

Research article

**Open Access**

# Metabolic analysis of the soil microbe *Dechloromonas aromatica* str. RCB: indications of a surprisingly complex life-style and cryptic anaerobic pathways for aromatic degradation

Kennan Kellaris Salinero\*<sup>1,4</sup>, Keith Keller<sup>2</sup>, William S Feil<sup>1,5</sup>, Helene Feil<sup>1,5</sup>, Stephan Trong<sup>3</sup>, Genevieve Di Bartolo<sup>3,6</sup> and Alla Lapidus<sup>3</sup>

Address: <sup>1</sup>Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA, <sup>2</sup>Physical Biosciences Division, Lawrence Berkeley National Laboratories, Berkeley, CA 94710, USA, <sup>3</sup>Genomics Division, DOE Joint Genome Institute, Walnut Creek, CA 94598, USA, <sup>4</sup>Yamana Science and Technology, Washington, DC 20009, USA, <sup>5</sup>Land for Urban Wildlife, Concord, CA 94527, USA and <sup>6</sup>Boards 'N More, Brentwood, CA 94513, USA

Email: Kennan Kellaris Salinero\* - [kellarkv@nature.berkeley.edu](mailto:kellarkv@nature.berkeley.edu); Keith Keller - [kkeller@lbl.gov](mailto:kkeller@lbl.gov); William S Feil - [bhfeil@nature.berkeley.edu](mailto:bhfeil@nature.berkeley.edu); Helene Feil - [bhfeil@nature.berkeley.edu](mailto:bhfeil@nature.berkeley.edu); Stephan Trong - [trong1@lbl.gov](mailto:trong1@lbl.gov); Genevieve Di Bartolo - [gen.dibartolo@gmail.com](mailto:gen.dibartolo@gmail.com); Alla Lapidus - [ALapidus@lbl.gov](mailto:ALapidus@lbl.gov)

\* Corresponding author

Published: 3 August 2009

Received: 1 December 2008

*BMC Genomics* 2009, **10**:351 doi:10.1186/1471-2164-10-351

Accepted: 3 August 2009

This article is available from: <http://www.biomedcentral.com/1471-2164/10/351>

© 2009 Salinero et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

**Background:** Initial interest in *Dechloromonas aromatica* strain RCB arose from its ability to anaerobically degrade benzene. It is also able to reduce perchlorate and oxidize chlorobenzoate, toluene, and xylene, creating interest in using this organism for bioremediation. Little physiological data has been published for this microbe. It is considered to be a free-living organism.

**Results:** The *a priori* prediction that the *D. aromatica* genome would contain previously characterized "central" enzymes to support anaerobic aromatic degradation of benzene proved to be false, suggesting the presence of novel anaerobic aromatic degradation pathways in this species. These missing pathways include the benzylsuccinate synthase (*bssABC*) genes (responsible for fumarate addition to toluene) and the central benzoyl-CoA pathway for monoaromatics. In depth analyses using existing TIGRFam, COG, and InterPro models, and the creation of *de novo* HMM models, indicate a highly complex lifestyle with a large number of environmental sensors and signaling pathways, including a relatively large number of GGDEF domain signal receptors and multiple quorum sensors. A number of proteins indicate interactions with an as yet unknown host, as indicated by the presence of predicted cell host remodeling enzymes, effector enzymes, hemolysin-like proteins, adhesins, NO reductase, and both type III and type VI secretory complexes. Evidence of biofilm formation including a proposed exopolysaccharide complex and exosortase (*epsH*) are also present. Annotation described in this paper also reveals evidence for several metabolic pathways that have yet to be observed experimentally, including a sulphur oxidation (*soxFCDYZAXB*) gene cluster, Calvin cycle enzymes, and proteins involved in nitrogen fixation in other species (including RubisCo, ribulose-phosphate 3-epimerase, and *nif* gene families, respectively).

**Conclusion:** Analysis of the *D. aromatica* genome indicates there is much to be learned regarding the metabolic capabilities, and life-style, for this microbial species. Examples of recent gene duplication events in signaling as well as dioxygenase clusters are present, indicating selective gene family expansion as a relatively recent event in *D. aromatica*'s evolutionary history. Gene families that constitute metabolic cycles presumed to create *D. aromatica*'s environmental 'foot-print' indicate a high level of diversification between its predicted capabilities and those of its close relatives, *A. aromaticum* str EbN1 and *Azoarcus* BH72.

## Background

*D. aromatica* strain RCB is a gram negative Betaproteobacterium found in soil environments [1]. Other members of the Betaproteobacteria class are found in environmental samples (such as soil and sludge) or are pathogens (such as *Ralstonia solanacearum* in plants and *Neisseria meningitidis* in humans) and in general the genus *Dechloromonas* has been found to be ubiquitous in the environment.

A facultative anaerobe, *D. aromatica* was initially isolated from Potomac River sludge contaminated with BTEX compounds (benzene, toluene, ethylbenzene and xylene) based on its ability to anaerobically degrade chlorobenzoate [1]. This microbe is capable of aromatic hydrocarbon degradation and perchlorate reduction, and can oxidize Fe(II) and H<sub>2</sub>S [2]. Although several members of the Rhodocyclales group of Betaproteobacteria are of interest to the scientific community due to their ability to anaerobically degrade derivatives of benzene, *D. aromatica* is the first pure culture capable of anaerobic degradation of the stable underivitized benzene molecule to be isolated. This, along with its ability to reduce perchlorate (a teratogenic contaminant introduced into the environment by man) and inquiry into its use in biocells [3] has led to interest in using this organism for bioremediation and energy production. Since the isolation of *D. aromatica*, other species of *Azoarcus* have been found to possess the ability to anaerobically degrade benzene, but have not been genomically sequenced [4].

The pathway for anaerobic benzene degradation has been partially deduced [5], but the enzymes responsible for this process have yet to be identified, and remain elusive even after the intensive annotation efforts described here-in. Conversely, central anaerobic pathways for aromatic compounds described in various other species were not found to be present in this genome [6].

## Methods

### Sequencing

Three libraries (3 kb, 8 kb and 30 kb) were generated by controlled shearing (Hydroshear, Genomic Solutions, Ann Arbor, MI) of spooled genomic DNA isolated from *D. aromatica* strain RCB and inserted into pUC18, pCUGIblu21, and pcc1Fos vectors, respectively. Clonal DNA was amplified using rolling circular amplification <http://www.jgi.doe.gov/> and sequenced on ABI 3700 capillary DNA sequencers (Applied Biosystems, Foster City, CA) using BigDye technology (Perkin Elmer Corporation, Waltham, MA). Paired end-reads [7] were used to aid in assembly, and proved particularly useful in areas of repeats.

The Phrap algorithm [8,9] was used for initial assembly. Finishing and manual curation was conducted on

CONSED v14 software [10], supplemented with a suite of finishing analysis tools provided by the Joint Genome Institute. *In silico* cross-over errors were corrected by manual creation of fake reads to guide the assembly by forcing the consensus to follow the correct path.

Gaps were closed through a combination of primer walks on the gap-spanning clones from the 3 and 8 kb libraries (identified by paired-end analysis in the CONSED software) as well as sequencing of mapped, unique PCR products from freshly prepared genomic DNA.

The final step required to create a finished single chromosomal sequence was to determine the number of tandem repeats for a 672 base DNA sequence of unknown length. This was done by creating the full tandem repeat insert from unique upstream and downstream primers using long-range PCR. We then determined the size of product (amplified DNA) between the unique sequences.

### Protein sequence predictions/orfs

Annotation done at Oak Ridge National Laboratory consisted of gene calls using CRITICA [11], glimmer [12], and Generation <http://compbio.ornl.gov/>. Annotation at the Virtual Institute for Microbial Stress and Survival <http://www.microbesonline.org> used bidirectional best hits as well as recruitment to TIGRfam hidden Markov models (HMMs), as described in Alm et al. [13]. Briefly, protein coding predictions derived from NCBI, or identified using CRITICA, with supplemental input from Glimmer, were analyzed for domain identities using the models deposited in the InterPro, UniProt, PRODOM, Pfam, PRINTS, SMART, PIR SuperFamily, SUPERFAMILY, and TIGRfam databases [13]. Orthologs were identified using bidirectional unique best hits with greater than 75% coverage. RPS-BLAST against the NCBI COGs (Clusters of Orthologous Genes) in the CDD database were used to assign proteins to COG models when the best hit E-value was <1e<sup>-5</sup> and coverage was >60%.

### Manual curation

Each and every predicted protein in the VIMSS database <http://www.microbesonline.org> [13] was assessed to compare insights obtained from recruitment to models from several databases (TIGRfams, COGs, EC and InterPro). Assignments that offered the most definitive functional assignment were captured in an excel spreadsheet with data entries for all proteins predicted in the VIMSS database. Extensive manual curation of the predicted protein set was carried out using a combination of tools including the VIMSS analysis tools, creation and assessment of HMMs, and phylogenomic analysis, as described [see Additional files 1, 2]. Changes in gene functional predictions and naming were captured in the excel spreadsheet, and predictions with strong phylogenetic evidence of

function posted using the interactive VIMSS web-based annotation interface.

#### **Phylogenomic analysis: Flower Power, SCI PHY and HMM scoring**

Hidden Markov models were generated for a large subset of proteins of interest, as detailed [see Additional file 1], to predict functional classification with the highest confidence measures currently available. The HMMs allowed recruitment of proteins to phylogenetic tree alignments that most closely reflect evolutionary relatedness across species. The proteins were assembled within clades of proteins that are aligned along their full length (no missing functional domains), and that allow high confidence of shared function in each species.

#### **Gene Family Expansion**

A clustered set of paralogs [see Additional files 1, 2] was used to search for recent gene duplication events. After an initial assessment of the VIMSS gene information/homolog data, candidate proteins were used as seed sequences for Flower Power and internal tree-viewing tools or SCI-PHY analyses. These two approaches employed neighbor-joining trees using the Scoredist correction setting in the Belvu alignment editor, or the SCI-PHY utility and tree viewer. In either case resulting phylogenomic tree builds were reviewed, and contiguous protein alignments of two or more proteins from *D. aromatica* were considered to be candidates for a gene duplication event, either in the *D. aromatica* genome or in a predecessor species.

#### **Resequencing to verify absence of plasmid structure**

After finishing the *D. aromatica* genome, analysis of the annotated gene set revealed the notable absence of several anaerobic aromatic degradation pathways that were expected to be present, due to their presence in *A. aromaticum* EbN1 (an evolutionary near-neighbor, as determined by 16sRNA phylogeny). Because many catabolic pathways are encoded on plasmid DNA, we felt it was important to preclude this possibility. We re-isolated DNA from a clonal preparation of *D. aromatica* that experimentally supported anaerobic benzene degradation, using three different plasmid purification protocols, each based on different physical parameters. All three generated a single band of DNA. The protocol that generated the highest yield of DNA was used to create a complete, new library of 2 kb inserts, and the library was submitted to sequence analysis using the protocols previously cited.

## **Results**

### **Overview of Gene and Protein Features**

The finished sequence for *D. aromatica* reveals a single circular, closed chromosome of 4,501,104 nucleotides created from 130,636 screened reads, with an average G+C

content of 60% and an extremely high level of sequence coverage (average depth of 24 reads/base [see Additional file 3]). Specific probing for plasmids confirmed no plasmid structure was present in the clonal species sequenced, which supports anaerobic benzene degradation. It is noted however that the presence of two *tra* clusters (putative conjugal transfer genes; VIMSS582582-582597 and VIMSS582865-582880), as well as plasmid partitioning proteins, indicates this microbial species is likely to be transformationally competent and thus likely to be able to support plasmid DNA structures.

The Virtual Institute for Microbial Stress and Survival (VIMSS, <http://www.microbesonline.org>) and the Joint Genome Institute [http://genome.jgi-psf.org/finished\\_microbes/decar/decar.home.html](http://genome.jgi-psf.org/finished_microbes/decar/decar.home.html) report 4170 and 4204 protein coding genes, respectively [see Additional file 3]. Cross-database comparisons were done to assure the highest probability of capturing candidate orfs for analysis. The majority of proteins are shared between data sets. Variations in N-termini start sites were noted, both between JGI and VIMSS datasets and between initial and later annotation runs (approximately 200 N-termini differences between four runs of orf predictions were noted for the initial two annotation runs, Joint Genome Institute's, done at Oak Ridge National Laboratories – ORNL, and VIMSS).

The most definitive functional classification, TIGRfams, initially defined approximately 10% of the proteins in this genome; as of this writing, 33% of predicted proteins in the *D. aromatica* genome are covered by TIGRfams, leaving 2802 genes with no TIGRfam classification [see Additional file 4]. Many proteins in the current and initial non-covered sets were investigated further using K. Sjölander's HMM building protocols (many of which are available at <http://phylogenomics.berkeley.edu>), to supplement TIGRfams. The Clusters of Orthologous Genes (COG) assignments were used for classification in the families of signaling proteins, but specific function predictions for these proteins also required further analyses. The metabolic and signaling pathways are discussed below, and the identity of orthologs within these pathways are based on analysis of phylogenomic profiles of clusters obtained by HMM analysis, with comparison to proteins having experimentally defined function.

### **Anaerobic aromatic degradation – absence of known enzymes indicates novel pathways**

One of the more striking findings is the absence of known key enzymes for monoaromatic degradation under anaerobic conditions. One of the primary metabolic capabilities of interest for this microbe is anaerobic degradation of benzene. Fumarate addition to toluene via benzylsuccinate synthase (BssABCD) is recognized as the common

mechanism for anaerobic degradation by a phylogenomically diverse population of microbes [14-16] and has been called "the paradigm of anaerobic hydrocarbon oxidation" [17]. Benzoyl CoA is likewise considered a central intermediate in anaerobic degradation, and is further catabolized via benzoyl CoA reductase (BcrAB) [17]. Populated KEGG maps in the IMG and VIMSS databases, based on BLAST analyses, indicate the presence of some of the enzymes previously characterized as belonging to the Bss pathway in *D. aromatica*, yet more careful analysis shows the candidate enzymes to be members of a general family, rather than true orthologs of the enzymes in question. The majority of catabolic enzymes of interest for *D. aromatica* are not covered by TIGRfams or COGs families. For this reason Flower Power clustering, SCI-PHY subfamily clade analysis, and HMM scoring were used to ascertain the presence or absence of proteins of interest (for a detailed description, see Additional file 1). The most reliable prediction-of-function approaches for genomically sequenced protein orfs are obtained using the more computationally intensive HMM modelling and scoring utilities. This allows the protein in question to be assessed by phylogenetic alignment to protein families or sub-families with experimentally known function, providing much more accurate predictions [18,19].

To explore the apparent lack of anaerobic aromatic degradation pathways expected to be present in this genome, all characterized anaerobic aromatic degradation pathways from *A. aromaticum* EbN1 [20] were defined by HMMs to establish presence or absence of proteins in both the *D. aromatica* and *Azoarcus* BH72 genomes (these three genomes comprise nearest-neighbor species in currently sequenced species [see Additional file 1]). In *A. aromaticum* EbN1, ten major catabolic pathways have been found for anaerobic aromatic degradation, and nine of the ten converge on benzoyl-CoA [21]. A key catalytic enzyme or subunit for each enzymatic step was used as a seed sequence to recruit proteins from a non-redundant set of Genbank proteins for phylogenetic analysis. Benzylsuccinate synthase, present in *A. aromaticum* EbN1 [20,22] as well as *Thauera aromatica* [6], and *Geobacter metallireducens* [23], is not present in either the *D. aromatica* or *Azoarcus* BH72 genomes (see Table 1). Benzoyl-CoA reductase and benzylsuccinate synthase, previously denoted as "central" to anaerobic catabolism of aromatics, are likewise absent. The set of recruited proteins for both benzylsuccinate synthase and benzoyl-CoA reductase indicate they are not as universally present as has been suggested. *D. aromatica* does encode a protein in the pyruvate formate lyase family, but further analysis shows that it is more closely related to the *E. coli* homolog of this protein (which is not involved in aromatic catabolism) than to BssA. Anaerobic reduction of ethylbenzene is carried out by ethylbenzene dehydrogenase (EbdABCD1, 2) in *A. aromaticum*. This

complex belongs to the membrane bound nitrate reductase (NarDKGHJI) family. In *D. aromatica*, this complex of proteins is only present as the enzymatically characterized perchlorate reductase (PcrABCD; [24]) which utilizes perchlorate, rather than nitrate, as the electron acceptor. EbdABCD proteins in *A. aromaticum* (VIMSS814904-814907 and VIMSS816928-816931) occur in operons that include (S)-1-phenylethanol dehydrogenases (Ped; VIMSS 814903 and 816927) [25], both of which are absent from *D. aromatica*, as is the acetophenone carboxylase that catalyzes ATP-dependent carboxylation of acetophenone produced by Ped.

For all pathways except the ubiquitous phenylacetic acid catabolic cluster, which is involved in the aerobic degradation of phenylalanine, and the PpcAB phenylphosphate carboxylase enzymes involved in phenol degradation via 4-hydroxybenzoate, all key anaerobic aromatic degradation proteins present in *A. aromaticum* EbN1 are missing from the *D. aromatica* genome (Table 1), and the majority are also not present in *Azoarcus* BH72. The lack of overlap for genes encoding anaerobic aromatic enzymes between these two species was completely unexpected, as both *A. aromaticum* EbN1 and *D. aromatica* are metabolically diverse degraders of aromatic compounds. In general *Azoarcus* BH72 appears to share many families of proteins with *D. aromatica* that are not present in *A. aromaticum* EbN1 (eg signaling proteins, noted below).

Anaerobic degradation of benzene occurs at relatively sluggish reaction rates, indicating that the pathways incumbent in *D. aromatica* for aromatic degradation under anaerobic conditions might serve in a detoxification role. Another intriguing possibility is that oxidation is dependent on intracellularly produced oxygen, which is likely to be a rate-limiting step. *Alicyclophilus denitrificans* strain BC couples benzene degradation under anoxic conditions with chlorate reduction, utilizing the oxygen produced by chlorite dismutase in conjunction with a monooxygenase and subsequent catechol degradation for benzene catabolism [26]. A similar mechanism may account for anaerobic benzene oxidation coupled to perchlorate and chlorate reduction in *D. aromatica*. However, anaerobic benzene degradation coupled with nitrate reduction is also utilized by this organism, and remains enigmatic [5].

The extremely high divergence of encoded protein families in this functional grouping differs from the general population of central metabolic and housekeeping genes: *Azoarcus* BH72, *Azoarcus aromaticum* EbN1 and *D. aromatica* are evolutionarily near-neighbors within currently sequenced genomes, as defined both by the high level of protein similarity within house-keeping genes (defined by the COG J family of proteins), and 16sRNA sequence. *Azo-*

**Table 1: Anaerobic aromatic degradation enzymes in near-neighbor *Aromatoleum aromaticum* EbN1.**

Proteins involved in the anaerobic aromatic pathways in <i>Aromatoleum aromaticum</i> str. EbN1	<i>A. aromaticum</i> EbN1 – representative protein used for HMM models	<i>Azoarcus</i> BH72 ortholog	<i>D. aromatica</i> RCB ortholog
1) phenylalanine			
Pat	VIMSS813888:pat (COG1448; EC 2.6.1.57)	-	-
Pdc	VIMSS817385:pdc (COG3961)	-	-
Pdh	VIMSS816687:pdh (COG1012)	-	-
IorAB	VIMSS813644:iorA (COG4321)	-	+
2) phenylacetate			
PadBCD	VIMSS816693:padB	-	-
PadEFGHI	VIMSS816700:padI	-	-
PadJ	VIMSS816701:padJ	-	-
3) benzyl alcohol/benzaldehyde			
Adh	VIMSS815388:adh (COG1062)	-	-
Ald	VIMSS816847:ald (COG1012; EC1.2.1.28)	+	-
4) p-cresol			
PchCF	VIMSS813733:pchC (EC: 1.17.99.1)	-	-
PchA	VIMSS815385:pchC	-	-
	VIMSS813734:pchF (EC 1.1.3.38)	-	-
	VIMSS815387:pchF	-	-
	VIMSS815384:pchA (COG1012)	-	-
5) phenol			
PpsABC	VIMSS816923:ppsA phenylphosphate synthase	-	-
PpcABCD	VIMSS815367:ppcA	-	-
6) 4-hydroxybenzoate			
PcaK	VIMSS816471:pcaK (COG2271)	-	-
HbcL	VIMSS816681:hbcL1 4-hydroxybenzoate CoA ligase	-	-
HcrCBA	VIMSS815644:hcrB	-	-
	VIMSS815645:hcrA	-	-
7) toluene			
BssDCABEFGH	VIMSS814633:bssA	-	-
BbsABCDEFGHI(IJ)	VIMSS814644:bbsH	-	-
	VIMSS814645:bbsG	-	-
	VIMSS814647:bbsF	-	-
	VIMSS814649:bbsD	-	-
	VIMSS814651:bbsB	-	-
8) ethylbenzene			
EbdABC	VIMSS814907:ebdA	-	+ (PcrA)
Ped	VIMSS814906:ebdB	-	+ (PcrB)
	VIMSS814905:ebdC	-	+ (PcrC)
	VIMSS814904:ebdD	-	+ (PcrD)
	VIMSS814903:ped	-	-
9) benzoate			
BenK	VIMSS816652:benK	-	-
BclA	VIMSS815152:bclA	+	-
BcrCBAD	VIMSS813961:bcrB	-	-
Dch Had Oah	VIMSS813959:bcrA	-	-

**Table 1: Anaerobic aromatic degradation enzymes in near-neighbor *Aromatoleum aromaticum* EbN1. (Continued)**

10) 3-Hydroxybenzoate HbcL	VIMSS813951:hbcL 3-hydroxybenzoate CoA ligase	-	-
BcrADB'C'			

Anaerobic aromatic degradation enzymes in near-neighbor *Aromatoleum aromaticum* EbN1 are largely absent from *Azoarcus* BH72 and *D. aromatica* RCB. Protein profiles (HMMs) were used to detect the presence or absence of anaerobic enzymes involved in degradation of aromatic compounds in *Aromatoleum aromaticum*. The last two columns denote presence (+) or absence (-) of homologs in either the *Azoarcus* BH72 or the *D. aromatica* RCB genome of the proteins listed that are present in the *A. aromaticum* EbN1 genome. The ethylbenzene dehydrogenase molybdenum-containing enzyme complex (ebdABCD) is described by TIGRFams 3479, 3478 and 3482, which define a type II DMSO reductase family of enzymes that includes *D. aromatica*'s perchlorate reductase PcrABCD subunits.

*arcus* BH72 and *A. aromaticum* EbN1 display the highest percent similarity between housekeeping proteins within this triad, with 138 of the 156 COG J proteins in *A. aromaticum* EbN1 displaying highest similarity to their BH72 counterparts. On average these two genomes display 83.5% amino acid identity across shared COG J proteins. *D. aromatica* is an outlier in the triad, with higher similarity to *Azoarcus* BH72 than *A. aromaticum* EbN1 (43 of *D. aromatica*'s 169 COG J proteins are most homologous to *A. aromaticum* EbN1 orthologs with an average 71% identity, and 67 are most homologous to *Azoarcus* BH72 with an average 72% identity).

Comparative genomics have previously established that large amounts of DNA present in one species can be absent even from a different strain within the same species [27]. In addition, the underestimation of the diversity of aromatic catabolic pathways (both aerobic and anaerobic) has been noted previously [28], and a high level of enzymatic diversity has been seen for pathways that have the same starting and end products, including anaerobic benzoate oxidation [29].

#### Aerobic aromatic degradation

*D. aromatica* encodes several aerobic pathways for aromatic degradation, including six groups of oxygenase clusters that each share a high degree of sequence similarity to the phenylpropionate and phenol degradation (Hpp and Mhp) pathways in *Comamonas* species [30,31]. The mhp genes of *E. coli* and *Comamonas* are involved in catechol and protocatechuate pathways for aromatic degradation via hydroxylation, oxidation, and subsequent ring cleavage of the dioxygenated species. Only one of the clusters in *D. aromatica* encodes an mhpA-like gene; it begins with VIMSS584143 MhpC, and is composed of orthologs of MhpABCDEF&R, and is in the same overall order and orientation as the *Comamonas* cluster as well as the *E. coli* mhp gene families [32] (see Fig. 1, cluster 3). These pathways are also phylogenomically related to the biphenyl/polychlorinated biphenyl (Bhp) degradation pathways in *Pseudomonad* species [32]. For *Comamonas testosteroni*, this pathway is thought to be associated with lignin degradation [31]. Hydroxyphenyl propionate (HPP), an alkanolic acid of phenol, is the substrate for Mhp, and is also pro-

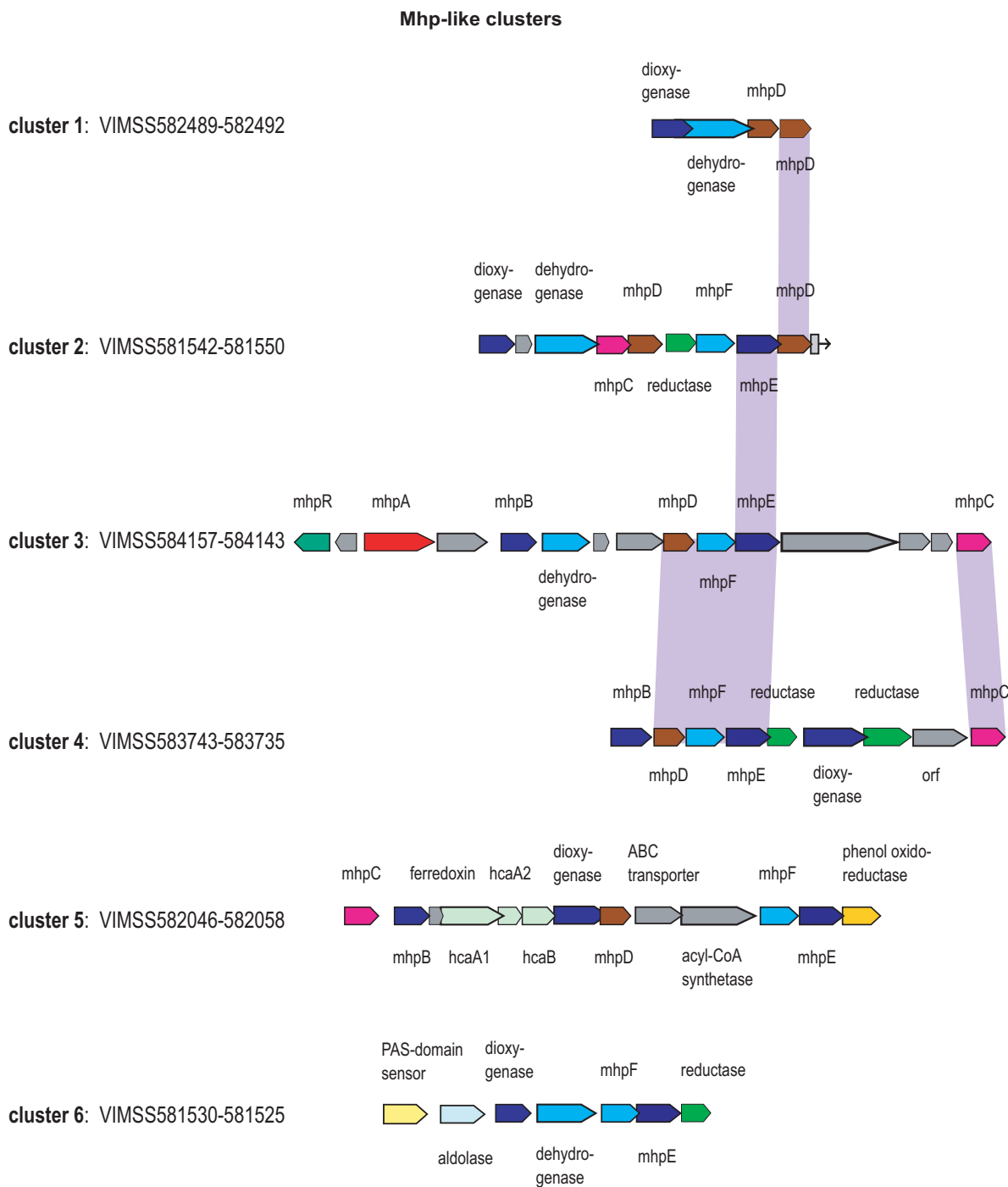
duced by animals in the digestive breakdown of polyphe-nols found in seed components [33]. Each gene cluster appears to represent a multi-component pathway, and is made up of five or more of various combinations of dioxygenase, hydroxylase, aldolase, dehydrogenase, hydratase, decarboxylase and thioesterase enzymes.

The single predicted MhpA protein in *D. aromatica* (VIMSS584155), which is predicted to support an initial hydroxylation of a substituted phenol substrate, shares 64.4% identity to *Rhodococcus* OhpB 3-(2-hydroxyphenyl) propionate monooxygenase (GI:8926385) vs. 26.4% for *Comamonas testosteroni* (GI:5689247), yet the remainder of the ohp genes in the *Rhodococcus ohp* clade do not share synteny with the *D. aromatica* mhp gene cluster.

#### Other aromatic oxygenases

Two chromosomally adjacent monooxygenase clusters, syntenic to genes found in *Burkholderia* and *Ralstonia* spp, indicate that *D. aromatica* might have broad substrate hydroxylases that support the degradation of toluene, vinyl chlorides, and TCE (Fig. 2 and Table 2), and are thus candidates for benzene-activating enzymes in the presence of oxygen.

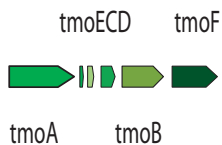
One monooxygenase gene cluster, composed of VIMSS581514 to 581519 ('tbc2 homologs,' Fig. 2), is orthologous to the *tbuA1UBVA2C/tmoAECDBF/touABCDEF/phlKLMNOP* and *tbc2ABCDEF* gene families (from *P. stutzeri*, *R. pickettii*, and *Burkholderia* JS150). This gene cluster includes a transport protein that is orthologous to *TbuX/TodX/XylN* (VIMSS581520). Specificity for the initial monooxygenase is not established, but phylogenetic analysis places VIMSS581514 monooxygenase with near-neighbors *TbhA* [34], reported as a toluene and aliphatic carbohydrate monooxygenase (76.5% sequence identity), and *BmoA* [35], a benzene monooxygenase of low regiospecificity (79.6% sequence identity). The high level of similarity to the *D. aromatica* protein is notable. The region is also highly syntenic with, and homologous to, the *tmoAECDBF* (AY52601) gene cluster responsible for *P. mendocina*'s ability to utilize toluene as a sole carbon and energy source [36].



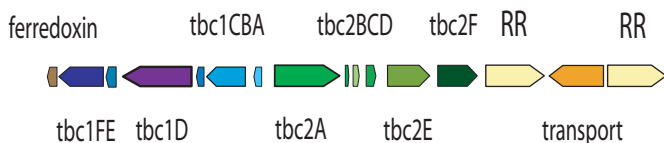
**Figure 1**  
**Aerobic degradation of aromatic compounds: multiple Mhp-like dioxygenase clusters.** Each of the six *mhp*-like gene clusters in the *D. aromatica* genome is depicted. Recent gene duplications between individual proteins are shown by a purple connector between duplicates. Naming convention was chosen for simplicity and consistency, and names all proteins paralogous to a given Mhp protein with the Mhp name (MhpABCDEF or R), but does not imply enzymatic specificity for the substrates listed here-in, though the general enzymatic reaction is highly likely to be conserved. Mhp: *meta* cleavage of hpp, (hydroxyphenyl)propionate. MhpA, 3HPP hydroxylase; MhpB, DHPP 1,2-dioxygenase; MhpC, 2-hydroxy-6-ketono-2,4-dienedioate hydrolase; MhpD, 2, decto-4-pentenoate hydratase; MhpE, 4-hydroxy-2-ketovalerate aldolase; MhpF, acetaldehyde dehydrogenase.

### Monooxygenase clusters

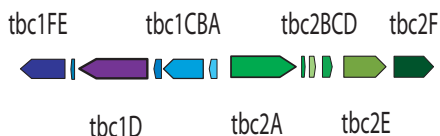
*Pseudomonas mendocina*



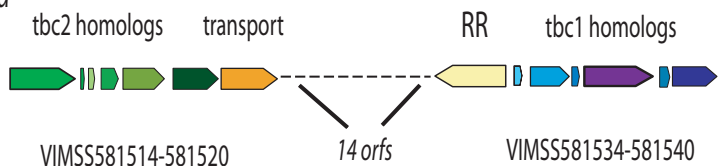
*Ralstonia eutropha* JMP134



*Burkholderia* strain JS150



*Dechloromonas aromatica*



**Figure 2**  
**Catabolic oxygenases of aromatic compounds: Synteny between *D. aromatica*, *P. mendocina*, *Burkholderia* and *R. eutropha*.** Orthologous gene clusters for *P. mendocina*, *R. eutropha* JMP134, *Burkholderia* JS150 and *D. aromatica* are shown. *D. aromatica* possesses two oxygenase gene clusters that are syntenic to the *tbc1* and *tbc2* catabolic gene clusters of *Burkholderia* JS150, but with an inversion and insertion in the chromosome. Also shown are the *tmo* (toluene mono-oxygenase) toluene degradative cluster of *P. mendocina* and the *tbc1* & *tbc2*-like (*tcb*: toluene, chlorobenzene, and benzene utilization) gene cluster of *R. eutropha* (VIMSS 896207–896222, *Burkholderia* protein names were used for consistency). The first seven orfs (encoding a *tbc1*-like cluster) of *R. eutropha* JMP134 are orthologous to the PoxABCDEFGH (phenol hydroxylase) and P0123456 genes of *Ralstonia* sp E2 and *R. eutropha* H16, respectively. Orthologs can be identified as having the same size and color scheme.



**Table 2: Aromatic degradation in *D. aromatica*: Mono- and Di-oxygenases.**

VIMSS id	Orthologs	Putative function	Size aas
581514	TbuA1/TmoA/TouA/PhIK/Tbc2A	methane/phenol/toluene hydroxylase	501
581515	TbuU/TmoE/TouB/PhIL/Tbc2B	toluene-4-monooxygenase	88
581516	TbuB/TmoC/TouC/PhIM/Tbc2C	ferredoxin subunit of ring-hydroxylating dioxygenase	111
581517	TbuV/TmoD/TouD/PhIN/Tbc2C	monooxygenase	146
581518	TbuA2/TmoB/TouE/PhIO/Tbc2E	hydroxylase	328
581519	TbuC/TmoF/TouF/PhIP/Tbc2F	flavodoxin reductase	338
581520	TbuX/TodX/XylIN	membrane protein; transport	464
581521	histidine kinase	signal transduction	963
581522	NarL	cheY like protein	208
581523	methyl-accepting chemotaxis protein	chemotaxis sensory transducer, membrane bound	532
581524	4-oxalocrotonate tautomerase	tautomerase	144
581525	oxidoreductase	oxidoreductase/dehydrogenase	254
581526	MhpE	4-hydroxy-2-oxovalerate aldolase	354
581527	MhpF	EC1.2.1.10 Acetaldehyde dehydrogenase (acetylating)	305
581528	2-hydroxyruconic semialdehyde dehydrogenase	NAD <sup>+</sup> -dependent dehydrogenase (EC1.2.1.60)	489
581529	ring-cleaving extradiol dioxygenase	catechol 2,3 dioxygenase (1.13.11.2)	311
581530	aldolase	4-hydroxyphenylacetic acid catabolism pathway	266
581531	S box domain	signal transduction	143
584293	orf	unknown	63
581532	orf	unknown	80
584294	EAL domain containing protein (obsolete in current VIMSS database)	diguanylate phosphodiesterase; signaling	65
581533	transcriptional regulator	LysR-type	300
581534	response regulator, tbuT family	activator of aromatic catabolism	558
812947	PhcK/DmpK/PhhK/PheA1/Tcb1A/AphK	monooxygenase	89
581535	PhcL/DmpL/PhhL/PheA2/Tcb1B/AphL	hydroxylase	329
581536	PhcM/DmpM/PhhM/PheA3/Tcb1C/AphM	monooxygenase	89
581537	PhcN/DmpN/PhhN/PheA4/Tcb1D	aromatic hydroxylase	517
581538	PhcO/DmpO/PhhO/PheA5/Tcb1E/AphO	aromatic hydroxylase	118
581539	PhcP/DmpP/PhhP/PheA6/Tcb1F/AphQ	hydroxylase reductase	353
581540	ferredoxin	2Fe-2S ferredoxin, iron-sulfur binding site	112
581541	transcriptional regulator	IPR000524: Bacterial regulatory protein GntR, HTH	235
581542	ring-cleaving extradiol dioxygenase	catechol 2,3 dioxygenase (EC1.13.11.2)	308
581543	orf	unknown	142
581544	2-hydroxyruconic semialdehyde dehydrogenase	NAD <sup>+</sup> -dependent dehydrogenase (EC1.2.1.60)	484
581545	MhpC	2-hydroxy-6-ketono-2,4-dienedioic acid hydrolase	274
581546/3337834	MhpD	2-keto-4-pentenoate hydratase	260
581547	oxidoreductase	3-oxoacyl-[acyl-carrier-protein] reductase (EC1.1.1.100)	264
581548	MhpF	acetaldehyde dehydrogenase (acetylating; EC1.2.1.10)	304
581549	MhpE	4-hydroxy-2-oxovalerate aldolase	343
581550	hydratase/decarboxylase	4-oxalocrotonate decarboxylase	262
581551	tautomerase	4-oxalocrotonate tautomerase	63

The large cluster of aromatic degradation enzymes in the *D. aromatica* genome shown includes two mono-oxygenase clusters in a linear array on the *D. aromatica* chromosome, with 17 predicted genes intergenically inserted, which encode *mhp* 'cluster 6' and several predicted signaling proteins. The second monooxygenase cluster is followed by *mhp* 'cluster 2' (for an overview of *mhp* clusters see Fig. 1).

Just downstream on the chromosome is a *phc/dmp/phh/phe/aph*-like cluster of genes, composed of the genes VIMSS812947 and VIMSS 581535 to 581540 ('*tbc1* homologs,' Fig. 2). Overall, chromosomal organization is somewhat different for *D. aromatica* as compared to *Ralstonia* and *Burkholderia*. *D. aromatica* has a fourteen gene insert that encodes members of the *mhp*-like family of aromatic oxygenases between the tandem *tbc 1* and 2-like oxygenase clusters (see Table 2), with an inversion of the second region compared to *R. eutropha* and *Burkholderia*. Clade analysis indicates a broad substrate phenol degra-

ation pathway in this cluster, with high sequence identity to the TOM gene cluster of *Bradyrhizobium*, which has the ability to oxidize dichloroethylene, vinyl chlorides, and TCE [37,38]. The VIMSS581522 response regulator gene that occurs between the two identified monooxygenase gene clusters shares 50.3% identity to the *Thaurea aromatica* *tutB* gene and 48.2% to the *Pseudomonas* sp. Y2 styrene response regulator (occupying the same clade in phylogenetic analysis). VIMSS581522 is likely to be involved in the chemotactic response in conjunction with VIMSS581521 (histidine kinase) and VIMSS581523

(methyl accepting chemotaxis protein), which would confer the ability to display a chemotactic response to aromatic compounds.

Overall, several mono- and di-oxygenases were found in the genome, indicating *D. aromatica* has diverse abilities in the aerobic oxidation of heterocyclic compounds.

There are several gene clusters indicative of benzoate transport and catabolism. All recognized pathways are aerobic. The benzoate dioxygenase cluster BenABCDR is encoded in VIMSS582483-582487, and is very similar to (and clades with) the xylene degradation (*xyXYZ*) cluster of *Pseudomonas*.

There is also an *hcaA* oxygenase gene cluster, embedded in one of the *mhp* clusters (see cluster 5, Fig. 1). Specificity of the large subunit of the dioxygenase (VIMSS582049) appears to be most likely for a bicyclic aromatic compound, as it shows highest identity to dibenzothiophene and naphthalene dioxygenases.

**Dechloromonas aromatica's sensitivity to the environment**

*Cell Signaling*

*D. aromatica* has a large number of genes involved in signaling pathways, with 314 predicted signaling proteins categorized in COG T (signal transduction mechanisms) and a total of 395 proteins (nearly 10% of the genome) either recruited to COG T or possessing annotated signal transduction domains. Signaling appears to be an area that has undergone recent gene expansion, as nine recent gene duplication events in this functional group are pre-

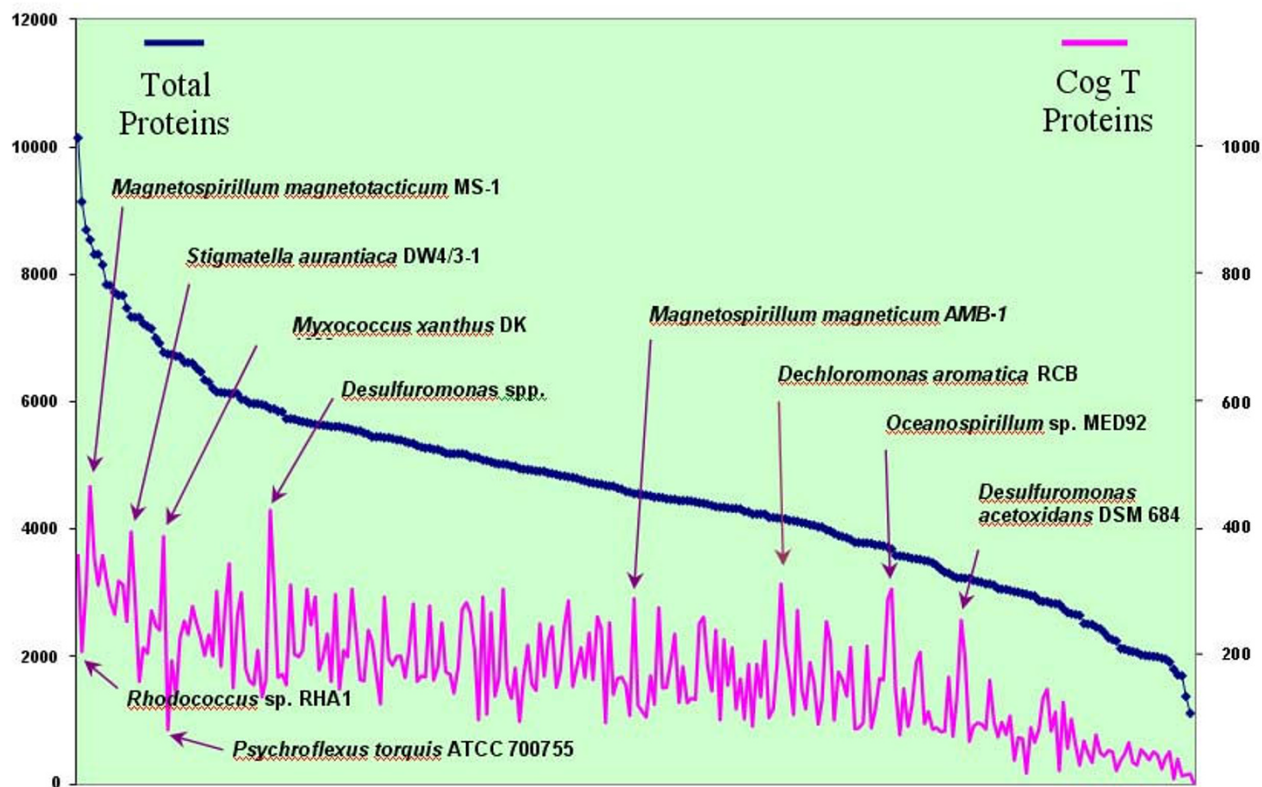
dicted by phylogenetic analysis, as described in a later section (shown in Table 3).

Complex lifestyles are implicated in large genomes with diverse signaling capability, and in general genomes with a very large number of annotated open reading frames (orfs) have high numbers of predicted signal transducing proteins, as shown in Fig. 3, though some species, such as *Rhodococcus* RHA1 and *Psychroflexus torques* are notable exceptions to this trend. However, assessment of COG T population size relative to other genomes with a similar number of predicted orfs (Fig. 3) indicates that *D. aromatica* is one of a handful of species that have a large relative number of signaling proteins vs similarly sized genomes. Other organisms displaying this characteristic include *Magnetospirillum magnetotacticum* MS-1, *Stigmatella aurantiaca*, *Myxococcus Xanthus* DK1622, *Magnetospirillum magneticum* AMB-1, *Oceanospirillum* sp. MED92, and *Desulfuromonas acetoxidans*. Within the Betaproteobacteria, *Chromobacterium violaceum* and *Thiobacillus denitrificans* have a relatively large number of signaling cascade genes, but still have far fewer than found in *D. aromatica*, with 262 predicted COG T proteins (6% of the genome) and 137 COG T proteins (4.8% of the genome), respectively. Histidine kinase encoding proteins are particularly well-represented, with only *Stigmatella aurantiaca* DW4/3-1, *Magnetococcus* sp. MC-1, *Myxococcus xanthus* DK 1622, and *Nostoc punctiforme* reported as having more. The sixty-eight annotated histidine kinases include a large number of nitrate/nitrogen responsive elements. Furthermore, the presence of 47 putative histidine kinases predicted to contain two transmembrane (TM) domains, likely to encode membrane-bound sensors (see Fig. 4), suggests that *D.*

**Table 3: Candidates for gene expansion in the *D. aromatica* genome.**

Protein/protein family function	Number of duplicates	Number of triplicates
Transport (membrane)	12	
Signal transduction or regulatory – includes:	9	
FlhD homolog	(1)	
FlhC homolog	(1)	
Nitrogen regulatory protein PII homolog	(1)	
Hydrolase/transhydrogenase or hydratase	4	1
Cytochromes	3	2
Mhp family	2	2
Phospholipase/phosphohydrolase	2	1
Phasin	1	
Dioxygenase	1	
NapH homolog	1	
NosZ homolog	1	
Unknown function	7	

Proteins within the genome that show evidence of possible recent gene duplication are tabulated by general functional group, or, in some cases, specific proteins (Naph, NosZ, FlhCD, Nitrogen regulatory protein PII). Duplicates and triplicates were determined by adjacent clustering of the *D. aromatica* proteins in a phylogenomic tree profile. The percent identity between the *D. aromatica* duplicate and triplicate candidates is higher than identity to other species' protein candidates, indicating a possible gene family expansion event. Areas of duplicated clusters of proteins (for instance, the regions surrounding VIMSS582581, 582612, 582641, 582657, 582863, 583914 and 583592), including phage elements and Tra-type conjugation proteins, are not included in this table. Parentheses indicate these duplication events have been tabulated in the general category of signal transduction or regulatory proteins – individual protein types of particular interest are noted separately by protein name.



**Figure 3**

**Number of predicted signaling proteins versus total protein count.** Microbial genomes, displaying total number of predicted open reading frames (orfs, left axis) and total number of predicted signaling proteins (defined as COG T, right axis). Microbes displaying a high number of signaling orfs relative to total predicted proteins are labelled (above COG T line), as well as two large-sized genomes having a relatively low number of annotated COG T proteins (labelled below COG T line).

*aromatica* is likely to be highly sensitive to environmental signals. Nearly half (48%) of the predicted histidine kinases are contiguous to a putative response regulator on the chromosomal DNA, indicating they likely constitute functionally expressed kinase/response regulator pairs. This is atypically high for contiguous placement on the chromosome [39].

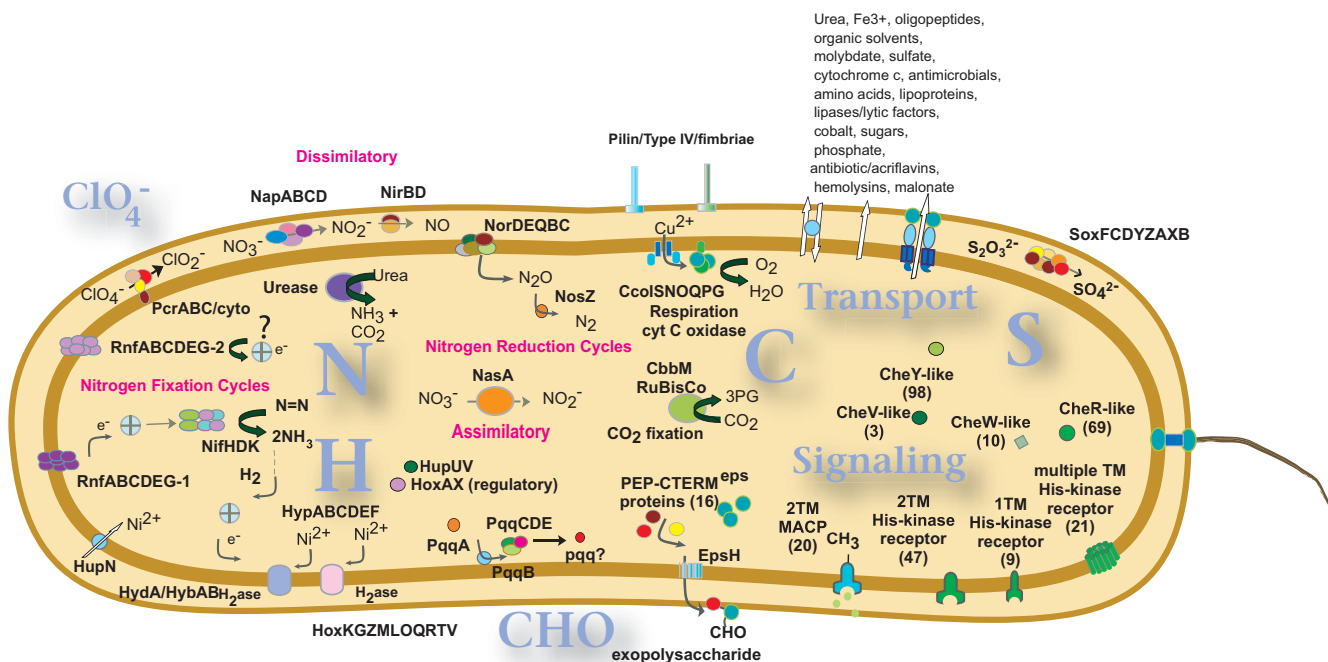
A relatively high level of diguanylate cyclase (GGDEF domain [40-42]) signaling capability is implied in *D. aromatica* by the presence of 57 proteins encoding a GGDEF domain (Interpro IPR000160 [see Additional file 5]) and an additional 10 with a GGDEF response regulator (COG1639) [40]. *E. coli*, for comparison, encodes 19. This gene family also appears to have undergone recent expansion in this microbe's evolutionary history. Microbes having a large number of proteins or even a diverse array of COG T elements do not *a priori* encode a large number of GGDEF elements, as *Stigmatella aurantiaca*, *Myxococcus*, *Xanthus* DK1622 and *Burkholderia pseudomallei*, by contrast, have very large genomes with extensive COG T pop-

ulations, yet each have 20 or fewer proteins identified as having GGDEF domains [see Additional file 5], and *Prochlorococcus* spp. appear to have none. Conversely, *Oceanospirillum* has a relatively small genome, yet has 112 proteins identified as likely GGDEF domain/IPR000160 proteins. GGDEF/EAL domain response regulators have been implicated in root colonization in *Pseudomonas putida* (Matilla et al. 2007); in *E. coli* the GGDEF domain-containing YddV protein upregulates the transcription of a number of cell wall modification enzymes [42], and in point of fact, *D. aromatica*'s VIMSS581804, a GGDEF domain containing homolog of the YddV *E. coli* protein, occurs upstream of a cluster of sixteen cell wall division proteins (encoded by VIMSS581805-581820).

#### **Cellular interactions with community/environment – secretion**

##### *Typel* secretion

Fifteen transport clusters include a TolC-like outer membrane component, and recent gene family expansion is noted within several families of ABC transporters for this

**Figure 4**

**Overview of predicted metabolic cycles, membrane transporters and signaling proteins in *D. aromatica*.** Various metabolic cycles, secretory apparatus and signaling cascades predicted in the annotation process are depicted. TM: transmembrane. Gene names are discussed in the relevant sections of this paper. Areas of the cell depicting Nitrogen, Hydrogen, Carbon and Sulfur cycles are indicated by "N," "H," "C," and "S."

genome. TolC was originally identified in *E. coli* as the channel that exports hemolysin [43], and hemolysin-like proteins are encoded in this genome. Two groups of ABC transporters occur as a cluster of five transport genes; these five-component transporters have been implicated in the uptake of external macromolecules [44].

The presence of putative lytic factors, lipases, proteases, antimicrobials, invasins, hemolysins, RTXs and colicins near potential type I transport systems indicate that these might be effector molecules used by *D. aromatica* for interactions with host cells (eg. for cell wall remodeling). Iron acquisition is likely to be supported by a putative FeoAB protein cluster (VIMSS583997, 583998), as well as several siderophore-like receptors and a putative FhuE protein (outer membrane receptor for ferric iron uptake; VIMSS583312). Other effector-type proteins, likely to be involved in cell/host interactions (and which in some species have a role in pathogenicity [45]), are present in this genome. Adhesins, haemagglutinins, and oxidative stress neutralizers are relatively abundant in *D. aromatica*. A number of transporters occur near the six putative soluble lytic murein transglycosylases, indicating possible cell wall remodeling capabilities for host colonization in conjunction with the potential effector molecules noted

above. Homologs of these transporters were shown to support invasion-type functions in other microbes [45]. Interaction with a host is further implicated by: VIMSS581582, encoding a potential cell wall-associated hydrolase, VIMSS581622, encoding a predicted ATPase, and VIMSS3337824/formerly 581623, encoding a putative membrane-bound lytic transglycosylase.

Eleven tandem copies of a 672 nucleotide insert comprise a region of the chromosome that challenged the correct assembly of the genome, and finishing this region was the final step for the sequencing phase of this project (see Methods). Unexpectedly, analysis of this region revealed a potential open reading frame encoding a very large protein that has been variously predicted at 4854, 2519 or 2491 amino acids in size during sequential automated protein prediction analyses (VIMSS3337779/formerly 582095). This putative protein, even in its smallest configuration, contains a hemolysin-type calcium-binding region, a cadherin-like domain, and several RTX domains, which have been associated with adhesion and virulence. Internal repeats of up to 100 residues with multiple copies have also been found in proteins from *Vibrio*, *Colwellia*, *Bradyrhizobium*, and *Shewanella* spp. (termed "VCBS" proteins as defined by TIGRfam1965).

Other potential effector proteins include: three hemolysin-like proteins adjacent to type I transporters, eight proteins with a predicted hemolysin-related function, including VIMSS583067, a hemolysin activation/secretion protein, VIMSS580979, hemolysin A, VIMSS583372, phospholipase/hemolysin, VIMSS581868, a homolog of hemolysin III, predicted by TIGRfam1065 to have cytolytic capability, VIMSS582079, a transport/hemolysin, and VIMSS581408, a general hemolysin. Five predicted proteins have possible LysM/invasin domains, including: VIMSS580547, 581221, 581781, 582766, and 583769. One gene, VIMSS583068, encodes a putative 2079 amino acid filamentous haemagglutinin, as well as a hasA-like domain, making it a candidate for hasA-like function (hasA is a hemophore that captures heme for iron acquisition [46]).

#### Type II secretion

Besides the constitutive Sec and Tat pathways, *D. aromatica* has several candidates for dedicated export secretions of unknown function, with 3–4 putative orthologs of PulDEFG interspersed with a lytic transglycosylase and a hemolysin (VIMSS582071-582085). The region from VIMSS581889 to VIMSS581897 includes *pulDEFG* type subunits and an *exeA* ATPase like protein. It is bracketed by signaling components comprised of a histidine kinase, adenylate cyclase, and a protein bearing similarity to the nitrogen response regulator *glnG* (VIMSS581898), which has been shown to be involved in NH<sub>3</sub> assimilation in other species [47].

In addition, there is a nine-gene cluster that encodes several proteins related to toluene resistance (VIMSS581899 to 581906).

A pilus-like gene cluster (which can also be classified as type IV secretion) occurs in VIMSS580547-580553, encoding a putative lytic transglycosylase, ABC permease, cation transporter, pilin peptidase, pilin ATPase and PulF-type protein. This assembly resembles other pilin assemblies associated with attachment to a substrate, such as the pilus structure responsible for chitin/host colonization in *Vibrio cholerae* [48].

Another large pilus-like cluster (VIMSS584160-584173) occurs in close proximity to the *mhp*CEFD BAR oxygenase genes (see eg VIMSS584157, *mhpR*).

#### Type III secretion

*D. aromatica* has been shown to be chemotactic under various circumstances. The flagellar proteins (FlIAEFGHIJKLMNOPQR, FlaABCDEFGHIJK) are followed by an additional cluster of 15 chemotaxis/signal transduction genes (VIMSS580462-580476), and homologs of FlhC and D regulatory elements required for the expression of

flagellar proteins (VIMSS582640 and 582641) [49], identified by phylogenetic clustering, are also present. Since *D. aromatica* has a flagellum and displays chemotactic behavior, it is likely that the flagellar gene cluster is solely related to locomotion, though type III secretion systems can also encode dedicated protein translocation machineries that deliver bacterial pathogenicity proteins directly to the cytosol of eukaryotic host cells [50].

#### Type IV secretion

There are two copies of a twenty-one gene cluster that includes ten putative conjugal transfer (Tra) sex-pilus type genes in the *D. aromatica* genome (VIMSS582582-582601 and VIMSS582864-582884), indicating a typeIV secretion structure that is related to non-pathogenic cell-cell interactions [51].

#### Type VI secretion

A large cluster of transport proteins that is related to the virulence associated genetic locus HIS-1 of *Pseudomonas aeruginosa* and the VAS genes of *V. cholerae* [52,53] includes homologs of hcp1, IcmF and clpV (as VIMSS583005, 582995 and 583009, respectively, in *D. aromatica* [see Additional file 6]). This IcmF-associated (IAHP) cluster has been associated with mediation of host interactions, via export of effector proteins that lack signal sequences [53]. Further evidence for type VI secretion is found in the presence of three proteins containing a Vgr secretion motif modeled by TIGRfam3361, which is found only in genomes having type VI secretory apparatus. Though most bacteria that contain IcmF clusters are pathogenic agents that associate with eukaryotic cell hosts [54], it has been reported that the host interactions supported by this cluster are not restricted to pathogens [55].

The type IV pili systems might be involved in biofilm development, as interactions with biofilm surfaces are affected by force-generating motility structures, including type IV pili and flagella [56]. Quorum sensing is a deciding input for biofilm formation, and the presence of an exopolysaccharide synthetic cluster lends further support for biofilm formation. Further, derivatives of nitrous oxide, which is an evident substrate for *D. aromatica*, are a key signal for biofilm formation vs cell dispersion in the microbe *P. aeruginosa* [57].

#### Cellular interactions with community – quorum sensing

Quorum sensing uses specific membrane bound receptors to detect autoinducers released into the environment. It is involved in both intra- and inter-species density detection [58,59]. Cell density has been shown to regulate a number of cellular responses, including bioluminescence, swarming, expression of virulence factors, secretion, and motility (as reviewed in Withers et al. 2001 [60]).

*D. aromatica* encodes six histidine kinase receptor proteins that are similar to the quorum sensing protein QseC of *E. coli* (VIMSS580745, 582451, 582897, 583274, 3337577 (formerly 583538), and 583893), five of which co-occur on the chromosome with homologs of the CheY like QseB regulator, and two of which appear to be the product of a recent duplication event (VIMSS583893 & 3337577). Of the six QseC homologs, phylogenetic analysis indicates VIMSS582451 is most similar to QseC from *E. coli*, where the QseBC complex regulates motility via the FlhCD master flagellar regulators (VIMSS582640 and 582641). The presence of several *qseC/B* gene pairs indicates the possibility of specific responses that are dependant on different sensing strategies. In other species, expression of ABC exporters is regulated by quorum sensing systems [46]; gene family expansion is indicated in the ABC export gene pool as well as the *qseC/B* sensors in *D. aromatica*.

N-acyl-homoserine lactone is the autoinducer typical for gram negative bacteria [61], yet *D. aromatica* lacks any recognizable AHL synthesis genes. *Ralstonia* Betaproteobacteria likewise encode several proteins in the *qseC* gene family and display a diversity of candidate cell density signaling compounds other than AHL [62]. The utility of having a diverse array of quorum sensing proteins remains to be determined, but appears likely to be associated with a complex, and possibly symbiotic, lifestyle for *D. aromatica*.

#### Cellular interactions with the environment – stress

##### Carbon Storage

Poly-hydroxyalkanoates (PHAs) store carbon energy, are synthesized from the catabolism of lipids, and constitute up to almost 90% of the dry weight of the Betaproteobacteria species *Comamonas testosteroni* [63]. These lipid-like carbon/energy storage polymers are found in granular inclusions. PhaR candidate VIMSS583509 is likely to be a regulatory protein for PHA synthesis, and is found near other proteins associated with PHA granule biosynthesis and utilization in *D. aromatica* (VIMSS583511-513).

Phasins are relatively small proteins (180–200 aas) that have been shown to associate with PHA inclusions [64]. There are six copies of phasin-type proteins, with indications of recent gene duplication for three of the phasin-type proteins (VIMSS581881, 582264, and 3337571 (formerly 583582)). There are also three homologs of the active subunit poly-B-hydroxybutyrate polymerase (PhaC orthologs) and two Pha reductase candidates present in a direct repeat, which is also found in *Legionella pneumophila*. Interestingly, one PhaC-like protein, VIMSS583511, is 70% identical to NodG of *Azospirillum brasilense*, a nodulation protein [65]. No PhaA-like ketothiolase ortholog is present. The presence of an amplified gene pool for carbon storage granules in *D. aromatica* may confer the ability to survive under low nutrient

conditions, and poly-3-hydroxybutyrate accumulation has recently been observed in *A. aromaticum* EbN1 cultures displaying reduced growth [66].

##### Phosphate

Inorganic polyphosphate storage appears likely, as both polyphosphate kinase (Ppk, VIMSS582444) and exopolyphosphatase (Ppx, VIMSS583870) are present. These genes are similar to those encoded in *Pseudomonas aeruginosa*, in that they are in disparate regions of the chromosome [67]. Polyphosphate has been implicated in stress response due to low nutrients in the environment [68], and also in DNA uptake [69].

Phosphate transport appears to be encoded in a large cluster of genes (VIMSS581746-581752), and response to phosphate starvation is likely supported by the PhoH homolog VIMSS583854.

##### Biofilm formation

There is a large cluster of exopolysaccharide export (*eps*) associated genes, including a proposed exosortase (*epsH*, VIMSS582792). Presence of the *eps* family proteins (VIMSS582786, VIMSS582790-582801) indicates capsular exopolysaccharide production, associated with either host cell interactions (including root colonization [70]) or biofilm production in soil sediments [71]. *D. aromatica* is one of a small number of species (19 out of 280 genomes assessed by Haft et al. [71]) that also encodes the PEP-CTERM export system. The PEP-CTERM signal, present in sixteen proteins in this genome, is proposed to be exported via a potential exportase, represented in this genome by *epsH* (VIMSS582792). Additionally, the presence of proteins encoding this putative exportase is seen only in genomes also encoding the *eps* genes.

##### Metabolic Cycles

##### Nitrogen

*D. aromatica* closely reflects several metabolic pathways of *R. capsulatus*, which is present in the rhizosphere, and its assimilatory nitrate/nitrite reductase cluster is highly similar to the *R. capsulatus* cluster [72]. Encoded nitrate response elements also indicate a possible plant association for this microbe, as nitrate can act as a terminal electron acceptor in the oxygen-limited rhizosphere. Alternatively, nitrous oxide (NO) reduction can indicate the ability to respond to anti-microbial NO production by a host (used by the host to mitigate infection [73]). Several gene families are present that indicate interactions with a eukaryotic host species, including response elements that potentially neutralize host defense molecules, in particular nitric oxide and other nitrogenous species.

Nitrate is imported into the cytosol by NasDEF in *Klebsiella pneumoniae* [74] and expression of nitrate and nitrite reductases is regulated by the *nasT* protein in *Azotobacter*

*vinelandii* [75]. A homologous set of these genes are encoded by the cluster VIMSS580377-580380 (Nas-DEFT), and a homolog of *narK* is immediately downstream at VIMSS580384, and is likely involved in nitrite extrusion. Upstream, a putative *nasA/nirBDC* cluster (assimilatory nitrate and nitrite reduction) is encoded near the *narXL*-like nitrate response element. VIMSS580393 encodes a nitrate reductase that is homologous to the NasA cytosolic nitrate reductase of *Klebsiella pneumoniae* [76]. Community studies have correlated the presence of NasA-encoding bacteria with the ability to use nitrate as the sole source of nitrogen [77]. The large and small subunits of nitrite reductase (VIMSS580391 *nirB* and VIMSS580390 *nirD*) are immediately adjacent to a transporter with a putative nitrite transport function (VIMSS580389 NirC-like protein). The NirB orf is also highly homologous to both NasB (nitrite reductase) and NasC (NADH reductase which passes electrons to NasA) of *Klebsiella pneumoniae*. HMMs created from alignments seeded by the NasB and NasC genes scored at  $3.2e^{-193}$  and  $4.0e^{-159}$ , respectively, to the VIMSS580391 NirB protein. *D. aromatica* is similar to *Methylococcus capsulatus*, *Ralstonia solanacearum*, *Polaromonas*, and *Rhodospirillum rubrum* for *nasA*, *nirB* and *nirD* gene clusters. However, the presence of the putative transporter *nirC* (VIMSS580389) shares unique similarity to the *E. coli* and *Salmonella nirBCD* clusters.

Putative periplasmic, dissimilatory nitrate reduction, which is a candidate for denitrification capability [78], is encoded by the *napDABC* genes (VIMSS 3337807/581796-581799). A probable cytochrome *c'*, implicated in nitric oxide binding as protection against potentially toxic excess NO generated during nitrite reduction [79], is encoded by VIMSS582015. Although most denitrifiers are free living, plant-associated denitrifiers do exist [80]. There is no dissimilatory nitrate reductive complex *narGHIJ*, but rather, NarG and NarH-like proteins are found in the evolutionarily-related perchlorate reductase alpha and beta subunits [24]. These proteins are present in the *pcrABCDcld* cluster, VIMSS582649-582652 and VIMSS584327, as previously reported for *Dechloromonas* species [81].

Ammonia incorporation appears to be metabolically feasible via a putative glu-ammonia ligase (VIMSS581081), an enzyme that incorporates free ammonia into the cell via ligation to a glutamic acid. An ammonium transporter and cognate regulator are likely encoded in the Amt and GlnK-like proteins VIMSS581101 and 581102.

Urea catabolism as a further source of nitrogen is suggested by two different urea degradation enzyme clusters. The first co-occurs with a urea ABC-transport system, just upstream of a putative nickel-dependent urea amidohy-

drolase (urease) enzyme cluster (VIMSS583666, 583671-583674, and VIMSS583677-583683; see Table 4). The second pathway is suggested by a cluster of urea carboxylase/allophanate hydrolase enzymes (VIMSS581083-581085, described by TIGRFams 1891, 2712, 2713, 3424 and 3425), which comprise four proteins involved in urea degradation to ammonia and carbon dioxide in other species, as well as an amidohydrolase [82].

#### Nitric oxide (NO) reductase

The chromosomal region around *D. aromatica*'s two *nosZ* homologs is notably different from near-neighbors *A. aromaticum* EbN1 and *Ralstonia solanacearum* which encode a *nosRZDFYL* cluster. *D. aromatica*'s *nosRZDFYL* operon lacks the *nosRFYL* genes, and displays other notable differences with most nitrate reducing microbes. In *D. aromatica*, two identical *nosZ* reductase-like genes (annotated as *nosZ1* and *nosZ2*, VIMSS583543 and VIMSS583547) are adjacent to two cytochrome *c553s*, a ferredoxin, and a transport accessory protein, and are uniquely embedded within a histidine kinase/response regulator cluster and include *nosD* and a *napGH*-like pair that potentially couples quinone oxidation to cytochrome *c* reduction. This indicates the NO response might be involved in cell signaling and as a possible general detoxification mechanism for nitric oxide.

The Epsilonproteobacteria *Wolinella succinogenes* is quite similar to *D. aromatica* for nitric oxide reductase genes (both have two *nosZ* genes, a *nosD* gene and a *napGH* pair in the same order and orientation [83]), but the *W. succinogenes* genome lacks the embedded signaling protein cluster. Further, nitric oxide reductase homologs NOR-DEBC (VIMSS582097, 582100-582103), along with the cytochrome *c'* protein (VIMSS582015), which has been shown to bind nitric oxide (NO) prior to its reduction [79], are all present, and potentially act in detoxification roles. It has been shown that formation of anaerobic biofilms of *P. aeruginosa* (which cause chronic lung infections in cystic fibrosis) require NO reductase when quorum has been reached [84], so a role in signaling and complex cell behavior is possible.

*W. succinogenes* shares other genome features with *D. aromatica*. It encodes only 2042 orfs, yet has a large number of signaling proteins, histidine kinases, and GGDEF proteins relative to its genome size. It also encodes *nif* genes, several genes similar to virulence factors, and similarity in the nitrous oxide enzyme cluster noted above. *W. succinogenes* is evolutionarily related to two pathogenic species (*Helicobacter pylori* and *Campylobacter jejuni*), and displays eukaryotic host interactions, yet is not known to be pathogenic [85]. The distinction between effector molecules causing a pathogenic interaction and a symbiotic one is unclear.

**Table 4: Putative nitrogen fixation gene cluster in *D. aromatica***

VIMSS id	Ortholog	Size, aas
583652	FldA, flavodoxin typical for nitrogen fixation	186
583653	hypothetical protein	86
583654	NafY-1, nitrogenase accessory factor Y	247
583655	NifB, nitrogenase cofactor biosynthesis protein	500
583656	4Fe-4S ferredoxin	92
583657	nitrogenase-associated protein	159
583658	flavodoxin	423
583659	ferredoxin, nitric oxide synthase	95
583660	2Fe-2S ferredoxin	120
583661	NifQ	190
583662	DraG	326
583663	histidine kinase	1131
583664	Che-Y like receiver	308
583666	UrtA urea transport	420
583667/3337562	CynS cyanate lyase	147
583668	S-box sensor, similar to oxygen sensor arcB	794
583669	ABC transporter	393
3337561	Protein of unknown function involved in nitrogen fixation	72
583671	UrtB urea transport	525
583672	UrtC urea transport	371
583673	UrtD urea transport	278
583674	UrtE urea transport	230
583677	UreH urease accessory protein	288
583678	Urea amidohydrolase gamma	100
583679	Urea amidohydrolase beta	101
583680	Urea amidohydrolase alpha/UreC urease accessory protein	569
583681	UreE urease accessory protein	175
583682	UreF urease accessory protein	228
583683	UreG urease accessory protein	201
583685	nitroreductase	558
583686	ferredoxin, subunit of nitrite reductase	122
583691	DraT	328
583692	NifH nitrogenase iron protein (EC1.18.6.1)	296
583693	NifD nitrogenase molybdenum-iron protein alpha chain (EC1.18.6.1)	490
583694	NifK nitrogenase molybdenum-iron protein beta chain (EC1.18.6.1)	522
583695/3337559	NifT	80
3337558	ferredoxin	63
583696	NafY-2 nitrogenase accessory factor Y	243
583710/3337556	NifW nitrogen fixation protein	113
3337555	NifZ	151
583711/3337554	NifM	271

The annotated nitrogen fixation homologs (Nif proteins) are embedded with a cluster of urea transport and degradation genes (Ure, Urea amidohydrolase and Urt transport families).

#### Nitrogen Fixation

Nitrogen fixation capability in *D. aromatica* is indicated by a complex of *nif*-like genes (see Table 4), that include putative nitrogenase alpha (NifD, VIMSS583693) and beta (NifK, VIMSS583694) subunits of the molybdenum-iron protein, an ATP-binding iron-sulfur protein (NifH, VIMSS583692), and the regulatory protein NifL (VIMSS583623), that share significant sequence similarity and synteny to the free-living soil microbe *Azotobacter vinelandii*. *D. aromatica* further encodes a complex that is likely to transport electrons to the nitrogenase, by using a six subunit *mf*ABCDGE-like cluster (VIMSS583616-

583619, 583621 and 583622) that is phylogenomically related to the *Rhodobacter capsulatus* complex used for nitrogen fixation [86]. There is a second *mf*-like NADH oxidoreductase complex composed of VIMSS583911-583916, of unknown involvement (see Fig. 4). *A. aromaticum* EbN1 and *Azoarcus* BH72 each encode two *mf*-like clusters as well.

Embedded in the putative nitrogen fixation cluster are two gene families involved in urea metabolism (Table 4). This includes the urea transport proteins (UrtABCDE) and urea hydrolase enzyme family (Ure protein family).



*Hydrogenases associated with nitrogen fixation*

Uptake hydrogenase is involved in the nitrogen fixation cycle in root nodule symbionts where it is thought to increase efficiency via oxidation of the co-produced hydrogen (H<sub>2</sub>) [87]. *D. aromatica* encodes a cluster of 13 predicted orfs (Hydrogenase-1 cluster, VIMSS581358-581370; Table 5) that includes a hydrogenase cluster syntenic to the *hox*KGZMLQQR(T)V genes found in *Azotobacter vinelandii*, which reversibly oxidize H<sub>2</sub> in that organism [88]. This cluster is followed by a second hydrogenase (Hydrogenase-2 cluster, VIMSS581373-581383). The hydrogenase assembly proteins, *hyp*ABF and CDE are included (VIMSS581368-581370 and 581380-581381, and VIMSS3337851 (formerly 581382)) as well as proteins related to the hydrogen uptake (*hup*) genes of various rhizobial microbes [87]. The second region, with the *hyp* and *hyd*-like clusters, lacks overall synteny to any one genome currently sequenced. It does, however, display regions of genes that share synteny with *Rhodospirillum rubrum*, which displays the highest percent identity across the cluster, both in terms of synteny and protein identity.

VIMSS581384 encodes a homolog of the HoxA hydrogenase transcriptional regulator, which has been shown to be expressed only during symbiosis in some species [89]. Regulation is indicated by homologs of NtrX (VIMSS581123) and NtrY (VIMSS581124); the NtrXY pathway comprises a two-component signaling system involved in the regulation of nitrogen fixation in *Azorhizobium caulinodans* ORS571 [90].

*Carbon Fixation via the Calvin-Benson-Bassham cycle*

The genes indicative of carbon fixation, using the Calvin cycle, are present in the *D. aromatica* genome. This includes Ribulose 1,5-bisphosphate carboxylase (RuBisCo, VIMSS581681), phosphoribulokinase (*cbbP*/PrkB, VIMSS581690), and a fructose bisphosphate (*fba*, VIMSS581693) of the Calvin cycle subtype. The RuBisCo *cbbM* gene is of the fairly rare type II form. *D. aromatica* CbbM displays a surprisingly high 77% amino acid identity to CbbM found in the deep-sea tube worm *Riftia pachyptila* symbiont [91]. In a recent study of aquatic sediments, *Rhodospirillum rubrum*, *Rhodospirillum rubrum*

**Table 5: Hydrogenase clusters associated with nitrogen fixation.**

VIMSS id	Orthologs	Putative function	Size, aas
581358	HoxK/HyaA/HupS	hydrogenase-I small subunit	363
581359	HoxG/HyaB/HupL	hydrogenase-I, nickel-dependent, large subunit	598
581360	HoxZ/HyaC/HupC	Ni/Fe-hydrogenase I b-type cytochrome subunit	234
581361	HoxM/HyaD/HupD	hydrogenase expression/formation protein	204
581362	HoxL/HypC/HupF	hydrogenase assembly chaperone	100
581363	HoxO/HyaE/HupG	hydrogenase-I expression	152
581364	HoxQ/HyaF/HupH	nickel incorporation into hydrogenase-I proteins	287
581365	HoxR/HupI	rubredoxin-type Fe(Cys) <sub>4</sub> protein	66
581366	HupJ/(similar to HoxT)	hydrogenase accessory protein	156
581367	HoxV/HupV	membrane-bound hydrogenase accessory protein	308
581368	HypA	hydrogenase nickel insertion protein	113
581369	HypB	hydrogenase accessory factor Ni(2+)-binding GTPase	352
581370	HypF	hydrogenase maturation protein	763
581371	ABC protein	periplasmic component, ABC transporter	260
581372	GGDEF domain	signal transduction, GGDEF	523
581373	Hyb0	hydrogenase-2 small subunit	394
581374	HybA	Fe-S-cluster-containing hydrogenase component	351
581375	HybB	cytochrome Ni/Fe component of hydrogenase-2	386
581376	HybC/HynA	hydrogenase-2 large subunit	570
581377	HybD/HynC	Ni, Fe-hydrogenase maturation factor	159
581378	HupF/HypC	hydrogenase assembly chaperone	96
581379	HybE/HupJ	hydrogenase accessory protein	183
581380	HypC	hydrogenase maturation protein	81
581381	HypD	hydrogenase maturation protein	374
581382/3337851	HypE	hydrogenase maturation protein	330
581383	HoxX/HypX	formation of active hydrogenase	558
581384	HoxA	response regulator with CheY domain (signal transduction)	495
581385	HoxB/HupU	regulatory [NiFe] Hydrogenase small subunit (sensor)	333
581386	HoxC/HupV	regulatory [NiFe] Hydrogenase large subunit (sensor)	472
581397	HupT	histidine kinase with PAS domain sensor	448
581398	HoxN/HupN/NixA	nickel transporter	269

A large region of putative hydrogenases associated with nitrogen fixation is noted. Hydrogenase-I is composed of *hox/hup* genes, hydrogenase-2 of the *hyb* genes, and the nickel insertion/maturation complex, *hyp* (present in two clusters: *hyp*ABF and *hyp*CDE).

and *R. rubrum* were also found to possess the *cbbM* type II isoform of RuBisCo [92]; this sub-type is shared by a only a few microbial species.

Further putative Cbb proteins are encoded by VIMSS581680 & 581688, candidates for CbbR (regulator for the *cbb* operon) and CbbY (found downstream of RuBisCo in *R. sphaeroides* [93]), respectively.

The presence of the *cbbM* gene suggests the ability to carry out the energetically costly fixation of CO<sub>2</sub>, though such functionality has yet to be observed, and carbon dioxide fixation capability has been found in only a few members of the microbial community.

There is a potential glycolate salvage pathway indicated by the presence of two isoforms of phosphoglycolate phosphatase (*gph*, VIMSS583850 and 581830). In other organisms, phosphoglycolate results from the oxidase activity of RuBisCo in the Calvin cycle, when concentrations of carbon dioxide are low relative to oxygen. In *Ralstonia (Alcaligenes) eutropha* and *Rhodobacter sphaeroides*, the *gph* gene (*cbbZ*) is located on an operon along with other Calvin cycle enzymes, including RuBisCo. In *D. aromatica*, the *gph* candidates for this gene (VIMSS583850 and 581830), are removed from the other *cbb* genes on the chromosome in *D. aromatica*; however VIMSS581830 is adjacent to a homolog of Ribulose-phosphate 3-epimerase (VIMSS581829, *rpe*).

The *ccoSNOQP* gene cluster codes for a *cbb*-type cytochrome oxidase that functions as the terminal electron donor to O<sub>2</sub> in the aerobic respiration of *Rhodobacter capsulatus* [94]. These genes are present in a cluster as VIMSS580484-580486 and VIMSS584273-584274; note that these genes are present in a large number of Betaproteobacteria.

Other carbon cycles, such as the reverse TCA cycle and the Wood-Ljungdahl pathways, are missing critical enzymes in this genome, and are not present as such.

#### Sulfur

Sulfate and thiosulfate transport appear to be encoded in the gene cluster composed of an OmpA type protein (VIMSS581631) followed by orthologs of a sulfate/thiosulfate specific binding protein Sbp (VIMSS581632), a CysU or T sulfate/thiosulfate transport system permease T protein (VIMSS581633), a CysW ABC-type sulfate transport system permease component (VIMSS581634), and a CysA ATP-binding component of sulfate permease (VIMSS581635).

In addition, candidates for the transcriptional regulator of sulfur assimilation from sulfate are present and include:

CysB, CysH, and CysI (VIMSS582364, 582360 and 582362, respectively).

A probable sulfur oxidation enzyme cluster is present and contains homologs of SoxFRCDYZAXB [95], with a putative SoxCD sulfur dehydrogenase, SoxF sulfide dehydrogenase, and SoxB sulfate thiohydrolase, which is predicted to support thiosulfate oxidation to sulphate (see Fig. 5). Functional predictions are taken from Friedrich et al. [95] [see Additional file 7]. A syntenic *sox* gene cluster is also found in *Anaeromyxobacter dehalogens* (although it lacks *soxFR*) and *Ralstonia eutropha*, but not in *A. aromaticum* EbN1. Thiosulfate oxidation, however, has not been reported under laboratory conditions tested thus far, and experimental support for this physiological capability awaits further investigation.

Conversely, the cytoplasmic SorAB complex [96] is not present in *D. aromatica* nor *A. aromaticum* EbN1, although it is found in several other Betaproteobacteria, including *R. metallidurans*, *R. eutropha*, *R. solanacearum*, *C. violaceum*, and *B. japonicum*.

#### Gene Family Expansion

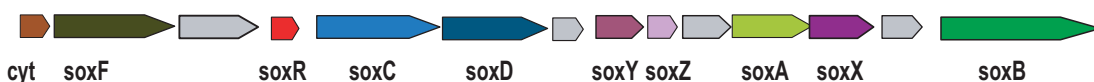
To determine candidates for recent gene duplication events, extensive phylogenomic profile analyses were conducted for all sets of paralogs in the genome. Flower Power recruitment and clustering against the non-redundant Genbank protein set was done, and the resulting alignments were analyzed using the tree-building SCI-PHY or Belvu based neighbor-joining utilities. The alignment of two or more *D. aromatica* protein sequences in a clade such that they displayed higher % identity to each other than to orthologs present in other species was interpreted as an indication of a probable recent duplication event, either in the *D. aromatica* genome itself or in a progenitor species. Results of this analysis are shown in Table 3.

Potential gene family expansion is indicated in several functional groups, including the following: signaling proteins (including cAMP signaling, histidine kinases, and others), Mhp-like aromatic oxidation complexes, nitrogen metabolism proteins and transport proteins.

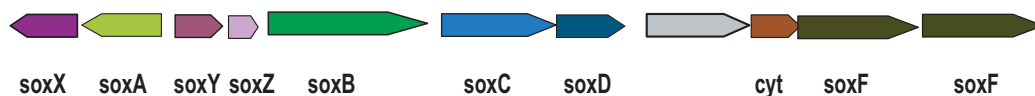
Most duplications indicate that a single gene, rather than sets of genes, were replicated. An exception is the Tra/Type IV transport cluster (VIMSS582581-582601 and VIMSS582864-582884) noted previously. In the protein sets for the histidine kinase/response regulator, duplication of histidine kinase appears to occur without duplication of the adjacent response regulator. The paralogs created by recent duplication events are typically found well-removed from one another on the chromosome, although some tandem repeats of single genes were noted.

### Sulfur Oxidation

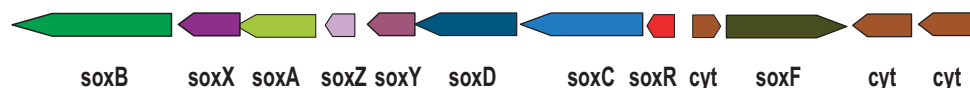
*Ralstonia eutropha* JMP



*Rhodopseudomonas palustris* HaA2



*Dechloromonas aromatica* RCB



**Figure 5**  
**Sulfur oxidation (thiosulfate to sulfate) candidates in *R. eutropha*, *R. palustris*, and *D. aromatica*.** Proposed model for this periplasmic complex is as follows: SoxXA, oxidatively links thiosulfate to SoxY; SoxB, potential sulfate thiohydrolase, interacts with SoxYZ (hydrolyzes sulfate from SoxY to regenerate); SoxCD, a sulfur dehydrogenase; oxidizes persulfide on SoxY to cysteine-S-sulfate and potentially yields 6 electrons per sulphate; SoxC, sulfite oxidase/dehydrogenase with homology to nitrate reductase, induced by thiosulfate; SoxDE, both c-type cytochromes with two heme-type binding sites; and SoxF, a FAD flavoprotein with sulfide dehydrogenase activity. Cyt, cytochrome.

However, the highest percent identity was not found between pairs of genes in tandem repeats.

#### Discussion

Discussion of results and analyses concerning aromatic degradation, various predicted metabolic cycles, secretion, signaling, quorum-sensing and gene family expansion are included in the relevant sections, above.

#### Conclusion

*Dechloromonas aromatica* strain RCB appears to support a highly complex lifestyle which might involve biofilm formation and interaction with a eukaryotic host. It lacks pre-

dicted enzyme families for anaerobic aromatic catabolism, though it supports degradation of several aromatic species in the absence of oxygen. The enzymes responsible for this metabolic function remain to be identified and characterized. It also encodes proteins suggestive of the ability to fix nitrogen and CO<sub>2</sub>, as well as thiosulfate oxidation. Converse to aromatic degradation, these enzymatic functionalities have yet to be experimentally demonstrated. In short, this genome was full of surprises.

The utility of TIGRfams and COGs families in these analyses cannot be overstated. New releases of TIGRfams dur-

ing the course of this analysis provided new insights and identified new functionality (malonate degradation cluster, PEP-Cterm transport and the epsH putative translocon, and urea degradation all were identified in the TIGRfam 7.0 additions). The HMM model building and assessment utilized as the major annotation approach for this study was employed to cover those protein families of interest that are not currently covered by TIGRfams. We utilized K. Sjölander's modelling and analysis tools, which are highly similar to those used to produce TIGRfams models. Overall, the extensive use of HMMs during this analysis allowed high confidence in predicted protein function, as well as certainty that several families of previously characterized anaerobic degradation enzymes for aromatic compounds are not present (eg Bss-ABCD and BcrAB).

### Authors' contributions

AL coordinated and oversaw the assembly of the genome. WSF, HF and GDB did the initial assembly of the genome. KKS conducted genome assembly, sequence finishing and gap closure activities, and created the final assembly. ST provided internal Joint Genome Institute assembly and analysis tools, and support in their use. KK was involved in the semi-automated genome annotation, and provided support for the VIMSS dataset and data lists from that set. KKS conducted all manual annotation work and protein family analyses described here-in. All authors have read and approved the final manuscript.

### Additional material

#### Additional file 1

*Phylogenomic analysis: Flower Power, SCI PHY and HMM scoring. The methodologies used for the annotation of predicted proteins in the D. aromatica genome, via hidden Markov model generation and assessment, are described in detail.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-351-S1.doc>]

#### Additional file 2

*Use of HMM pipelines to assign putative function to D. aromatica proteins and determine protein cohort. A flow diagram is used to depict each of the four approaches employed using HMM generation for determining the presence or absence of specific proteins or enzymes in the D. aromatica genome, as well as for annotation of the predicted protein set.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-351-S2.doc>]

#### Additional file 3

*Dechloromonas aromatica RCB genome assembly statistics. Statistics of the finishing process are shown in table format.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-351-S3.doc>]

#### Additional file 4

*Statistics for open reading frame predictions. Number of proteins from this genome having TIGRfam annotations (as of TIGRfam release 7.0) is shown.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-351-S4.doc>]

#### Additional file 5

*Number of predicted proteins with annotated diguanylate cyclase domains (IPR000160) in various genomes. The highly variable number of GGDEF domains predicted in proteins from a number of microbial species is shown.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-351-S5.doc>]

#### Additional file 6

*Type VI secretion cluster. Effector proteins in the IcmF-like Type VI secretion cluster annotated in D. aromatica are listed.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-351-S6.doc>]

#### Additional file 7

*Putative sulfur oxidation (Sox) cluster. A number of proteins capable of supporting sulfur oxidation in other species have homologs in the D. aromatica genome, and are listed.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-351-S7.doc>]

### Acknowledgements

KKS sincerely thanks Tanja Woyke for her very helpful suggestions and direction for creation of the tables and figures, Patrick Chain for helpful suggestions on the manuscript, Dan Kirshner for technical help on computational work, Nandini Krishnamurthy for building an internally clustered data-set of *D. aromatica* proteins as well as help with computational tools, Ching Shang for ideas regarding biochemical pathways, Paul Richardson for general support during the finishing phase and Frank W. Larimer for a cogent and extremely helpful critique of the manuscript. The genome finishing portion of this study was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231 and Los Alamos National Laboratory under Contract No. W-7405-ENG-36. The majority of the annotation was done as an independent project by KKS. Considerable intellectual support, computational and data analysis tools were provided by Adam Arkin, Katherine Huang, Morgan Price, Eric Alm, Dan Kirshner, and Kimmen Sjölander – sufficient gratitude cannot be expressed for their generous help.

### References

1. Coates JD, Chakraborty R, Lack JG, O'Connor SM, Cole KA, Bender KS, Achenbach LA: **Anaerobic benzene oxidation coupled to nitrate reduction in pure culture by two strains of Dechloromonas.** *Nature* 2001, **411(6841)**:1039-1043.
2. Chakraborty R, O'Connor SM, Chan E, Coates JD: **Anaerobic degradation of benzene, toluene, ethylbenzene, and xylene**

- compounds by *Dechloromonas* strain RCB. *Appl Environ Microbiol* 2005, **71**(12):8649-8655.
3. Thrash JC, Van Trump JI, Weber KA, Miller E, Achenbach LA, Coates JD: **Electrochemical stimulation of microbial perchlorate reduction.** *Environ Sci Technol* 2007, **41**(5):1740-1746.
  4. Kasai Y, Takahata Y, Manefield M, Watanabe K: **RNA-based stable isotope probing and isolation of anaerobic benzene-degrading bacteria from gasoline-contaminated groundwater.** *Appl Environ Microbiol* 2006, **72**(5):3586-3592.
  5. Chakraborty R, Coates JD: **Hydroxylation and carboxylation – two crucial steps of anaerobic benzene degradation by *Dechloromonas* strain RCB.** *Appl Environ Microbiol* 2005, **71**(9):5427-5432.
  6. Beller HR, Spormann AM: **Analysis of the novel benzylsuccinate synthase reaction for anaerobic toluene activation based on structural studies of the product.** *J Bacteriol* 1998, **180**(20):5454-5457.
  7. Edwards A, Voss H, Rice P, Civitello A, Stegemann J, Schwager C, Zimmermann J, Erfle H, Caskey CT, Ansoorge W: **Automated DNA sequencing of the human HPRT locus.** *Genomics* 1990, **6**(4):593-608.
  8. Ewing B, Hillier L, Wendl MC, Green P: **Base-calling of automated sequencer traces using phred. I. Accuracy assessment.** *Genome Res* 1998, **8**(3):175-185.
  9. Ewing B, Green P: **Base-calling of automated sequencer traces using phred. II. Error probabilities.** *Genome Res* 1998, **8**(3):186-194.
  10. Gordon D, Abajian C, Green P: **Consed: a graphical tool for sequence finishing.** *Genome Res* 1998, **8**(3):195-202.
  11. Badger JH, Olsen GJ: **CRITICA: coding region identification tool invoking comparative analysis.** *Mol Biol Evol* 1999, **16**(4):512-524.
  12. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL: **Improved microbial gene identification with GLIMMER.** *Nucleic Acids Res* 1999, **27**(23):4636-4641.
  13. Alm EJ, Huang KH, Price MN, Koche RP, Keller K, Dubchak IL, Arkin AP: **The MicrobesOnline Web site for comparative genomics.** *Genome Res* 2005, **15**(7):1015-1022.
  14. Chakraborty R, Coates JD: **Anaerobic degradation of monoaromatic hydrocarbons.** *Appl Microbiol Biotechnol* 2004, **64**(4):437-446.
  15. Eglund PG, Pelletier DA, Dispensa M, Gibson J, Harwood CS: **A cluster of bacterial genes for anaerobic benzene ring biodegradation.** *Proc Natl Acad Sci USA* 1997, **94**(12):6484-6489.
  16. Heider J, Fuchs G: **Anaerobic metabolism of aromatic compounds.** *Eur J Biochem* 1997, **243**(3):577-596.
  17. Boll M, Fuchs G, Heider J: **Anaerobic oxidation of aromatic compounds and hydrocarbons.** *Curr Opin Chem Biol* 2002, **6**(5):604-611.
  18. Storm CE, Sonnhammer EL: **Comprehensive analysis of orthologous protein domains using the HOPS database.** *Genome Res* 2003, **13**(10):2353-2362.
  19. Eisen JA: **Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis.** *Genome Res* 1998, **8**(3):163-167.
  20. Kuhner S, Wohlbrand L, Fritz I, Wrruck W, Hultschig C, Hufnagel P, Kube M, Reinhardt R, Rabus R: **Substrate-dependent regulation of anaerobic degradation pathways for toluene and ethylbenzene in a denitrifying bacterium, strain EbNI.** *J Bacteriol* 2005, **187**(4):1493-1503.
  21. Rabus R, Kube M, Heider J, Beck A, Heitmann K, Widdel F, Reinhardt R: **The genome sequence of an anaerobic aromatic-degrading denitrifying bacterium, strain EbNI.** *Arch Microbiol* 2005, **183**(1):27-36.
  22. Kube M, Heider J, Amann J, Hufnagel P, Kuhner S, Beck A, Reinhardt R, Rabus R: **Genes involved in the anaerobic degradation of toluene in a denitrifying bacterium, strain EbNI.** *Arch Microbiol* 2004, **181**(3):182-194.
  23. Kane SR, Beller HR, Legler TC, Anderson RT: **Biochemical and genetic evidence of benzylsuccinate synthase in toluene-degrading, ferric iron-reducing *Geobacter metallireducens*.** *Biodegradation* 2002, **13**(2):149-154.
  24. Bender KS, Shang C, Chakraborty R, Belchik SM, Coates JD, Achenbach LA: **Identification, characterization, and classification of genes encoding perchlorate reductase.** *J Bacteriol* 2005, **187**(15):5090-5096.
  25. Rabus R, Kube M, Beck A, Widdel F, Reinhardt R: **Genes involved in the anaerobic degradation of ethylbenzene in a denitrifying bacterium, strain EbNI.** *Arch Microbiol* 2002, **178**(6):506-516.
  26. Weelink SA, Tan NC, ten Broeke H, Kieboom C van den, van Doesburg W, Langenhoff AA, Gerritse J, Junca H, Stams AJ: **Isolation and characterization of *Alicyclophilus denitrificans* strain BC, which grows on benzene with chlorate as the electron acceptor.** *Appl Environ Microbiol* 2008, **74**(21):6672-6681.
  27. Sibley MH, Raleigh EA: **Cassette-like variation of restriction enzyme genes in *Escherichia coli* C and relatives.** *Nucleic Acids Res* 2004, **32**(2):522-534.
  28. Laurie AD, Lloyd-Jones G: **The phn genes of *Burkholderia* sp. strain RP007 constitute a divergent gene cluster for polycyclic aromatic hydrocarbon catabolism.** *J Bacteriol* 1999, **181**(2):531-540.
  29. Pelletier DA, Harwood CS: **2-Hydroxycyclohexanecarboxyl coenzyme A dehydrogenase, an enzyme characteristic of the anaerobic benzoate degradation pathway used by *Rhodospseudomonas palustris*.** *J Bacteriol* 2000, **182**(10):2753-2760.
  30. Arai H, Ohishi T, Chang MY, Kudo T: **Arrangement and regulation of the genes for meta-pathway enzymes required for degradation of phenol in *Comamonas testosteroni* TA441.** *Microbiology* 2000, **146**(Pt 7):1707-1715.
  31. Arai H, Yamamoto T, Ohishi T, Shimizu T, Nakata T, Kudo T: **Genetic organization and characteristics of the 3-(3-hydroxyphenyl)propionic acid degradation pathway of *Comamonas testosteroni* TA441.** *Microbiology* 1999, **145**(Pt 10):2813-2820.
  32. Hofer B, Eltis LD, Dowling DN, Timmis KN: **Genetic analysis of a *Pseudomonas* locus encoding a pathway for biphenyl/polychlorinated biphenyl degradation.** *Gene* 1993, **130**(1):47-55.
  33. Ward NC, Croft KD, Puddey IB, Hodgson JM: **Supplementation with grape seed polyphenols results in increased urinary excretion of 3-hydroxyphenylpropionic Acid, an important metabolite of proanthocyanidins in humans.** *J Agric Food Chem* 2004, **52**(17):5545-5549.
  34. Ma YaH, S D: **The catechol 2,3-dioxygenase gene and toluene monooxygenase genes from *Burkholderia* sp. AA1, an isolate capable of degrading aliphatic hydrocarbons and toluene.** *J Ind Microbiol Biotechnol* 2000, **25**:127-131.
  35. Kitayama A, Kawakami Y, Nagamune T: **Gene organization and low regioselectivity in aromatic-ring hydroxylation of a benzene monooxygenase of *Pseudomonas aeruginosa* J1104.** *J Ferment Bioeng* 1996, **82**:421-425.
  36. Tao Y, Fishman A, Bentley WE, Wood TK: **Oxidation of benzene to phenol, catechol, and 1,2,3-trihydroxybenzene by toluene 4-monooxygenase of *Pseudomonas mendocina* KR1 and toluene 3-monooxygenase of *Ralstonia pickettii* PKO1.** *Appl Environ Microbiol* 2004, **70**(7):3814-3820.
  37. Zhang H, Luo H, Kamagata Y: **Characterization of the Phenol Hydroxylase from *Burkholderia kururiensis* KP23 Involved in Trichloroethylene Degradation by Gene Cloning and Disruption.** *Microbes Environ* 2003, **18**:167-173.
  38. Canada KA, Iwashita S, Shim H, Wood TK: **Directed evolution of toluene ortho-monooxygenase for enhanced 1-naphthol synthesis and chlorinated ethene degradation.** *J Bacteriol* 2002, **184**(2):344-349.
  39. Mizuno T, Kaneko T, Tabata S: **Compilation of all genes encoding bacterial two-component signal transducers in the genome of the cyanobacterium, *Synechocystis* sp. strain PCC 6803.** *DNA Res* 1996, **3**(6):407-414.
  40. Galperin MY: **Structural classification of bacterial response regulators: diversity of output domains and domain combinations.** *J Bacteriol* 2006, **188**(12):4169-4182.
  41. Ryjenkov DA, Tarutina M, Moskvina OV, Gomelsky M: **Cyclic diguanylate is a ubiquitous signaling molecule in bacteria: insights into biochemistry of the GGDEF protein domain.** *J Bacteriol* 2005, **187**(5):1792-1798.
  42. Mendez-Ortiz MM, Hyodo M, Hayakawa Y, Membrillo-Hernandez J: **Genome-wide transcriptional profile of *Escherichia coli* in response to high levels of the second messenger 3',5'-cyclic diguanylic acid.** *J Biol Chem* 2006, **281**(12):8090-8099.
  43. Wandersman C, Deleplaire P: **ToxC, an *Escherichia coli* outer membrane protein required for hemolysin secretion.** *Proc Natl Acad Sci USA* 1990, **87**(12):4776-4780.

44. Richarme G, Caldas TD: **Chaperone properties of the bacterial periplasmic substrate-binding proteins.** *J Biol Chem* 1997, **272(25)**:15607-15612.
45. Binet R, Letoffe S, Ghigo JM, Delepelaire P, Wandersman C: **Protein secretion by Gram-negative bacterial ABC exporters – a review.** *Gene* 1997, **192(1)**:7-11.
46. Omori K, Idei A: **Gram-negative bacterial ATP-binding cassette protein exporter family and diverse secretory proteins.** *J Biosci Bioeng* 2003, **95(1)**:1-12.
47. Pahel G, Tyler B: **A new glnA-linked regulatory gene for glutamine synthetase in Escherichia coli.** *Proc Natl Acad Sci USA* 1979, **76(9)**:4544-4548.
48. Meibom KL, Li XB, Nielsen AT, Wu CY, Roseman S, Schoolnik GK: **The Vibrio cholerae chitin utilization program.** *Proc Natl Acad Sci USA* 2004, **101(8)**:2524-2529.
49. Tomimaga A, Lan R, Reeves PR: **Evolutionary changes of the flhDC flagellar master operon in Shigella strains.** *J Bacteriol* 2005, **187(12)**:4295-4302.
50. Hueck CJ: **Type III protein secretion systems in bacterial pathogens of animals and plants.** *Microbiol Mol Biol Rev* 1998, **62(2)**:379-433.
51. Marsh JW, Taylor RK: **Genetic and transcriptional analyses of the Vibrio cholerae mannose-sensitive hemagglutinin type 4 pilus gene locus.** *J Bacteriol* 1999, **181(4)**:1110-1117.
52. Mougous JD, Cuff ME, Raunser S, Shen A, Zhou M, Gifford CA, Goodman AL, Joachimiak G, Ordonez CL, Lory S, et al.: **A virulence locus of Pseudomonas aeruginosa encodes a protein secretion apparatus.** *Science* 2006, **312(5779)**:1526-1530.
53. Pukatzki S, Ma AT, Sturtevant D, Krastins B, Sarracino D, Nelson WC, Heidelberg JF, Mekalanos JJ: **Identification of a conserved bacterial protein secretion system in Vibrio cholerae using the Dictyostelium host model system.** *Proc Natl Acad Sci USA* 2006, **103(5)**:1528-1533.
54. Parsons DA, Heffron F: **sciS, an icmF homolog in Salmonella enterica serovar Typhimurium, limits intracellular replication and decreases virulence.** *Infect Immun* 2005, **73(7)**:4338-4345.
55. Bingle LE, Bailey CM, Pallen MJ: **Type VI secretion: a beginner's guide.** *Curr Opin Microbiol* 2008, **11(1)**:3-8.
56. Watnick P, Kolter R: **Biofilm, city of microbes.** *J Bacteriol* 2000, **182(10)**:2675-2679.
57. Barraud N, Hassett DJ, Hwang SH, Rice SA, Kjelleberg S, Webb JS: **Involvement of nitric oxide in biofilm dispersal of Pseudomonas aeruginosa.** *J Bacteriol* 2006, **188(21)**:7344-7353.
58. Federle MJ, Bassler BL: **Interspecies communication in bacteria.** *J Clin Invest* 2003, **112(9)**:1291-1299.
59. Fuqua C, Parsek MR, Greenberg EP: **Regulation of gene expression by cell-to-cell communication: acyl-homoserine lactone quorum sensing.** *Annu Rev Genet* 2001, **35**:439-468.
60. Withers H, Swift S, Williams P: **Quorum sensing as an integral component of gene regulatory networks in Gram-negative bacteria.** *Curr Opin Microbiol* 2001, **4(2)**:186-193.
61. Huang JJ, Han JI, Zhang LH, Leadbetter JR: **Utilization of acyl-homoserine lactone quorum signals for growth by a soil pseudomonad and Pseudomonas aeruginosa PAO1.** *Appl Environ Microbiol* 2003, **69(10)**:5941-5949.
62. Cha C, Gao P, Chen YC, Shaw PD, Farrand SK: **Production of acyl-homoserine lactone quorum-sensing signals by gram-negative plant-associated bacteria.** *Mol Plant Microbe Interact* 1998, **11(11)**:1119-1129.
63. Thakor N, Trivedi U, Patel KC: **Biosynthesis of medium chain length poly(3-hydroxyalkanoates) (mcl-PHAs) by Comamonas testosteroni during cultivation on vegetable oils.** *Bioresour Technol* 2005, **96(17)**:1843-1850.
64. Potter M, Madkour MH, Mayer F, Steinbuechel A: **Regulation of phasin expression and polyhydroxyalkanoate (PHA) granule formation in Ralstonia eutropha H16.** *Microbiology* 2002, **148(Pt 8)**:2413-2426.
65. Delledonne M, Porcari R, Fogher C: **Nucleotide sequence of the nodG gene of Azospirillum brasilense.** *Nucleic Acids Res* 1990, **18(21)**:6435.
66. Trautwein K, Kuhner S, Wohlbrand L, Halder T, Kuchta K, Steinbuechel A, Rabus R: **Solvent stress response of the denitrifying bacterium "Aromatoleum aromaticum" strain EbN1.** *Appl Environ Microbiol* 2008, **74(8)**:2267-2274.
67. Zago A, Chugani S, Chakrabarty AM: **Cloning and characterization of polyphosphate kinase and exopolyphosphatase genes from Pseudomonas aeruginosa 8830.** *Appl Environ Microbiol* 1999, **65(5)**:2065-2071.
68. Rao NN, Kornberg A: **Inorganic polyphosphate supports resistance and survival of stationary-phase Escherichia coli.** *J Bacteriol* 1996, **178(5)**:1394-1400.
69. Reusch RN, Sadoff HL: **Putative structure and functions of a poly-beta-hydroxybutyrate/calcium polyphosphate channel in bacterial plasma membranes.** *Proc Natl Acad Sci USA* 1988, **85(12)**:4176-4180.
70. Santaella C, Schue M, Berge O, Heulin T, Achouak W: **The exopolysaccharide of Rhizobium sp. YAS34 is not necessary for biofilm formation on Arabidopsis thaliana and Brassica napus roots but contributes to root colonization.** *Environ Microbiol* 2008, **10(8)**:2150-2163.
71. Haft DH, Paulsen IT, Ward N, Selengut JD: **Exopolysaccharide-associated protein sorting in environmental organisms: the PEP-CTERM/EpsH system. Application of a novel phylogenetic profiling heuristic.** *BMC Biol* 2006, **4**:29.
72. Cabello P, Pino C, Olmo-Mira MF, Castillo F, Roldan MD, Moreno-Vivian C: **Hydroxylamine assimilation by Rhodobacter capsulatus EIF1. requirement of the hcp gene (hybrid cluster protein) located in the nitrate assimilation nas gene region for hydroxylamine reduction.** *J Biol Chem* 2004, **279(44)**:45485-45494.
73. Anjum MF, Stevanin TM, Read RC, Moir JW: **Nitric oxide metabolism in Neisseria meningitidis.** *J Bacteriol* 2002, **184(11)**:2987-2993.
74. Lin JT, Goldman BS, Stewart V: **The nasFEDCBA operon for nitrate and nitrite assimilation in Klebsiella pneumoniae M5a1.** *J Bacteriol* 1994, **176(9)**:2551-2559.
75. Gutierrez JC, Ramos F, Ortner L, Tortolero M: **nasST, two genes involved in the induction of the assimilatory nitrite-nitrate reductase operon (nasAB) of Azotobacter vinelandii.** *Mol Microbiol* 1995, **18(3)**:579-591.
76. Ogawa K, Akagawa E, Yamane K, Sun ZW, LaCelle M, Zuber P, Nakano MM: **The nasB operon and nasA gene are required for nitrate/nitrite assimilation in Bacillus subtilis.** *J Bacteriol* 1995, **177(5)**:1409-1413.
77. Allen AE, Booth MG, Frischer ME, Verity PG, Zehr JP, Zani S: **Diversity and detection of nitrate assimilation genes in marine bacteria.** *Appl Environ Microbiol* 2001, **67(11)**:5343-5348.
78. Siddiqui RA, Warnecke-Eberz U, Hengsberger A, Schneider B, Kostka S, Friedrich B: **Structure and function of a periplasmic nitrate reductase in Alcaligenes eutrophus H16.** *J Bacteriol* 1993, **175(18)**:5867-5876.
79. Turner SM, Moir JW, Griffiths L, Overton TW, Smith H, Cole JA: **Mutational and biochemical analysis of cytochrome c', a nitric oxide-binding lipoprotein important for adaptation of Neisseria gonorrhoeae to oxygen-limited growth.** *Biochem J* 2005, **388(Pt 2)**:545-553.
80. Baek SH, Shapleigh JP: **Expression of nitrite and nitric oxide reductases in free-living and plant-associated Agrobacterium tumefaciens C58 cells.** *Appl Environ Microbiol* 2005, **71(8)**:4427-4436.
81. Waller AS, Cox EE, Edwards EA: **Perchlorate-reducing microorganisms isolated from contaminated sites.** *Environ Microbiol* 2004, **6(5)**:517-527.
82. Kanamori T, Kanou N, Atomi H, Imanaka T: **Enzymatic characterization of a prokaryotic urea carboxylase.** *J Bacteriol* 2004, **186(9)**:2532-2539.
83. Simon J, Einsle O, Kroneck PM, Zumft WG: **The unprecedented nos gene cluster of Wolinella succinogenes encodes a novel respiratory electron transfer pathway to cytochrome c nitrous oxide reductase.** *FEBS Lett* 2004, **569(1-3)**:7-12.
84. Yoon SS, Hennigan RF, Hilliard GM, Ochsner UA, Parvatiyar K, Kamani MC, Allen HL, DeKievit TR, Gardner PR, Schwab U, et al.: **Pseudomonas aeruginosa anaerobic respiration in biofilms: relationships to cystic fibrosis pathogenesis.** *Dev Cell* 2002, **3(4)**:593-603.
85. Baar C, Eppinger M, Raddatz G, Simon J, Lanz C, Klimmek O, Nandakumar R, Gross R, Rosinus A, Keller H, et al.: **Complete genome sequence and analysis of Wolinella succinogenes.** *Proc Natl Acad Sci USA* 2003, **100(20)**:11690-11695.

86. Schmehl M, Jahn A, Meyer zu Vilsendorf A, Hennecke S, Masepohl B, Schuppler M, Marxer M, Oelze J, Klipp W: **Identification of a new class of nitrogen fixation genes in Rhodobacter capsulatus: a putative membrane complex involved in electron transport to nitrogenase.** *Mol Gen Genet* 1993, **241(5-6)**:602-615.
87. Baginsky C, Brito B, Imperial J, Palacios JM, Ruiz-Argueso T: **Diversity and evolution of hydrogenase systems in rhizobia.** *Appl Environ Microbiol* 2002, **68(10)**:4915-4924.
88. Menon AL, Mortenson LE, Robson RL: **Nucleotide sequences and genetic analysis of hydrogen oxidation (hox) genes in Azotobacter vinelandii.** *J Bacteriol* 1992, **174(14)**:4549-4557.
89. Durmowicz MC, Maier RJ: **Roles of HoxX and HoxA in biosynthesis of hydrogenase in Bradyrhizobium japonicum.** *J Bacteriol* 1997, **179(11)**:3676-3682.
90. Pawlowski K, Klosse U, de Bruijn FJ: **Characterization of a novel Azorhizobium caulinodans ORS571 two-component regulatory system, NtrY/NtrX, involved in nitrogen fixation and metabolism.** *Mol Gen Genet* 1991, **231(1)**:124-138.
91. Robinson JJ, Stein JL, Cavanaugh CM: **Cloning and sequencing of a form II ribulose-1,5-biphosphate carboxylase/oxygenase from the bacterial symbiont of the hydrothermal vent tube-worm Riftia pachyptila.** *J Bacteriol* 1998, **180(6)**:1596-1599.
92. Oda Y, Wanders W, Huisman LA, Meijer WG, Gottschal JC, Forney LJ: **Genotypic and phenotypic diversity within species of purple nonsulfur bacteria isolated from aquatic sediments.** *Appl Environ Microbiol* 2002, **68(7)**:3467-3477.
93. Gibson JL, Tabita FR: **Analysis of the cbbXYZ operon in Rhodobacter sphaeroides.** *J Bacteriol* 1997, **179(3)**:663-669.
94. Thony-Meyer L, Beck C, Preisig O, Hennecke H: **The ccoNOQP gene cluster codes for a cb-type cytochrome oxidase that functions in aerobic respiration of Rhodobacter capsulatus.** *Mol Microbiol* 1994, **14(4)**:705-716.
95. Friedrich CG, Bardischewsky F, Rother D, Quentmeier A, Fischer J: **Prokaryotic sulfur oxidation.** *Curr Opin Microbiol* 2005, **8(3)**:253-259.
96. Kappler U, Friedrich CG, Truper HG, Dahl C: **Evidence for two pathways of thiosulfate oxidation in Starkeya novella (formerly Thiobacillus novellus).** *Arch Microbiol* 2001, **175(2)**:102-111.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

