers of c-di-GMP-controlling enzymes in different organisms, (b) c-di-GMP-controlling enzymes binding with sensor domains.