Complete genome sequence of the phenanthrene-degrading soil bacterium *Delftia acidovorans* Cs1-4

Ameeha R. Shetty¹, Vidya de Gannes², Chioma C. Obi³, Susan Lucas⁴, Alla Lapidus⁵, Jan-Fang Cheng⁶, Lynne A. Goodwin⁶, Samuel Pitluck⁷, Linda Peters⁴, Natalia Mikhailova⁴, Hazuki Teshima⁶, Cliff Han⁶, Roxanne Tapia⁶, Miriam Land⁷, Loren J. Hauser¹, Nikos Kyrpides⁴, Natalia Ivanova⁴, Ioanna Pagani⁴, Patrick S. G. Chain⁶, Vincent J. Denef⁴, Tanya Woyke⁴ and William J. Hickey¹*

**Abstract**

Polycyclic aromatic hydrocarbons (PAH) are ubiquitous environmental pollutants and microbial biodegradation is an important means of remediation of PAH-contaminated soil. *Delftia acidovorans* Cs1-4 (formerly *Delftia* sp. Cs1-4) was isolated by using phenanthrene as the sole carbon source from PAH contaminated soil in Wisconsin. Its full genome sequence was determined to gain insights into a mechanisms underlying biodegradation of PAH. Three genomic libraries were constructed and sequenced: an Illumina GAii shotgun library (916,416,493 reads), a 454 Titanium standard library (770,171 reads) and one paired-end 454 library (average insert size of 8 kb, 508,092 reads). The initial assembly contained 40 contigs in two scaffolds. The 454 Titanium standard data and the 454 paired end data were assembled together and the consensus sequences were computationally shredded into 2 kb overlapping shreds. Illumina sequencing data was assembled, and the consensus sequence was computationally shredded into 1.5 kb overlapping shreds. Gaps between contigs were closed by editing in Consed, by PCR and by Bubble PCR primer walks. A total of 182 additional reactions were needed to close gaps and to raise the quality of the finished sequence. The final assembly is based on 253.3 Mb of 454 draft data (averaging 38.4 X coverage) and 590.2 Mb of Illumina draft data (averaging 89.4 X coverage). The genome of strain Cs1-4 consists of a single circular chromosome of 6,685,842 bp (66.7 %G+C) containing 6,028 predicted genes; 5,931 of these genes were protein-encoding and 4,425 gene products were assigned to a putative function. Genes encoding phenanthrene degradation were localized to a 232 kb genomic island (termed the *phn* island), which contained near its 3' end a bacteriophage P4-like integrase, an enzyme often associated with chromosomal integration of mobile genetic elements. Other biodegradation pathways reconstructed from the genome sequence included: benzoate (by the acetyl-CoA pathway), styrene, nicotinic acid (by the maleamate pathway) and the pesticides Dicamba and Fenitrothion. Determination of the complete genome sequence of *D. acidovorans* Cs1-4 has provided new insights the microbial mechanisms of PAH biodegradation that may shape the process in the environment.

**Keywords:** *Delftia acidovorans* Cs1-4, Genome, *phn* island, Phenanthrene, polycyclic aromatic hydrocarbons, Nanopods

* Correspondence: wjhickey@wisc.edu

¹O.N. Allen Laboratory for Soil Microbiology, Department of Soil Science, University of Wisconsin-Madison, Madison, WI 53706, USA

Full list of author information is available at the end of the article

© 2015 Shetty et al. Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
Introduction
Polycyclic aromatic hydrocarbons (PAH) are ubiquitous environmental pollutants and microbial biodegradation is an important means of remediation of PAH-contaminated soil. *Delftia acidovorans* Cs1-4 (formerly *Delfia* sp. Cs1-4) was isolated using phenanthrene as the sole carbon source from PAH contaminated soil in Wisconsin [1] and its genome sequence was determined to gain insights into the mechanisms and pathways of PAH metabolism. *D. acidovorans* Cs1-4 was also unique as the strain in which the novel extracellular structures called nanopods were discovered [2]. Nanopods are extensions of the cell that consist of a surface layer protein encasing outer membrane vesicles (Fig. 1 [2]). The specific functions of nanopods in *D. acidovorans* Cs1-4 are unknown. But, some connection to phenanthrene degradation is likely as growth on that compound induces production of these structures [2] and mutants disabled in production of nanopods are impaired in growth on phenanthrene [2, 3].

Bacterial degraders of PAH can be grouped based on the amino acid similarities in the large subunit of the ring hydroxylating dioxygenase, which initiates PAH metabolism. Based on those RHD similarities, several groups of PAH-degrading bacteria have been resolved [4, 5], and *D. acidovorans* Cs1-4 is the first representative of the group designated as the Phn family to have a full genome sequence determined. A draft genome sequence has also been determined for a second representative of the Phn family, *Burkholderia* sp. Ch1-1 (GenBank ADNR00000000; [6]).

Bacteria belonging to different RHD groups typically differ in other characteristics relevant to PAH metabolism including the range of PAH degraded, pathways for PAH metabolism and the structure of gene clusters encoding PAH degradation [7]. Furthermore, lateral transfer of RHD genes of different phylogenetic groups appears to be mediated by different types of mobile genetic elements [8]. The full genome sequence of *D. acidovorans* Cs1-4 can thus provide insights into a variety of aspects of PAH metabolism in general, and phenanthrene degradation in particular.

Organism information
Classification and features
The genus *Delftia* is a phylogenetically cohesive group within the *Comamonadaceae* family of the *Betaproteobacteria* (Table 1, Fig. 2). At the time of writing, permanent draft genome sequences were publically available for four draft or finished genome sequences of *D. acidovorans* (including *D. acidovorans* Cs1-4). But, of those four genomes, only those of strains Cs1-4 and SPH1 (NC_0100002) appear to be *bona fide* representatives of this species, as 16S rRNA sequences of the other two strains, CCUG 247B and CCUG 15835, best match to *Delftia tsuruhatensis* (Table 2). These species affiliations were supported by a phylogenetic tree of the *Comamonadaceae*, which resolved three species clusters within *Delftia: D. acidovorans, D. tsuruhatensis* and *D. litopenaei* (Fig. 2). Strains CCUG 247B and CCUG 15835 clustered with *D. tsuruhatensis* instead of *D. acidovorans*. A fourth species, *D. lacustris*, clustered with *D. tsuruhatensis* and thus, in this analysis, did not have strong phylogenetic support as a distinct species (Fig. 2).

*Delftia acidovorans* Cs1-4 was isolated from soil contaminated by coal tar at the former site of a manufactured gas plant in Chippewa Falls, WI by using an enrichment culture with phenanthrene as the sole added carbon source [1]. Strain Cs1-4 has since been used as the model organism for the study of nanopods [2, 3]. *Delftia acidovorans* Cs1-4 is deposited in the culture collection of the United States Department of Agriculture, Agricultural Research Service (Peoria, IL) as strain NRRL B-65277.

The phospholipid fatty acid profile of *D. acidovorans* Cs1-4 (cells grown on Tryptic Soy Broth and PLFA prepared by the Bligh and Dyer method, [9]) was 19:0 CYCLO ω6c, 2 %; 18:1 ω7c, 6 %; 18:0, 9 %; 17:0 CYCLO, 23 %; 16:1 ω7c, 11 %; 16:0, 36 %; sum of minor PLFA, 14 %. The major quinone is ubiquinone. The glycolipid composition of lipopolysaccharide was (mole %): rhamnose, 49.1 %, glucose, 38.0 %, N-acetylglucosamine, 107. % and N-acetylgalactosamine, 2.2 %. The LPS was prepared by a Tri reagent method [10] and monosaccharide content determined by liquid chromatography-mass spectrometry. A distinguishing characteristic of *D. acidovorans* is a novel surface layer protein, NpdA [2] and NpdA of *D. acidovorans* Cs1-4 was antigenically distinct form that of *D. acidovorans* 15668 and SPH1 (Fig. 3). Immunoblotting was done following the method of Gallager et al. 2001 [11] with anti-NpdA prepared as described by Shetty et al. 2011 [2].

Biolog GN plates (Biolog, Hayward, CA) were used to assay metabolic characteristics of *D. acidovorans* Cs1-4, as well as *D. acidovorans* strains SPH1 and ATCC 15668. All three were positive for metabolism of: Tween 80, Tween 40, D-mannitol, D-psicose, pyruvic acid methyl ester, succinic acid mono-methyl-ester, cis-aconitic acid, formic acid, D-gluconic acid, α-hydroxybutyric acid, β-hydroxybutyric acid, γ-hydroxybutyric acid, β-hydroxyphenylacetic acid, itaconic acid, α-ketobutyric acid, α-ketoglutaric acid, α-ketovaleric acid, L-lactic acid, malonic acid, propionic acid, quinic acid, D-saccharic acid, sebacic acid, succinic acid, bromosuccinic acid, L-alaenamide, D-alanine, L-alanine, L-alanylglycine, L-aspartic acid, L-histidine, L-leucine, L-phenylalanine, L-proline, L-pyroglutamic acid, L-threonine, γ-amino butyric acid, urocanic acid and glycerol.

Strains Cs1-4, SPH1 and ATCC 15668 were negative for metabolism of: α-cyclodextrin, dextrin, glycogen, N-acetyl-D-galactosamine, N-acetyl-D-glucosamine, adonitol, L-arabinose, D-arabitol, D-cellulbiose, L-fructose,
D-galactose, gentiobiose, α-D-glucose, α-D-lactose, lactulose, maltose, D-mannose, D-melibiose, β-methyl-D-glucoside, D-raffinose, L-rhamnose, D-sorbitol, sucrose, D-trehalose, turanose, xylitol, D-galactonic acid lactone, D-galacturonic acid, D-glucosaminic acid, D-glucuronic acid, glycyrl-L-aspartic acid, glycyrl-L-glutamic acid, hydroxyl-L-proline, L-ornithine, D-serine, L-serine, D-melibiose, L-carnitine, inosine, uridine, thymidine,
phenylethylamine, putrescine, 2-aminoethanol, 2,3-butanediol, L-α-glycerol phosphate, α-D-glucose-1-phosphate and D-glucose-6-phosphate.

*Delphlia acidovorans* strains Cs1-4, SPH1 and ATCC 15688 varied in metabolism of six substrates: *i*-erythritol, D-fructose, m-inositol, citric acid, glucuronamide and L-asparagine. Strain Cs1-4 was positive for only *i*-erythritol, strain ATCC 15688 was positive for D-fructose and citric acid, while strain SPH1 metabolized all six of these compounds.

### Genome sequencing and annotation

#### Genome project history

The genome sequence was completed in May 2011 and presented for public access on December 2011 and is available in GenBank (NC_015563). Quality assurance of the genomic DNA preparation used for sequencing was done in the laboratory of W.J. Hickey at the University of Wisconsin-Madison. Sequencing, finishing and annotation were performed by the U.S. DOE JGI. A summary of the project information is shown in Table 3.

#### Growth conditions and genomic DNA preparation

*Delphlia acidovorans* Cs1-4 was grown aerobically in Nutrient Broth at 30 °C. DNA was isolated from 100 mL of overnight culture by a CTAB method [12]. Cells were collected by centrifugation (10,000 × g 10 min) and then resuspended in Tris-EDTA buffer to OD600 of 1.0. Lysozyme (100 mg/mL) was added followed by 10 % SDS and proteinase K (10 mg/mL). After incubation for 1 h at 37 °C, 5 M NaCl and CTAB/NaCl were added, and the solution incubated at 65 °C for 10 min. DNA was purified by phenol chloroform extraction, and then re-suspended in TE buffer with RNase (10 mg/mL). For quality confirmation, the DNA preparation was run on a 1 % agarose gel with phage λ DNA as a mass standard (DOE JGI).

#### Genome sequencing and assembly

The draft genome of *D. acidovorans* Cs1-4 was generated at the DOE JGI using a combination of Illumina and 454 technologies. An Illumina GAii shotgun library which generated 16,416,493 reads totaling 591 Mb, a 454 Titanium standard library which generated 16,416,493 reads totaling 591 Mb, and 454 technologies. An Illumina GAii shotgun library which generated 16,416,493 reads totaling 591 Mb, a 454 Titanium standard library which generated 770,171 reads totaling 234.2 Mb of 454 data, was constructed and sequenced. The initial draft assembly contained 40 contigs in 2 scaffolds. The 454 Titanium standard data and the 454 paired end data were assembled together with Newbler, version 2.3. The Newbler consensus sequences were computationally shredded into 2 kb overlapping shreds. Illumina sequencing data was assembled with VELVET, version 0.7.63, and the consensus sequence was computationally shredded into 1.5 kb overlapping shreds. We integrated the 454 Newbler consensus shreds, the Illumina VELVET consensus shreds and the read pairs in the 454 paired end library using parallel phrap, version SPS - 4.24 (High Performance Software, LLC).
Fig. 2 Phylogenetic tree highlighting the position of Delftia acidovorans Cs1-4 relative to other Delftia species as well as genera within the family Comamonadaeae. The 16S rRNA gene sequences were obtained from either type strains (strain designation ends with "T") and/or have full genome sequence data that is publicly available (GenBank accession numbers are indicated in parentheses) were aligned by using MUSCLE (http://www.drive5.com/muscle/). MEGA v.5.05 was used for phylogenetic reconstruction via the maximum-likelihood method (tree inference by nearest neighbor interchange), with the Tamura-Nei substitution model. Uniform bootstrap values obtained from 1000 replicate analyses are indicated at nodes. 

Table 2 Summary of currently available genome sequence data from bacteria identified as Delftia acidovorans

| Genome Name            | Taxon ID    | Status   | 16S rRNA Gene IDa | 16S rRNA Size(bp) | BLAST-N Match | Identity | Accession Number of BLAST-N Match |
|------------------------|-------------|----------|-------------------|-------------------|---------------|----------|----------------------------------|
| Delftia acidovorans Cs1-4b | 650716032    | Finished | 650858110         | 1521              | Delftia acidovorans Cs1-4 | 100 %   | NR_074626                        |
| Delftia acidovorans SPH-1   | 641228489    | Finished | 641295730         | 1526              | Delftia acidovorans SPH-1 | 100 %   | NR_074691                        |
| Delftia acidovorans CCUG 15835 | 2541046983  | Draft    | 2541343813        | 1521              | Delftia tsuruhatensisARB-1 | 100 %   | KC572558                         |
| Delftia acidovorans CCUG 2748 | 2541046984  | Draft    | 2541347625        | 1521              | Delftia tsuruhatensisARB-1 | 100 %   | KC572558                         |

aGene ID is as assigned by the JGI Integrated Microbial Genomes database
bGenome sequence determined under the name Delftia sp. Cs1-4
software Consed was used in the following finishing process. Illumina data was used to correct potential base errors and increase consensus quality using the software Polisher developed at JGI (Alla Lapidus, unpublished). Possible mis-assemblies were corrected using gapResolution (Cliff Han, unpublished), Dupfinisher, or sequencing cloned bridging PCR fragments with subcloning. Gaps between contigs were closed by editing in Consed, by PCR and by Bubble PCR primer walks (J-F Cheng, unpublished). A total of 182 additional reactions were necessary to close gaps and to raise the quality of the finished sequence. The total size of the genome is 6,685,842 bp and the final assembly is based on 253.3 Mb of 454 draft data which provides an average 38.4× coverage of the genome and 590.2 Mb of Illumina draft data which provides an average 89.4× coverage of the genome.

**Genome annotation**

Genes were identified using Prodigal as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline. The predicted CDSs were translated and used to search the National Center for Biotechnology Information non-redundant database, UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. These data sources were combined to assert a product description for each predicted protein. Non-coding genes and miscellaneous features were predicted using tRNAscan-SE, RNAMMer, Rfam, TMHMM, and signalP. The final genome sequence is deposited in GenBank under accession NC_015563.

**Genome properties**

The genome of strain Cs1-4 consists of a single circular chromosome of 6,685,842 bp (66.7 % G + C content) containing with 6,028 predicted genes (Fig. 4, Table 4). There were 5,931 protein-encoding genes; 4,425 of these gene products were assigned to a putative function with the remaining annotated as hypothetical proteins (Table 5). There were 80 tRNA genes and five 16S rRNA genes. For the latter, three were 100 % identical to each other, while two were 99 % identical (DelCs14_R0076, DelCs14_R0098).

**Insights from the genome sequence**

Comparisons with other sequenced *Delftia* genomes

The genome of strain Cs1-4 was similar to that of the other available *Delftia* genomes with respect to categories such as gene count and % G + C (Table 4). However, there was a clear distinction between these strains in the area of ribosomal genes, as their abundance was nearly identical in the genomes of strains Cs1-4 and SPH1, but much greater than that in...
strains CCUG 274B and CCUG 15835. There was also a marked difference between the organisms in the area of function prediction, with strains Cs1-4 and SPH1 having a much larger portion of genes with function predicted than did strains CCUG 274B and CCUG 15835. This divergence of the latter two strains from strains Cs1-4 and SPH1 would further support re-assignment of strains CCUG 274B and CCUG 15835 to a Delftia species other than acido- 

vorans (e.g., D. tsruhatensis).

The closest relative of D. acidovorans Cs1-4 with genome sequence data is D. acidovorans SPH1. Strain SPH1 was isolated from a sewage treatment plant in Germany as a part of a microbial consortium that degraded linear alkylbenzene sulfonates [13]. Strains Cs1-4 and SPH1 had 99 % identity over the full length of the 16S rRNA gene. Compared to the type strain D. acidovorans ATCC 15668, both strains Cs1-4 and SPH1 had 99 % identity over 99 % of the 16S rRNA gene. The genomes of strain Cs1-4 and strain SPH1 had an average nucleotide identity (ANI) of 97.48 %, which qualified the strains as the same species using the 95 % ANI threshold [14, 15]. Based on this high level of nucleotide sequence similarity, a common species is reasonable for these bacteria. But, while the genomic composition of strains SPH1 and

Table 4 Delftia acidovorans Cs1-4 genome statistics

| Attribute                     | Value | % of Total |
|-------------------------------|-------|------------|
| Genome size (bp)              | 6,685,842 | 100        |
| DNA coding (bp)               | 5,998,883 | 89.73      |
| DNA G + C content (bp)        | 4,460,657 | 66.72      |
| DNA scaffold                  | 1     | 0.00       |
| Total genes                   | 6,028 | 100.00     |
| Protein coding genes          | 5,931 | 98.39      |
| RNA genes                     | 97    | 1.61       |
| Pseudo genes                  | 70    | 1.16       |
| Genes in internal clusters    | 857   | 14.22      |
| Genes with function prediction| 4,425 | 73.41      |
| Genes assigned to COGs        | 4,845 | 80.37      |
| Genes assigned to Pfam domains| 5,061 | 83.96      |
| Genes with signal peptides    | 2,505 | 41.56      |
| Genes with transmembrane helices| 1,377 | 22.84     |
| CRISPR repeats                | 3     | NA         |

strains CCUG 274B and CCUG 15835. There was also a marked difference between the organisms in the area of function prediction, with strains Cs1-4 and SPH1 having a much larger portion of genes with function predicted than did strains CCUG 274B and CCUG 15835. This divergence of the latter two strains from strains Cs1-4 and SPH1 would further support re-assignment of strains CCUG 274B and CCUG 15835 to a Delftia species other than acido- 

vorans (e.g., D. tsruhatensis).

The closest relative of D. acidovorans Cs1-4 with genome sequence data is D. acidovorans SPH1. Strain SPH1 was isolated from a sewage treatment plant in Germany as a part of a microbial consortium that degraded linear alkylbenzene sulfonates [13]. Strains Cs1-4 and SPH1 had 99 % identity over the full length of the 16S rRNA gene. Compared to the type strain D. acidovorans ATCC 15668, both strains Cs1-4 and SPH1 had 99 % identity over 99 % of the 16S rRNA gene. The genomes of strain Cs1-4 and strain SPH1 had an average nucleotide identity (ANI) of 97.48 %, which qualified the strains as the same species using the 95 % ANI threshold [14, 15]. Based on this high level of nucleotide sequence similarity, a common species is reasonable for these bacteria. But, while the genomic composition of strains SPH1 and

Table 4 Delftia acidovorans Cs1-4 genome statistics

| Attribute                     | Value | % of Total |
|-------------------------------|-------|------------|
| Genome size (bp)              | 6,685,842 | 100        |
| DNA coding (bp)               | 5,998,883 | 89.73      |
| DNA G + C content (bp)        | 4,460,657 | 66.72      |
| DNA scaffold                  | 1     | 0.00       |
| Total genes                   | 6,028 | 100.00     |
| Protein coding genes          | 5,931 | 98.39      |
| RNA genes                     | 97    | 1.61       |
| Pseudo genes                  | 70    | 1.16       |
| Genes in internal clusters    | 857   | 14.22      |
| Genes with function prediction| 4,425 | 73.41      |
| Genes assigned to COGs        | 4,845 | 80.37      |
| Genes assigned to Pfam domains| 5,061 | 83.96      |
| Genes with signal peptides    | 2,505 | 41.56      |
| Genes with transmembrane helices| 1,377 | 22.84     |
| CRISPR repeats                | 3     | NA         |
Cs1-4 was similar, dot plots exhibited an “X” alignment (Fig. 5), which is indicative of inversions about the origin of replication [16].

Alignment of the *D. acidovorans* Cs1-4 and SPH1 genomes revealed in the former a large genomic island (232 kb) termed the *phn* island that was absent from strain SPH1 [6]. The island contained near its 3’ end a bacteriophage P4-like integrase, a type of enzyme often associated with chromosomal integration of mobile genetic elements [17]. Numerous close orthologs of the strain Cs1-4 integrase can be identified by BLAST-P searches of Genbank, and these may serve as starting points for the elucidation of mobile genetic elements possibly related to the *phn* island.

Conversely, genomic alignment of strains Cs1-4 and SPH1 revealed in strain SPH1 a mobile genetic element of ca. 67 kb, which was absent from strain Cs1-4. The loci bounding this element in strain SPH1 were (gene name, equivalent loci in strain Cs1-4) Daci_1694 (*rpoH*, Delcs14_4885) and Daci_1756 (*inclR*, Delcs14_4884). The mobile genetic element in strain SPH1 contained metal detoxification functions, and had extensive similarity to a region in the genomes of strains CCUG 274B and CCUG 15835.

The secretory machinery of strain Cs1-4 were similar to strain SPH1, as well as to strains CCUG 274B and CCUG 15835. These consisted of Type II, Type IV and Type VI secretion system (T6SS) along with the components of a Sec-Signal Recognition Particle Translocon and the *tatABC* genes of the Twin-Arginine Translocation pathway. For all strains, there was a single T6SS cluster. The functions of T6SS have been explored mostly in pathogenic bacteria, where a common feature is mediation of intercellular contact in antagonistic interactions [18]. Functions of T6SS in environmental bacteria such as the *D. acidovorans* strains discussed here are unknown.

*Delphi acidovorans* Cs1-4 produces a novel surface layer protein, which is essential for the formation of extracellular structure called nanopods [2]. The surface layer protein (termed Nanopod Protein A, NpdA) is encoded by Delcs14_5206 (*npdA*), and orthologs of *npdA* are present in the genomes of *D. acidovorans* SPH1 as well as *D. acidovorans* strains CCUG 274B and CCUG 15835.

### Table 5

| Code | Value | % of total | Description |
|------|-------|------------|-------------|
| J    | 207   | 3.78       | Translation |
| A    | 2     | 0.04       | RNA processing and modification |
| K    | 581   | 10.62      | Transcription |
| L    | 168   | 3.07       | Replication, recombination and repair |
| B    | 2     | 0.04       | Chromatin structure and dynamics |
| D    | 32    | 0.58       | Cell cycle control, mitosis and meiosis |
| V    | 71    | 1.3        | Defense mechanisms |
| T    | 362   | 6.62       | Signal transduction mechanisms |
| M    | 251   | 4.59       | Cell wall/membrane biogenesis |
| N    | 138   | 2.52       | Cell motility |
| U    | 145   | 2.65       | Intracellular trafficking and secretion |
| O    | 162   | 2.96       | Posttranslational modification, protein turnover, chaperones |
| C    | 346   | 6.32       | Energy production and conversion |
| G    | 242   | 4.42       | Carbohydrate transport and metabolism |
| E    | 479   | 8.76       | Amino acid transport and metabolism |
| F    | 88    | 1.61       | Nucleotide transport and metabolism |
| H    | 191   | 3.49       | Coenzyme transport and metabolism |
| I    | 261   | 4.77       | Lipid transport and metabolism |
| P    | 368   | 6.73       | Inorganic ion transport and metabolism |
| Q    | 172   | 3.14       | Secondary metabolites biosynthesis, transport and catabolism |
| R    | 621   | 11.35      | General function prediction only |
| S    | 580   | 10.60      | Function unknown |
| -    | 1183  | 19.63      | Not in COGs |

The total is based on the total number of protein coding genes in the annotated genome.
Profiles of metabolic networks and pathways

Characterization of phenanthrene catabolism genes

Genes for the entire phenanthrene catabolic pathway were identified on a novel 232 kb genomic island named the phn island [6] and were segregated into three clusters that were predicted to encode the metabolism of phenanthrene to ortho-phthalate (phn genes), ortho-phthalate to protocatechuate (oph genes) and meta-cleavage of protocatechuate to pyruvate and oxaloacetate (pmd genes). These clusters were non-contiguous; the oph and pmd clusters were separated from each other by ca. 5 kb and located 60 kb upstream of the phn cluster. The phn gene cluster had a %G + C content of ca. 50, which differed significantly from the 66.7 % G + C content of the Cs1-4 chromosome, while %G + C of the oph and pmd genes was similar to the chromosomal backbone. The G + C content of Comamonadaceae genotypes ranges from 60 % to 70 %, thus the phylogenetic origin of the phn genes is outside of this family and likely outside of the order Burkholderiales.

Styrene degradation via the phenyl acetate pathway

Styrene is often a soil pollutant and strain Cs1-4 possessed genes predicted to encode its degradation by the phenylacetate pathway. The putative styrene operon (conversion of styrene to phenyl acetate [19]), consisted of a regulatory element (marR, DelCs1_4846), dienelactone hydrolase (DelCs1_4847), flavin reductase (styB, DelCs1_4848), monoxygenase (styA, DelCs1_48449), short chain dehydrogenase, (DelCs1_4850) and AraC-type transcriptional regulator (DelCs1_4851).

The genetics of phenyl acetate metabolism has been studied in a variety of bacteria, and thirteen genes encoding its transformation to succinyl-CoA and acetyl-CoA have been identified [20, 21]. These genes often occur in a single cluster, but in strain Cs1-4, they are dispersed across the genome. The single largest cluster contained paaABCDE (DelCs1_5720-24) and paaK (DelCs1_5725), which were predicted to encode a phenylacetyl-CoA epoxidase and phenylacetate-CoA ligase, respectively. Other orthologs that were identified included an epoxycyclopheryl-CoA isomerase (paaG, DelCs1_0512) and a ring-opening enzyme (paaN, DelCs1_5726).

Benzoate degradation by the benzoyl-CoA pathway

Metabolism of benzoate is important in the degradation of many aromatic compounds, and benzoate degradation by aerobic bacteria is most commonly initiated by oxygenases. In contrast, growth of strain Cs1-4 on benzoate was predicted to proceed by an alternative pathway in which benzoyl-CoA is formed as the primary metabolite [22]. The gene cluster putatively conferring this function included an ABC-type transporter (DelCs1_0078-82), a benzoate-CoA ligase (DelCs1_0073) and a benzoate oxygenase (boxABC, DelCs1_0073-0075).

Nicotinic acid metabolism by the maleamate pathway

Biodegradation of nicotinic acid has been widely studied as a model for metabolism of N-heterocycles. Also, there is interest in the application of bioremediation of nicotine-contaminated soils at tobacco processing facilities. Jimenez et al. 2008 [23] conducted a detailed analysis of genes in Pseudomonas putida KT2440 encoding nicotinic degradation by the maleamate pathway, and strain Cs1-4 possessed a similar set of genes. The nic operon in strain Cs1-4 included a Nic regulator (marR, DelCs1_4781), an MFS transporter (nicT, DelCs1_4782), nicotinate dehydrogenase subunit B (nicB2, DelCs1_4783), nicotinate dehydrogenase subunit A (nicA, DelCs1_47840), salicylate 1-monoxygenase (nicC, DelCs1_4785), maleate isomerase (nicE, DelCs1_4786), N-formylmaleamate deformylase (nicD, DelCs1_4787), 2,5-dihydroxyppyridine 5,6-dioxygenase (nicX, DelCs1_4788), malemate amidohydrolase (nicF, DelCs1_4789) and an aldehyde dehydrogenase (nicB1, DelCs1_4790).

Pesticide degradation: Dicamba and Fenitrothion

Two gene clusters in the genome of strain Cs1-4 resembled clusters identified in other proteobacteria that coded for the metabolism of Dicamba (2-methoxy-3,6-dichlorobenzoate) or Fenitrothion (O,O-dimethyl O-3-methyl-4-nitrophenyl phosphorothioate). The Dicamba cluster in strain Cs1-4 encompassed a 0-demethylase that was composed of an oxygenase (DelCs1_5158), ferredoxin (DelCs1_5157) and reductase (DelCs1_5158), and the predicted product of this cluster would be 3,6-dichlorosalicylate. The pathway for 3,6-dichlorosalicylate metabolism is not well delineated, but strain Cs1-4 has putative salicylate oxygenases that could mediate ring hydroxylation (e.g., products of DelCs1_3111, DelCs1_3828).

For fenitrothion, biodegradation could occur via methyl hydroquinone as an intermediate, as it possessed a cluster of genes with similarity to those encoding fenitrothion metabolism in Burkholderia sp. NF100 [25]. These genes putatively encoded were a flavoprotein monoxygenase (mhqA, DelCs1_5563), extradiol dioxygenase (mhqB, DelCs1_5562), a ring hydroxylation oxygenase (DelCs1_5559), ferredoxin (DelCs1_5558) and hydrolase (mhqA, DelCs1_5554).

Conclusions

Determination of the complete genome sequence of D. acidovorans Cs1-4 achieved the objective of enabling new insights into the genes underlying PAH metabolism and evolutionary mechanisms that may shape the process in the environment. Furthermore, the genome sequence data suggested that biodegradative capacity of D. acidovorans Cs1-4 extended beyond PAH, and the organism was well equipped for growth in soils contaminated by a variety of compounds such as N-heterocycles and pesticides.
Abbreviations
PAH: Polycyclic aromatic hydrocarbons; RHD: Ring hydroxylating dehydrogenase; ANI: Average nucleotide identity; NpA: Nanopod protein A; DOE JGI: Department of Energy Joint Genome Institute; T6SS: Type VI secretion system; LPS: Lipopolysaccharide; PLFA: Phospholipid fatty acid.

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
Authors affiliated with the US-DOE were involved in aspects of genome production including sequencing (SL, J-FC, TW), finishing (AL, LAG, HT, CH, RT, PSCG), annotation (ML, LH, NK, NI, IP, NM) and Genbank submission (SP, LP). All other authors were involved in development of the initial proposal for sequencing of the D. acidovorans Cs1-4 genome (ARS, PSCG, VJD, WJH) and/or manuscript preparation (ARS, VG, CCO, WJH). All authors read and approved the final manuscript.

Acknowledgements
Sequencing, assembly and computational annotation of the D. acidovorans Cs1-4 genome was done by U.S Department of Energy, Joint Genome Institute through the Community Sequencing Project (CP9795673 to William J. Hickey). The work conducted by the U.S. Department of Energy, Joint Genome Institute was supported by U.S. Department of Energy, Office of Science under contract No. DE-AC02-05CH11231. Research in the corresponding author’s laboratory on genomic and phenotypic characteristics of D. acidovorans Cs1-4 was supported by the O.J. Allen Professorship in Soil Microbiology (to William J. Hickey).

Author details
1. O.J. Allen Laboratory for Soil Microbiology, Department of Soil Science, University of Wisconsin-Madison, Madison, WI 53706, USA. 2. Department of Food Production, University of the West Indies, St. Augustine, Trinidad and Tobago. 3. Department of Microbiology, University of Lagos, Lagos, Nigeria. 4. DOE Joint Genome Institute, Walnut Creek, CA, USA. 5. Algorithmic Biology Lab, St. Petersburg Academic University, St.Petersburg, Russia. 6. Oak Ridge National Laboratory, Oak Ridge, TN, USA. 7. Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA.

Received: 15 July 2014 Accepted: 15 July 2015
Published online: 15 August 2015

References
1. vacca DJ, Beam WF, Hickey WJ. Isolation of soil bacteria adapted to degrade humic acid-sorbed phenanthrene. Appl Environ Microbiol. 2005;71:3797–805.
2. Shetty A, Chen S, Tsocheva EI, Grant J, Hickey WJ. Nanopods: a new bacterial structure and mechanism for deployment of outer membrane vesicles. PLOS ONE. 2011;6, e20725.
3. Shetty A, Hickey WJ. Effects of outer membrane vesicle formation, surface-layer production and nanopod development on the metabolism of phenanthrene by Delfta acidovorans Cs1-4. PLoS ONE. 2014;9, e92143.
4. Stolz A. Molecular characteristics of xenobiotic-degrading sphingomonads. Appl Microbiol Biotechnol. 2009;81:793–811.
5. Moser R, Stahl U. Insights into the genetic diversity of initial dioxygenases from PAH-degrading bacteria. Appl Microbiol Biotechnol. 2001;55:609–18.
6. Hickey WJ, Chen S, Zhao J. The phi2 island: a new genomic island encoding catabolism of polyaromatic aromatic hydrocarbons. Front Microbiol. 2012;3:1–15.
7. Habe H, Omori T. Genetics of polycyclic aromatic hydrocarbon metabolism in diverse aerobic bacteria. Biosci Biotechnol Biochem. 2003;67:225–43.
8. Top EM, Springadi D. The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds. Curr Opin Biotechnol. 2003;14:262–9.
9. Bligh EG, Dyer WJ. A rapid method of total lipid extraction and purification. Can J Biochem Physiol. 1959;37:911–7.
10. Yi EC, Hackett M. Rapid isolation method for lipopolysaccharide and lipid A from Gram-negative bacteria. Analyst. 2000;125:651–6.
11. Gallagher S, Winston SE, Fuller SA, Hurrell JRG. Immunoblotting and Immunoelectrophoretic. In: Ausubel R, Bent RE, editors. Protocols in Molecular Biology. New York: John Wiley & Sons; 2008. p. 10.8.1–10.8.28.
12. Wilson K. Preparation of genomic DNA from bacteria. In: Ausubel R, Bent RE, editors. Current Protocols in Molecular Biology. New York: John Wiley & Sons; 2001. p. 2.4.1–5.
13. Schleheck D, Krequer TP, Fischer K, Cook AM. Mineralization of individual congeners of linear alkybenzenesulfonate by defined pairs of heterotrophic bacteria. Appl Environ Microbiol. 2004;70:6563–63.
14. Konstantinidis KT, Ramette A, Tiedje JM. The bacterial species definition in the genomic era. Phil Trans R Soc B. 2006;361:1929–40.
15. Kim M, Oh HS, Park SC, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int J Syst Evol Microbiol. 2014;64:364–51.
16. Eisen J, Heidelberg J, White O, Salzberg S. Evidence for symmetric chromosomal inversions around the replication origin in bacteria. Genome Biol. 2000;1:1–9.
17. Ravathan R, Studer S, Zehnder AJ, van der Meer JR. Int-B13, an unusual site-specific recombinase of the bacteriophage P4 integrase family, is responsible for chromosomal insertion of the 105-kilobase clc element of Pseudomonas sp. strain B13. J Bacteriol. 1998;180:505–14.
18. Basler M, Ho BT, Melakalos JJ. Ti/tar: Type VI secretion system counterattack during bacterial cell-cell interactions. Cell. 2013;154:884–94.
19. Tischler D, Grönig JAD, Kaschabak SB, Schömann M. One-component tyrosine monooxygenases: an evolutionary view on a rare class of flavoproteins. Appl Biochem Biotechnol. 2012;167:931–41.
20. Teufel R, Mascarei V, Ismail W, Voss M, Pereira J, Eisenreich W, et al. Bacterial phenylalanine and phenylacetate catalytic pathway revealed. Proc Natl Acad Sci U S A. 2010;107:14390–5.
21. Luengo JM, García JL, Olivera BR. The phenylacetyl-CoA catalyze a complex catalytic unit with broad biotechnological applications. Mol Microbiol. 2001;39:1434–42.
22. Hanwood CS, Burchhardt G, Herrmann H, Fuchs G. Anaerobic metabolism of aromatic compounds via the benzooyl-CoA pathway. FEMS Microbiol Rev. 1998;22:439–58.
23. Jiménez JL, Canales A, Jiménez-Barbero J, Ginalski K, Leszek R, Garcia JL, et al. Deciphering the genetic determinants for aerobic nicotinic acid degradation: the nic cluster from Pseudomonas putida KT2440. Proc Natl Acad Sci U S A. 2008;105:11329–34.
24. Wang X-Z, Li B, Herman PL, Weeks DP. A three component enzyme system catalyzes the O-demethylation of the herbicide Dicamba in Pseudomonas maltophilia D1-6. Appl Environ Microbiol. 1997;63:1623–6.
25. Tago K, Sato J, Takeha K, Kawagishi H, Hayatsu M. Characterization of methylhydroquinone-metabolizing oxygenase genes encoded on plasmid in Burkholderia sp. NF100. J Biosci Bioeng. 2005;100:157–23.
26. Woese CR, Kandler O, Woese RI. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria and Eucarya. Proc Natl Acad Sci U S A. 1990;87:4576–9.
27. Garrity GM, Bell JA, Pylum LT, WV. Proteobacteria phyl. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. Bergey’s Manual of Systematic Bacteriology, vol. 2, 2nd ed. Springer, New York; Part C; 2005. p. 1.
28. Garrity GM, Bell JA, Librum T, Class II. Betaproteobacteria class. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. Bergey’s Manual of Systematic Bacteriology, Second Edition, Volume 2, Part C. Springer: New York; 2005. p. 575.
29. Validisation List No. 107. List of new names and new combinations previously effectively, but not validly, published. Int J Syst Evol Microbiol. 2006;56:1–6.
30. Garrity GM, Bell JA, Librum T, Order I. Burkholderiales ord. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. Bergey’s Manual of Systematic Bacteriology, vol. 2, 2nd ed. Springer, New York; Part C; 2005. p. 575.
31. Willems A, De Ley J, Gills M, Kerssts K. Comamonadaceae, a new family encompassing the acidovorans RNA complex, including Kiovovorax paradoxus gen. nov., comb. nov., for Alcaligenes paradoxus (Davis 1969). Int J Syst Bacteriol. 1991;41:445–50.
32. Wen A, Fegan M, Hayward C, Chakraborty S, ly. Li. Phylogenetic relationships among members of the Comamonadaceae, and description of Delto acidovorans (den Dooren de Jong 1926 and Tamaka et al. 1981) gen. nov., comb. nov. Int J Syst Evol Microbiol. 1999;49:567–76.
33. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. The gene ontology consortium. Nat Genet. 2000;25:25–9.