

# Species and metabolic pathways involved in bioremediation of Vietnamese soil contaminated with Agent Orange. *Bordetella petrii* emerges as a key player in degradation of 2,4 dichlorophenoxyacetic acid

Thi Lan Anh Nguyen<sup>1</sup>, Thi Cam Ha Dang<sup>2</sup>, Jacco Koekkoek<sup>1</sup>, Martin Braster<sup>1</sup>, John Parsons<sup>3</sup>, Bram Brouwer<sup>1</sup>, Tjalf de Boer<sup>4</sup>, and R.J.M. van Spanning<sup>1</sup>

<sup>1</sup>Vrije Universiteit Amsterdam

<sup>2</sup>Vietnam Academy of Science and Technology

<sup>3</sup>Universiteit van Amsterdam

<sup>4</sup>MicroLife Solutions

July 16, 2020

## Abstract

Four bacterial strains were isolated from enrichment cultures inoculated with soil from Bien Hoa military base in Vietnam contaminated with the herbicides 2,4-dichlorophenoxyacetate (2,4-D) and 2,4,5-trichlorophenoxyacetate (2,4,5-T). They were classified as *Pseudomonas aeruginosa* BT1 2.2, *Sphingomonas histidinilytica* BT1 5.2, *Bordetella petrii* BT1 9.2, and *Achromobacter xylosoxidans* BT1 10.2, respectively. All 4 of them were able to degrade 2,4-D and 2,4,5-T during cultivation, but only the last 3 species used them as sole sources of carbon and free energy. We obtained a comprehensive insight into their degradation pathways by genomic analysis of these strains. A gene cluster with *tfdCDEF* genes was found in *A. xylosoxidans* BT1 10.2. The gene organization along with the amino acid sequences of the gene products are almost identical to those in *B. petrii* DSM12804. The *B. petrii* BT1 9.2 strain that we isolated has a full complement of the *tfdABCDEF* genes. Surprisingly, the gene organization along with the amino acid sequences of the gene products are virtually identical to those of *Cupriavidus pinatubonensis* JMP134, referred to as type I *tfd* genes, and clearly different from those of *A. xylosoxidans* and *B. petrii* DSM12804. Altogether, our enrichment approach has successfully resulted in boosting 3 different types of proteobacterial species that are equipped with metabolic pathways to use the herbicides as sole sources of carbon and free energy. We hypothesize that some of the corresponding genetic potential may have been recruited in recent mating events between these species and other members of the  $\beta$ - and  $\gamma$ -proteobacteria.

## 1. Introduction

During the Vietnam War from 1961 to 1971, Agent Orange, which is an herbicide and defoliant chemical, was used by the U.S. military army as part of its chemical warfare program. It is a mixture of equal parts of two herbicides, 2,4-D and 2,4,5-T (Stellman, Stellman, Christian, Weber, & Tomasallo, 2003; Westing, 1984). Many different types of bacterium isolated from different environments such as *C. pinatubonensis* JMP 134 (Filer & Harker, 1997; Laemmli, Leveau, Zehnder, & van der Meer, 2000), *Sphingomonas* sp. TFD44 (Smejkal, Vallaes, Burton, & Lappin-Scott, 2001; Thiel, Kaschabek, Gröning, Mau, & Schlömann, 2005), *A. xylosoxidans* subsp. *denitrificans* EST4002 (Vedler, Vahter, & Heinaru, 2004), *Bradyrhizobium* sp. HW13 (Kitagawa et al., 2002), *P. aeruginosa* PA01c and *Halomonass* sp. EF43 (Kleinstüber, Müller, & Babel, 2001) have the ability to degrade 2,4-D. They have been isolated mainly from soils contaminated with herbicides (Ka, Holben, & Tiedje, 1994; Tonso, Matheson, & Holben, 1995) and further genetic studies revealed their

possession of key genes encoding the enzymes for 2,4-D degradation. These bacteria have been classified into two groups based on their phylogeny and genetic potential to express the 2,4-D-degrading enzymes (Serbent, Rebelo, Pinheiro, Giongo, & Tavares, 2019). Group I bacteria belong to the subdivision of  $\beta$ - and  $\gamma$ -proteobacteria and include species from the genera *Achromobacter*, *Burkholderia*, *Delftia*, *Halomonas*, *Pseudomonas*, *Ralstonia*, *Rhodoferrax*, and *Variovorax* (Hoffmann, Kleinstuber, Müller, & Babel, 2003; Itoh et al., 2004). They usually have a full complement of a *tfdABCDEF* gene cluster for 2,4-D degradation. Group II bacteria include members of  $\alpha$ -proteobacteria, closely related to the genera *Sphingomonas* and *Bradyrhizobium* (Kitagawa et al., 2002; Zharikova, Iasakov, Zhurenko, Korobov, & Markusheva, 2018). Their organization of *tfd* genes is different from that of group I species and, importantly, *Sphingomonas* possesses a so-called *cadABCD* gene cluster encoding the subunits of a ring hydroxylating dioxygenase that replaces the function of TfdA (Kijima, Mita, Kawakami, & Amada, 2018).

The first description of genes involved in degradation of 2,4-D was based on studies on the  $\beta$ -proteobacterium *C. pinatubonensis* (strain JMP 134 / LMG 1197, previously termed *C. necator*, *Ralstonia eutropha* and originally *Alcaligenes eutrophus*), showing that it contained a so called *tfdABCDEF* gene cluster expressing the enzymes for the degradation of 2,4-D (Don & Pemberton, 1981; Streber, Timmis, & Zenk, 1987; Vallaeys, Fulthorpe, Wright, & Soulas, 1996). Subsequently, it was revealed that this species even had 2 copies of this gene cluster located adjacent to one another on plasmid pJP4 (Laemmli et al., 2000). The gene *tfdA* encodes an  $\alpha$ -ketoglutarate-dependent 2,4-dichlorophenoxyacetate dioxygenase, which catalyses the initial cleavage of the acetate side chain of 2,4-D into 2,4-dichlorophenol (2,4-DCP) (Fukumori & Hausinger, 1993a, 1993b). Subsequently, 2,4-dichlorophenol 6-monooxygenase which is encoded by *tfdB* converts 2,4-DCP into 3,5-dichlorocatechol (3,5-DCC). The substituted catechol is then sequentially degraded to 2-maleylacetate (2-MA) via an ortho-cleavage pathway by the enzymes encoded by the *tfdCDEF* genes, which include a chlorocatechol 1,2-dioxygenase, a dichloromuconate cycloisomerase, a carboxymethylenebutenolidase, and a maleylacetate reductase, successively (Laemmli et al., 2000; Liu & Chapman, 1984; Perkins, Gordon, Caceres, & Lurquin, 1990). 2-MA is then degraded to succinyl-CoA and acetyl-CoA in a sequence of reactions involving the products of the *tfdF* and *pcaIJF* genes, the last 3 of which encode a 3-oxoadipate CoA-transferase (PcaIJ) and a 3-oxoadipyl-CoA thiolase (PcaF), successively. Succinyl-CoA and acetyl-CoA may fuel the tricarboxylic acid (TCA) cycle. A list of enzymes according to the KEGG database is in supplementary Table S1.

In contrast to the diversity of 2,4-D-degrading bacteria, only a few aerobic 2,4,5-T degraders have been reported. The best known are members of genera *Burkholderia* (Danganan, Ye, Daubaras, Xun, & Chakrabarty, 1994; Huong, Itoh, & Suyama, 2007; Kellogg, Chatterjee, & Chakrabarty, 1981), *Nocardioide*s (Golovleva, Pertsova, Evtushenko, & Baskunov, 1990), *Sphingomonas* (Huong et al., 2007), and *Bradyrhizobium* (Hayashi, Sano, Suyama, & Itoh, 2016; Huong et al., 2007; Rice, Menn, Hay, Sanseverino, & Sayler, 2005), all of which were capable of utilizing 2,4,5-T as sole carbon and free energy source. *Burkholderia cepacia* AC1100 has 3 gene clusters involved in degradation of 2,4,5-T. The first one has *tftAB* genes encoding the two subunits of 2,4,5-trichlorophenoxyacetic acid oxygenase that converts 2,4,5-T to 2,4,5-trichlorophenol (2,4,5-TCP). The latter compound is further oxidized to 2,5-dichloro-p-benzoquinone (2,5-DCBQ) by a chlorophenol-4-monooxygenase encoded by the *tftCD* genes on the 2<sup>nd</sup> cluster (Danganan et al., 1994; Daubaras, Hersberger, Kitano, & Chakrabarty, 1995; Gisi & Xun, 2003; Hübner, Danganan, Xun, Chakrabarty, & Hendrickson, 1998; Zaborina et al., 1998). An unknown factor then converts DCBQ into 2,5-dichlorohydroquinone (2,5-DCHQ). The 3<sup>rd</sup> gene cluster, *tftEFGH*, encodes the enzymes for converting DCHQ to 3-oxoadipate (3-OXA). First, 2,5-dichlorohydroxyquinone reductive dehydrochlorinase (TftG) dechlorinates 2,5-DCHQ to chlorohydroquinone (CHQ) and then to hydroquinone (HQ) (Zaborina et al., 1998). Subsequent ring cleavage is then catalysed by hydroxyquinone 1,2-dioxygenase (TftH) to yield 4-hydroxymuconic semialdehyde (4-HMSA) which then reacts to maleylacetate (2-MA). The latter is converted to 3-OXA by a maleylacetate reductase encoded by the *tftE* gene, an orthologue of *tfdF*. *S. paucimobilis* strain B90 has LinE and LinD proteins that differ from TftGH proteins, yet their substrate and ultimate product are the same (Kumari et al., 2002). Here the Tft and Tfd degradation pathways merge, ultimately resulting in formation of succinyl-CoA and acetyl-CoA (Daubaras, Danganan, et al., 1996; Daubaras et al., 1995). In yet another

parallel pathway, 2,5-DCHQ is converted to 2-MA via the enzymes chlorophenol-4-monooxygenase (Tft-CD), 2-hydroxy-1,4-benzoquinone reductase (HbqR) and hydroxyquinol 1,2-dioxygenase (ChqB) (Ferraroni *et al.* 2005). These enzymes are found in f.e. *Burkholderia* and *Nocardioide*s species (Daubaras, Saido, & Chakrabarty, 1996; Ferraroni *et al.*, 2005; Takenaka *et al.*, 2011).

In this study, we isolated indigenous bacterial strains originally present in soil heavily contaminated with Agent Orange in Vietnam, all of which were dominant in enrichment cultures with 2,4-D and 2,4,5-T as sole sources of carbon and free energy, and both of which had genes involved in breakdown of these compounds. Here, we aimed at getting a more fundamental understanding of their physiological and genetic potential in that metabolic process. The approach in this study was to i) to culture the indigenous strains under defined conditions with 2,4-D and 2,4,5-T as sole carbon and free energy sources, ii) to monitor the degradation of these herbicides in time along with increases in cellular biomass, and iii) to unravel their genome sequences in order to obtain insight in their metabolic potential. These efforts resulted in an integrative view of species and pathways involved in the degradation of these chlorinated phenoxyacetates.

## 2. Materials and methods

### 2.1 Culturing conditions and isolation of bacterial strains

Herbicide contaminated soil for enrichment studies was collected from Bien Hoa airbase (10°58'14.3"N 106°48'19.3"E), Dong Nai Province, Vietnam. Ten grams of herbicide contaminated soil was added in basal salt medium (BS, 40 ml per culture) containing  $\text{KH}_2\text{PO}_4$  0.5 g/l,  $(\text{NH}_4)_2\text{SO}_4$  0.25 g/l,  $\text{MgSO}_4$  0.2 g/l,  $\text{CaCl}_2$  0.5 g/l and  $\text{NaNO}_3$  0.4 g/l (van der Zaan *et al.*, 2012), and they were supplemented with 100 mg/l 2,4-D and 100 mg/l 2,4,5-T. Both compounds were purchased from Sigma, more than 95% pure, and used as carbon- and free energy source during degradation assays. Cultures were incubated in the dark on a rotary shaker at 30°C and 200 rpm (the 1<sup>st</sup> enrichment). After cultivation for 26 days, these cultures were then used to inoculate a 2<sup>nd</sup> enrichment with 200 mg/l 2,4-D and 100 mg/l 2,4,5-T and further incubated for 25 days. Colony forming units (CFUs) were isolated on nutrient agar at 30°C by serial dilution of the cultures at different time points after the start of the experiment. Nutrient agar (NA) medium contained 1.0 g peptone, 1.0 g meat extract, 0.5 g NaCl, and 1.5 g agar per 100 ml. Bacterial strains that were isolated as dominant CFUs were pre-cultured overnight in nutrient broth (NB) medium. The cells were harvested by centrifugation at 10,000 rpm for 5 min, washed 2 times with sterilized BS and transferred into BS medium supplemented with 200 mg/l 2,4-D and 100 mg/l 2,4,5-T as inoculum for the degradation experiment under the same conditions as those for the enrichment cultures. A flask with dead cells amended with the herbicides was used as control. All experiments were independently performed in triplicate. Cell growth was determined by measuring the absorbance of cultures at 600 nm using a spectrophotometer (Multiskan GO, Thermo Scientific, U.S). Concentration of 2,4-D and 2,4,5-T in all culture samples were determined in time using tandem liquid chromatography mass spectrometry (LC-MS/MS) (Aligent Technologies, 1200 Series).

### 2.2 Analysis of 2,4-D and 2,4,5-T by LC-MS/MS

The 2,4-D and 2,4,5-T-degrading ability of bacterial communities was assessed by detecting the remaining concentration of herbicides by tandem liquid chromatography mass spectrometry (LC-MS/MS). Data analyses of the biodegradation tests were performed using Prism version 8.2.1. The herbicides were extracted using the QuEChERS (Quick, Easy, Cheap, Effective, Rugged, and Safe) method (Rejczak & Tuzimski, 2015). Briefly, aliquots of the cultures (2 ml) were transferred to 15-ml Falcon tubes, after which 2 ml of acetonitrile with 1% formic acid was added and vortexed for 15 min. A mixture of salts (2 g of  $\text{MgSO}_4$ , 0.5 g of NaCl, 0.5 g of  $\text{C}_6\text{H}_5\text{NaO}_7 \cdot 2\text{H}_2\text{O}$ , and 0.25 g of  $\text{C}_6\text{H}_6\text{Na}_2\text{O}_7 \cdot 1.5 \text{H}_2\text{O}$ ) was added to the homogenate and mixed for 5 min, followed by centrifugation for 5 min at 10,000 rpm. Then, 2 ml of the top layer was collected for the LCMS/MS analysis. The remaining concentrations of 2,4-D and 2,4,5-T were measured on an Elute UHPLC system coupled to an EVOQ triple quadrupole mass selective detector (both Bruker, Bremen Germany). The target compounds were separated on a 50 x 2 mm, 3  $\mu\text{m}$  Luna C18 column (Phenomenex, Utrecht the Netherlands), applying a gradient of 0.1% formic acid and methanol (both Biosolve, Valkenswaard the Netherlands). The TQ-MS was applying an electro spray ionization in a negative ion mode. The transitions

for 2,4-D were 221 > 163 and 219 > 161 (quantifier and qualifier, respectively). For 2,4,5-T the transitions were 253 > 195 and 255 > 197 (quantifier and qualifier, respectively). Before analysis, samples were diluted in deionized water and store at 4°C until injection. Data acquisition and analyses were performed with MS Data review software.

### 2.3 DNA isolation and 16S rRNA gene sequencing of isolated strains

The bacterial isolates were identified based on 16S rRNA gene sequence analysis. DNA of bacteria was extracted using the MoBio PowerSoil® DNA Isolation kit (Carlsbad, CA, USA) according to the manufacturer's instructions. Amplification was carried out with universal primers 8F (5'- AGAGTTTGATYMTGGCTCAG -3') and 1512R (5'- ACGGYTACCTTGTTACGACTT-3') as described previously (Weisburg, Barns, Pelletier, & Lane, 1991). Reactions were performed in a Thermocycler (Biometra, Analytik Jena, Germany). PCR products were purified and sequenced by MacroGen Europe. These sequences were compared with known 16S rRNA gene sequences deposited in the GenBank database using the BLAST search at the National Center for Biotechnology Information (<http://www.ncbi.nlm.gov.BLAST/>).

### 2.4 Genomic DNA extraction, genome sequencing, assembly, and annotation

Genomic DNA was extracted using the MoBio PowerSoil® DNA Isolation kit (Carlsbad, CA, USA). The quality and quantity of the extracted genomic DNA were checked by agarose gel electrophoresis and Qubit dsDNA HS Assay kit (ThermoFischer Scientific, cat. no. Q32851). Genome sequencing was carried out by BaseClear B.V (the Netherlands) according to their in-house protocol. Single-end or paired-end sequence reads were generated using the Illumina NovaSeq 6000 or MiSeq system. The sequences generated with the MiSeq system were performed under accreditation according to the scope of BaseClear B.V. (L457; NEN-EN-ISO/IEC 17025). When paired-end sequencing is being performed, the "Number of reads" noted in Table S2 is referring to read pairs. FASTQ read sequence files were generated using bcl2fastq2 version 2.18. Initial quality assessment was based on data passing the Illumina Chastity filtering. Subsequently, reads containing PhiX control signal were removed using an in-house filtering protocol. In addition, reads containing (partial) adapters were clipped (up to a minimum read length of 50 bp). The second quality assessment was based on the remaining reads using the FASTQC quality control tool version 0.11.5. Table 1 shows essential data of the sequencing and assembly of the genomes of the 4 selected isolates.

The raw Illumina reads were used directly for genome assembly using SPAdes (version 3.13.0). SPAdes is a De Bruijn graph assembler for short reads and includes its own read error correction (Bankevich et al., 2012). The following SPAdes settings were used: -careful, -kmer: 55,77,127. Assembly quality was determined using QUAST (Gurevich, Saveliev, Vyahhi, & Tesler, 2013) where we determined the number of scaffolds, total genome length and N50 as a measurement for genome completeness. We found that removal of scaffolds smaller than 5000 bps did not significantly impact the total assembled genome length so to better facilitate annotation these scaffolds were removed. The scaffold size adjusted genomes were annotated using a number of programs and databases. Gene prediction and initial annotation was done using Prokka (Seemann, 2014). Predicted genes were further annotated using Interproscan (Pfam and hamap databases) and the CAZy carbohydrate activated enzyme database (all families) (Jones et al., 2014; Lombard, Golaconda Ramulu, Drula, Coutinho, & Henrissat, 2014). Signal peptides were predicted with SignalP and secondary metabolite clusters using AntiSMASH (Almagro Armenteros et al., 2019; Weber et al., 2015). The closest homologues species was determined by BLAST analysis using the Clusters of Orthologues Groups (COG) database for each predicted gene separately and by taking the species with the most hits as closest species (Tatusov, Galperin, Natale, & Koonin, 2000). The NCBI-annotated genomes of BT1 2.2 (PRJNA586721), BT1 5.2 (PRJNA586745), BT1 9.2 (PRJNA586752), and BT1 10.2 (PRJNA587043) were deposited in GenBank with their Bioproject numbers listed in between brackets.

### 2.5 Phylogeny

The amino acid sequences of TftA homologues from the 4 isolated bacteria were compared with reference sequences available in the NCBI and KEGG database. The sequences were aligned and a phylogenetic tree (Fig. 3) was visualized using the MEGA version 7 software by the neighbor-joining method with 1,000

bootstrap replicates (Kumar, Stecher, & Tamura, 2016).

### 3. Results

#### Isolation and classification of 2,4-D- and 2,4,5-T-degrading bacteria

Fourteen different types of bacterial colony were isolated from herbicide contaminated soil communities enriched in cultures with 2,4-D and 2,4,5-T as sole sources of carbon and free energy. These strains varied in shape and colour of colony on nutrient agar plates. The results indicated that the dominant isolates were from genera *Rhodanobacter*, *Arthrobacter*, *Pseudomonas*, *Achromobacter*, *Sphingomonas*, *Bosea*, *Pedobacter*, *Sphingopyxis*, *Staphylococcus* and *Ancylobacter* (Table 2). Among these isolates, 4 Gram-negative strains that popped up as the most dominant species in the enrichments were selected to study their herbicide biodegradation capacities and their genetic make-up to achieve that. BLAST results showed that the 16S rRNA sequences of the isolates BT1 2.2, BT1 5.2, BT1 9.2, and BT1 10.2 were highly similar (>99% identity) to *P. aeruginosa*, *S. histidinilytica*, *B. petrii*, and *A. xylosoxidans*, respectively. Therefore, we classified them as *P. aeruginosa* BT1 2.2, *S. histidinilytica* BT1 5.2, *B. petrii* BT1 9.2, and *A. xylosoxidans* BT1 10.2, respectively.

#### Biodegradation of 2,4-D and 2,4,5-T by four strains isolated from enrichment cultures

We set out a growth experiment where the isolates were cultured in basal salt medium supplemented with 2,4-D and 2,4,5-T as sole sources of carbon and free energy. We determined their growth curves (Fig. 1A) and the removal of 2,4-D and 2,4,5-T in time (Figs 1B and C). *P. aeruginosa* BT1 2.2 showed poor growth with a small increase in biomass after 7 days of cultivation, which subsequently decreased again after 10 days, most likely as a result of cell death and lysis. It did, however, show complete removal of 2,4-D and 2,4,5-T within 10 days of cultivation, but we have no data on the products that were formed. The other 3 strains showed substantial growth and complete removal of the herbicides within 10 days of cultivation. The fastest growers were *B. petrii* BT1 9.2 and *A. xylosoxidans* BT1 10.2, but the latter one had a lag time of 2 days before it started growing. Also, its final biomass was slightly less than that of *B. petrii* BT1 9.2. In both cases, maximum growth was reached after 4 days of cultivation. *S. histidinilytica* BT1 5.2 needed a bit longer to reach its maximum growth yield. Due to its apparent slower growth rate, maximum biomass was obtained after 7 days, but the final yield was comparable to that of *B. petrii* BT1 9.2. The latter was the best performer with respect to rates and yields of growth along with its ability to completely degrade 2,4-D and 2,4,5-T. In order to get a more comprehensive view on its capacities, we studied the responses of *B. petrii* BT1 9.2 in cultures with only 2,4-D or only 2,4,5-T and compared the growth data (Fig. 2A) and removal of the herbicides (Figs 2B and C) with those from the culture with a mixture of both herbicides. Fastest growth and highest yields were observed in the cultures with only 2,4-D as sole source of carbon and free energy. Both growth rates and yields were slightly less in the mixture of the 2 compounds, and the poorest performance with regard to rates and yield was observed with 2,4,5-T only. These growth properties coincided in part with the degradation rates of 2,4-D and 2,4,5-T. The former compound was removed within 2 days in the cultures with only 2,4-D, while it took 10 days in those with the mixture. Degradation of 2,4,5-T took 4 days of cultivation, while in the mixture this required 10 days.

#### Pathways for herbicide degradation

We inspected the 4 genomes for their genetic potential to degrade 2,4-D and 2,4,5-T to benzoate and further on to succinyl-CoA and acetyl-CoA. To this end, we gathered the amino acid sequences of all the enzymes identified in the KEGG database as being involved in these metabolic pathways. A list of our selected enzymes along with their properties and some additional parameters is shown in Supplementary Table S1. The KEGG database amino acid sequences of the selected genes were then used to look for similar sequences in the genome sequences of our 4 isolates. The script for these analyses is described in the materials and methods section. The results of these analyses for each organism are listed in Supplementary Table S3 A-D. Percentages identity between the KEGG protein sequences and those expressed from the selected genomes are listed as well. The threshold for selection of the orthologous proteins was set at a minimum of 45% identity. Importantly, all 4 species showed (initial) degradation of 2,4-D and 2,4,5-T so they should at least

have a *tfdA* gene to degrade 2,4-D into 2,4-DCP, and a *tftAB* gene set to encode the dioxygenase for converting 2,4,5-T into 2,4,5-TCP, or as yet unknown enzymes that could take over that role. A phylogenetic tree of TftA homologues from the 4 isolates is shown in Fig. 3 along with the reference enzymes from the KEGG and NCBI databases. These reference enzymes also include CadA, BenA and AntA as we had reason to believe that some of the homologues that we found group closer to these proteins than to the TftA group. As such we were able to judge the minimum potential of the 4 species to degrade 2,4-D and 2,4,5-T and their further metabolites ultimately to succinyl-CoA and acetyl-CoA. We first constructed the reference pathways for degradation of the herbicides using the KEGG database maps ‘chlorocyclohexane and chlorobenzene degradation’ and ‘benzoate degradation’. A graphical representation of these pathways is shown in Fig. 4. We then reconstructed the pathways in our 4 isolated species by identifying their sequences orthologous to those of the reference pathways (Fig. S1). Below is a further description of the 4 isolates with emphasis on their genetic potential to convert 2,4-D and 2,4,5-T.

*A. xylosoxidans* BT1 10.2 has its *tfdCDEF* genes clustered along with a regulatory *tfdS* gene located upstream and divergently transcribed from it (Fig. 5A). Downstream are genes encoding the subunits of a ring hydroxylating dioxygenase most similar to a 2-aminobenzenesulfonate 2,3-dioxygenase first described in *aAlcaligenes* species (Mampel, Ruff, Junker, & Cook, 1999). The catalytic part is composed of AbsAa and AbsAb and the electron donor is the flavodoxin, AbsAc. This gene organization and corresponding amino acid sequences are almost identical to those of *B. petrii* DSM 12804. This high degree of identity may be the result of a recent gene transfer event between ancestors of these species. Its *tfdB* gene is located elsewhere on the genome (ORF2511; Table S3). We did not find an obvious *tfdA* gene nor *tftAB* genes for the initial conversion of 2,4-D and 2,4,5-T, respectively. A likely candidate for *tfdA* is ORF 3265 (Table S3, 35% identity with the KEGG reference protein from *C. pinatubonensis*). Those for *tftAB* are ORFs 2313 and 2314 (around 32% identities with their KEGG counterparts from *B. cepacia*). Other genes potentially involved in further degradation of 2,4,5-T breakdown products are *hbqR* and *chqB* expressing HbqR and ChqB for the conversion of 2-HBQ via BT to 2-MA. Yet, the genes encoding the enzymes to convert 2,4,5-T into HBQ were not identified in this species.

*B. petrii* BT1 9.2 has a full complement of *tfdABCDEF* genes encoding the enzymes for the sequential conversion of 2,4-D into 2-MA (Fig. 5A). They are clustered and transcribed in the order *tfdCDEFB* with a *lysR* -type *tfdS* homologue upstream and divergently transcribed from that. A bit more downstream separated by 4 genes and with the transcription direction opposed to the other structural genes is the *tfdA* gene. One of the four genes sandwiched by the *tfdA* and remaining *tfd* genes is a gene encoding a PcaF-type enzyme responsible for cleavage of 3-oxoadipate-CoA into succinyl- and acetyl-CoA. The identity of the *tfdBCDEF* encoded proteins with the KEGG reference proteins from *C. pinatubonensis* JMP134 (accession no. M35097.1) is remarkably high, all of which ranging from 99 to 100%. Also, the gene organizations are identical suggestive of some kind of horizontal gene transfer between ancestors of these 2  $\beta$ -proteobacteria during recent evolution. We also compared the genetic organization of the isolated Vietnamese species with the one with its genome sequence deposited in the databases, *B. petrii* DSM 12804 (accession no. AM902716.1). Their gene organizations are different (Fig. 5A). Also, the pairwise comparisons show that the identities are much lower than those between *B. petrii* BT1 9.2 and *C. pinatubonensis* (Table S4). A second copy of the *tfdA* gene, not clustered with other *tfd* genes and with the expressed protein having 45% identity to the reference TfdA, was found elsewhere on the genome (ORF 3407). It has also the *hbqR* gene so it may be able to convert 2-HBQ to BT. However, it lacks reasonable homologues of enzymes to produce the substrate or to consume the product. Also, the *hbqR* gene is not clustered with other genes involved in 2,4,5-T degradation. We did not find genes with high identity to *tftAB* genes but as this strain has the potential to perform the initial ring hydroxylating dioxygenase step, we assume that proteins related to the reference TftAB proteins are responsible for this reaction. Likely candidates to take over the role of TftA, all with identities to the reference TftA close to 30% are ORFs 524, 1633, 2507 and 3859. Their genes are all accompanied with homologues of *tftB*.

*S. histidinilytica* BT1 5.2 has the genetic potential to express all enzymes for degradation of 2,4-D to succinyl-CoA and acetyl-CoA. Importantly, it does not have an apparent *tfdA* gene. The function of the *tfdA* gene

product is likely taken over by a ring hydroxylating dioxygenase encoded by a *cadABCD* gene cluster that is sandwiched between the other *tfd* genes (Fig. 5B). A similar gene organization is seen in *Sphingomonas* sp. TFD44 (accession no. AY598949.1), *Sphingobium herbicidovorans* MH (accession no. AJ628861.1) and *Sphingomonas* sp. ERG5 (accession no. KF494257.1) (Fig. 4C) (Nielsen et al., 2017; Nielsen et al., 2013). A pairwise comparison of the gene products of the latter with those from our isolated strain indicates that they share a high degree of identity (more than 95%) both on the level of gene cluster organization and on the level of their amino acid sequences (Table S4). Also, both clusters contain *pcaIJF* genes to make the enzymes for conversion of 2-MA, the product of the Tfd proteins, into succinyl-CoA and acetyl-CoA, hence the potential of making a full pathway from 2,4-D all the way to TCA-cycle intermediates. In addition to the gene cluster with *tfd* and *cad* genes, it has another one with copies of *tfdCDEF* genes on it, along with another set of *pcaIJF* genes. This one may be switched on for growth on chlorocatechols. Importantly, the suggested *tfdD* and *tfdE* genes in both clusters share a low identity with their counterparts from the KEGG reference proteins (in between 32 and 42%, see Table S3). Nevertheless, we regard them as *tfd* genes because of their occurrence in the gene clusters with the complementary *tfd* genes. Its versatility in degrading chlorinated aromatics is further exemplified by the observation that it has a *catABCD* gene cluster under apparent control of a *catR* regulatory gene (Fig. 5C). Its gene organization looks quite similar to the one of *P. aeruginosa* BT1 2.2. Pairwise identities of the CatRABC proteins from both species are around 50% (Table S4). Upstream of the *cat* gene cluster are 2 sets of genes that resemble those encoding the 2 subunits of the family of ring hydroxylating dioxygenases like AntAB, BenAB and TftAB. It is tempting to speculate that these enzyme systems convert their as yet unknown aromatic substrates into catechol for further degradation to maleylacetate by the Cat proteins. It does not have obvious *tftAB* genes, yet it is able to convert it at least to 2,4,5-TCP. Hence another enzyme must be encoded by *S. histidinilytica* BT1 5.2. It might be that the CadABCD proteins can do that as they belong to the same family of ring hydroxylating dioxygenases. Alternatively, that reaction may be carried out by the gene products of ORFs 5533 and 5532, which have around 30% identity with their counterparts TftA and TftB, respectively, from *B. cepacia* that serve as the KEGG database reference proteins.

*P. aeruginosa* BT1 2.2 has *catAB* genes potentially involved in 2,4-D breakdown. Their genes are clustered. A map of this gene cluster is in Fig. 5C. A closer look at the gene cluster encompassing the *catAB* genes shows some interesting features as they make part of a typical *catABC* gene cluster, the proteins of which are involved in catechol degradation to 3-OXA. Transcribed in the same direction and sandwiched by the *catB* and *catA* genes is a gene homologous to *catC* and upstream but divergently transcribed a *lysR*-type regulatory gene possibly equivalent to *catR* (McFall, Chugani, & Chakrabarty, 1998; Nojiri et al., 2002). A similar gene organization is found in *P. putida* mt-2 (Jiménez, Pérez-Pantoja, Chavarría, Díaz, & de Lorenzo, 2014). Gene clusters composed of *catRABC* genes but organized differently are also found in *R. erythropolis* CCM2595 (accession no. FM995530.1) and *Burkholderia* sp. TH2 (accession no. AB035483.1). More upstream are 2 gene clusters encoding the subunits of ring-hydroxylating oxygenases, both of which under apparent control of AraC-type regulators, their genes flanking these 2 gene clusters. The one most close to the *cat* gene cluster has 2 structural genes with highest similarity with *antABC* genes to make an anthranilate 1,2-dioxygenase, AntAB as the catalytic core and AntC as the flavodoxin-type electron donor. The one more downstream is most likely *abenABCD* gene cluster encoding a benzoate dioxygenase (BenABC) and a dihydroxy cyclohexadiene carboxylate dehydrogenase (BenD). *Burkholderia* sp. TH2 has also a set of *benABCD* genes adjacent to its *cat* gene cluster (Fig. 5C). The phylogenetic tree with TftA homologues including AntA and BenA proteins shows their position in the sub-groups, which supports their suggested functions. Both enzyme systems, AntABC and BenABCD, convert their substrates, anthranilate and benzoate, respectively, to catechol. All together, these findings suggest that *P. aeruginosa* BT1 2.2 is specialized in degradation of benzoate and related compounds as it has *ben* and *cat* gene clusters for the sequential conversion of benzoate via catechol to 3-oxoadipate. Genes encoding PcaIJ and PcaF are found elsewhere on the genome for further breakdown of 3-oxoadipate to succinyl-CoA and acetyl-CoA. It is able to convert 2,4-D at least into 2,4-DCP, and we speculate that this is achieved by the product of ORF3804, which has 38% identity with the KEGG reference *tfdA* from species *C. pinatubonensis* JMP134. An alternative gene set for *tftAB* genes might be ORFs 1030 and 1031. The protein from ORF 1030 has 30% identity with the

KEGG reference TftA from *B. cepacia* AC1100.

## 4. Discussion

We have identified 3 species of proteobacteria as key players in the degradation of 2,4-D and 2,4,5-T in Vietnamese soil contaminated with Agent Orange. They were identified as *S. histidinilytica* BT1 5.2, *B. petrii* BT1 9.2 and *A. xylosoxidans* BT1 10.2, the first is an  $\alpha$ -proteobacterium, the latter 2 are  $\beta$ -proteobacteria. All 3 were shown to grow and increase their numbers during enrichments on these herbicides, were isolated as dominant species at the end of culturing (Nguyen et al., unpublished data) and grew in batch cultures with 2,4-D and 2,4,5-T as sole sources of carbon and free energy. Physiological studies along with a detailed inspection of their genomes revealed that they have the potential to express the key enzymes for the sequential degradation of 2,4-D all the way to carbon dioxide. The pathways for degradation of 2,4,5-T in these 3 species are less clear.

The breakdown of 2,4-D by *B. petrii* and its use as carbon and free energy source is a novel finding as there are, to our knowledge, no studies that show degradation of 2,4-D or 2,4,5-T by species of the genus *Bordetella*. There is only one report published that unravels its genome sequence, which displays genes involved in herbicide degradation, but these were not correlated to physiological studies (Gross et al., 2008). Like *B. petrii* BT1 9.2, also *S. histidinilytica* BT1 5.2 can grow on the herbicides and has a comparable set of *tfd* genes. A distinctive feature, however, is the observation that it does not have a *tfdA* gene to make an  $\alpha$ -ketoglutaric acid dependent 2,4-D dioxygenase. Instead, the *tfdBCDEF* genes in its cluster sandwich a set of 4 so-called *cadABCD* genes encoding the subunits of a ring hydroxylating dioxygenase that replaces the TfdA function. This seems to be quite common within the  $\alpha$ -proteobacteria (Nielsen et al., 2017; Nielsen et al., 2013). In a previous study, we detected a *tftA* gene fragment using PCR with degenerate *tftA* primers (Anh et al., unpublished data). We now know that it does not have an obvious set of *tftAB* genes, and that the PCR fragment is in fact from *cadA*. This result further underscores the overall view that members of the family of ring hydroxylating dioxygenases, such as TftA, BenA and CadA, are quite similar to one another yet they have specific tasks in different reactions and pathways. The versatility of *S. histidinilytica* BT1 5.2 with regard to the breakdown of the herbicides is further emphasized by the observation that it has a 2<sup>nd</sup> gene cluster with *tfdCDEF* genes along with a regulatory *lysR*-type gene just as *tfdS* and again with *pcaIJF* genes. Also notable is the finding of a 3<sup>rd</sup> gene cluster with *catABCD* genes and a regulatory *catR* gene, which would allow it to grow on catechols as well and emphasizes its versatility in growth on aromatic compounds. These findings were also reported in previous studies with *Sphingomonas* (Cycoń, Żmijowska, & Piotrowska-Seget, 2011; Huong et al., 2007). Previously, *S. herbicidovorans* MH was reported for the degradation of phenoxyalkanoic acid herbicides (Kohler, 1999; Müller, Byrde, Werlen, van der Meer, & Kohler, 2004), but this strain could not degrade 2,4,5-T (Kohler, 1999). Just like *B. petrii* BT1 9.2 and *S. histidinilytica* BT1 5.2, also *A. xylosoxidans* BT1 10.2 can grow on 2,4-D and use it as sole source of carbon and free energy. Its genetic potential, however, is less clear as it does have *tfdBCDEF* genes along with *pcaIJF* genes, but we could not identify a *tfdA* homologue with more than 45% identity to the reference homologue in the KEGG database. In order to explain disappearance of 2,4-D and growth of *A. xylosoxidans* BT1 10.2 in the culture, we assume that it carries a gene encoding an alternative enzyme or that it makes use of a TfdA homologue with lower identity.

We noticed also the disappearance of 2,4,5-T in all of the cultures, yet we are unaware of the fate of its metabolites. We did not recognize in either of the species a full suite of genes encoding the key enzymes for its sequential degradation. Perhaps as yet unknown enzymes may be recruited for 2,4,5-T breakdown and bacterial growth as we did see some increase in biomass of *B. petrii* BT1 9.2 growing on 2,4,5-T. It is more likely, however, that at least the initial step in degradation of 2,4,5-T is catalysed by a member of the TftA family, but with a relatively low identity with the KEGG counterpart. Indeed, we show that pairwise identities indicate that the orthologous enzymes in some cases substantially differ from one species to another. Hence, the KEGG proteins may not identify those that have such low identity, yet are able to carry out the same reaction. This is particularly true for the TfdD and TfdE proteins. If one looks at the pairwise identities of the type I and type II proteins of *C. pinatubonensis* for example than it is noteworthy



that they share only 15% (TfdE) and 35% (TfdD) identities, respectively (Laemmli et al., 2000). Hence, we could use this observation of low identity as a valid argument that *S. histidinilytica* BT1 5.2 may well have counterparts of the *tfdD* and *tfdE* genes from *C. pinatubonensis*, although they share an identity less than 45%.

Not all dominant species isolated from the enrichment cultures were able to use the herbicides as sources of carbon and free energy. *P. aeruginosa* BT1 2.2 has no obvious set of *tfd* genes, which explains its poor growth on the herbicides. This is in contrast with other studies, which showed that *Pseudomonas* species were able to biodegrade 2,4-D (Marrón-Montiel, Ruiz-Ordaz, Rubio-Granados, Juárez-Ramírez, & Galíndez-Mayer, 2006; Yang et al., 2017). *P. aeruginosa* BT1 2.2 does have *catAB* genes, though, but these are clustered with a *catC* gene and a regulatory *catR* gene, suggestive for a role in catechol metabolism. That would make sense as the same cluster is preceded by *benABCD* and *antABC* gene clusters, allowing it to convert benzoate and anthranilate, respectively, to catechol. A possible reason for its dominance in the enrichment cultures may be that it feeds on metabolites of the herbicides in the enrichment cultures. Alternatively, it manages to survive by feeding on the contents of lysed cells. As such, the enrichment culture may be regarded as a food web with different trophic levels ranging from primary consumers to scavengers.

We further noticed high degrees of similarities of gene cluster organizations along with the amino acid sequences of their gene products (more than 95%) between those of *B. petrii* BT1 9.2 and *C. pinatubonensis* JMP134 (gene cluster I) on the one hand and those of *B. petrii* DSM12804 and *A. xylosoxidans* BT1 10.2 on the other hand. That observation suggests recent events of horizontal gene transfer between members of these  $\beta$ -proteobacteria. It is surprising that the cluster and its gene products from *B. petrii* that we isolated from Vietnamese soils is different from the one that was already deposited in the database (*B. petrii* DSM 12804). Apparently, exchange between *B. petrii* species with their mating partners were independent events. The gene clusters from *C. pinatubonensis* and *B. petrii* DSM 12804 are on plasmids, but we have no information about the genetic location of the *tfd* clusters from our isolates as we have only a set of contigs. Since the identities are so high, we suggest that gene exchange events occurred quite recently, and that these events supported growth on chlorinated aromatics such as 2,4-D and 2,4,5-T, which were produced by the American Chemical Paint Company only after the 1940's. As a consequence, it now allows *B. petrii* BT1 9.2 and *C. pinatubonensis* JMP134 to degrade 2,4-D all the way to 2-MA as they have a full complement of the *tfdABCDEF* genes. The clusters from *A. xylosoxidans* BT1 10.2 and *B. petrii* DSM 12804 have only *tfdCDEF* genes, hence their products serve to convert (chlorinated) catechol into 2-MA. Catechols are much older than the phenoxyacetates as they are also formed during wood fires and breakdown of other aromatics (Baldwin, Staszak-Kozinski, & Davidson, 1994; Tao, Fishman, Bentley, & Wood, 2004). Hence these clusters might become useful via genetic transfer in species that already have some TfdA and TfdB activity to boost the 2,4-D degrading pathway. Altogether, our integrative systems ecology approach where we combined physiology with genomics and downstream analyses resulted in a more fundamental insight in degradation of 2,4-D and 2,4,5-T, the species involved in that and the types of pathway that they use. Such an approach may easily be applied for other types of bioremediation as well.

## 5. Acknowledgements:

This work is supported by a BE-Basic Foundation-FES grant from the Dutch Ministry of Economic Affairs and by a grant of the Ministry of Science & Technology Vietnam (MOST, project code 826/QD-BKHCN). We thank Brett Olivier and Chrats Melkonian for assistance in data management.

## 6. References

- Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., . . . Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, 37 (4), 420-423. doi:10.1038/s41587-019-0036-z
- Baldwin, I. T., Staszak-Kozinski, L., & Davidson, R. (1994). Up in smoke: I. Smoke-derived germination cues for postfire annual, *Nicotiana attenuata* torr. Ex. Watson. *Journal of Chemical Ecology*, 20 (9), 2345-2371.

- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., . . . Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology : A Journal of Computational Molecular Cell Biology*, 19 (5), 455-477. doi:10.1089/cmb.2012.0021
- Cycoń, M., Żmijowska, A., & Piotrowska-Seget, Z. (2011). Biodegradation kinetics of 2,4-D by bacterial strains isolated from soil. *Central European Journal of Biology*, 6 (2), 188-198. doi:10.2478/s11535-011-0005-0
- Danganan, C. E., Ye, R. W., Daubaras, D. L., Xun, L., & Chakrabarty, A. M. (1994). Nucleotide sequence and functional analysis of the genes encoding 2,4,5-trichlorophenoxyacetic acid oxygenase in *Pseudomonas cepacia* AC1100. *Applied and Environmental Microbiology*, 60 (11), 4100-4106.
- [dataset]Thi Lan Anh Nguyen, Tjalf de Boer and Rob J.M. van Spanning; 2020; Genome sequences; GenBank at NCBI; Persistent identifier to be added after acceptance.
- Daubaras, D. L., Danganan, C. E., Hübner, A., Ye, R. W., Hendrickson, W., & Chakrabarty, A. M. (1996). Biodegradation of 2,4,5-trichlorophenoxyacetic acid by *Burkholderia cepacia* strain AC1100: evolutionary insight. *Gene*, 179 (1), 1-8.
- Daubaras, D. L., Hershberger, C. D., Kitano, K., & Chakrabarty, A. M. (1995). Sequence analysis of a gene cluster involved in metabolism of 2,4,5-trichlorophenoxyacetic acid by *Burkholderia cepacia* AC1100. *Applied and Environmental Microbiology*, 61 (4), 1279-1289.
- Daubaras, D. L., Saido, K., & Chakrabarty, A. M. (1996). Purification of hydroxyquinol 1,2-dioxygenase and maleylacetate reductase: the lower pathway of 2,4,5-trichlorophenoxyacetic acid metabolism by *Burkholderia cepacia* AC1100. *Applied and Environmental Microbiology*, 62 (11), 4276-4279.
- Don, R. H., & Pemberton, J. M. (1981). Properties of six pesticide degradation plasmids isolated from *Alcaligenes paradoxus* and *Alcaligenes eutrophus* . *Journal of Bacteriology*, 145 (2), 681-686.
- Ferraroni, M., Seifert, J., Travkin, V. M., Thiel, M., Kaschabek, S., Scozzafava, A., . . . Briganti, F. (2005). Crystal structure of the hydroxyquinol 1,2-dioxygenase from *Nocardioides simplex* 3E, a key enzyme involved in polychlorinated aromatics biodegradation. *Journal of Biological Chemistry*, 280 (22), 21144-21154. doi:10.1074/jbc.M500666200
- Filer, K., & Harker, A. R. (1997). Identification of the inducing agent of the 2,4-dichlorophenoxyacetic acid pathway encoded by plasmid pJP4. *Applied and Environmental Microbiology*, 63 (1), 317-320.
- Fukumori, F., & Hausinger, R. P. (1993a). *Alcaligenes eutrophus* JMP134 "2,4-dichlorophenoxyacetate monooxygenase" is an alpha-ketoglutarate-dependent dioxygenase. *Journal of Bacteriology*, 175 (7), 2083-2086. doi:10.1128/jb.175.7.2083-2086.1993
- Fukumori, F., & Hausinger, R. P. (1993b). Purification and characterization of 2,4-dichlorophenoxyacetate/alpha-ketoglutarate dioxygenase. *Journal of Biological Chemistry*, 268 (32), 24311-24317.
- Gisi, M. R., & Xun, L. (2003). Characterization of chlorophenol 4-monooxygenase (TftD) and NADH:flavin adenine dinucleotide oxidoreductase (TftC) of *Burkholderia cepacia* AC1100. *Journal of Bacteriology*, 185 (9), 2786-2792. doi:10.1128/jb.185.9.2786-2792.2003
- Golovleva, L. A., Pertsova, R. N., Evtushenko, L. I., & Baskunov, B. P. (1990). Degradation of 2,4,5-Trichlorophenoxyacetic acid by a *Nocardioides simplex* culture. *Biodegradation*, 1 (4), 263-271. doi:10.1007/BF00119763
- Gross, R., Guzman, C. A., Sebahia, M., dos Santos, V. A., Pieper, D. H., Koebnik, R., . . . Martinez-Arias, R. (2008). The missing link: *Bordetella petrii* is endowed with both the metabolic versatility of environmental bacteria and virulence traits of pathogenic Bordetellae. *BMC Genomics*, 9 , 449. doi:10.1186/1471-2164-9-449

- Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QAST: quality assessment tool for genome assemblies. *Bioinformatics*, *29* (8), 1072-1075. doi:10.1093/bioinformatics/btt086
- Hayashi, S., Sano, T., Suyama, K., & Itoh, K. (2016). 2,4-Dichlorophenoxyacetic acid (2,4-D)- and 2,4,5-trichlorophenoxyacetic acid (2,4,5-T)-degrading gene cluster in the soybean root-nodulating bacterium *Bradyrhizobium elkanii* USDA94. *Microbiological Research*, *188-189*, 62-71. doi:10.1016/j.micres.2016.04.014
- Hoffmann, D., Kleinstaub, S., Müller, R. H., & Babel, W. (2003). A transposon encoding the complete 2,4-dichlorophenoxyacetic acid degradation pathway in the alkali-tolerant strain *Delftia acidovorans* P4a. *Microbiology*, *149* (9), 2545-2556. doi:10.1099/mic.0.26260-0
- Hübner, A., Danganan, C. E., Xun, L., Chakrabarty, A. M., & Hendrickson, W. (1998). Genes for 2,4,5-trichlorophenoxyacetic acid metabolism in *Burkholderia cepacia* AC1100: characterization of the tftC and tftD genes and locations of the tft operons on multiple replicons. *Applied and Environmental Microbiology*, *64* (6), 2086-2093.
- Huong, N. L., Itoh, K., & Suyama, K. (2007). Diversity of 2,4-dichlorophenoxyacetic acid (2,4-D) and 2,4,5-trichlorophenoxyacetic acid (2,4,5-T)-degrading bacteria in Vietnamese soils. *Microbes and Environments*, *22* (3), 243-256. doi:10.1264/jsme2.22.243
- Itoh, K., Tashiro, Y., Uobe, K., Kamagata, Y., Suyama, K., & Yamamoto, H. (2004). Root nodule *Bradyrhizobium* spp. harbor tfdA $\alpha$  and cadA, homologous with genes encoding 2,4-dichlorophenoxyacetic acid-degrading proteins. *Applied and Environmental Microbiology*, *70* (4), 2110-2118. doi:10.1128/aem.70.4.2110-2118.2004
- Jiménez, J. I., Pérez-Pantoja, D., Chavarría, M., Díaz, E., & de Lorenzo, V. (2014). A second chromosomal copy of the catA gene endows *Pseudomonas putida* mt-2 with an enzymatic safety valve for excess of catechol. *Environmental Microbiology*, *16* (6), 1767-1778. doi:10.1111/1462-2920.12361
- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., . . . Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, *30* (9), 1236-1240. doi:10.1093/bioinformatics/btu031
- Ka, J. O., Holben, W. E., & Tiedje, J. M. (1994). Genetic and phenotypic diversity of 2,4-dichlorophenoxyacetic acid (2,4-D)-degrading bacteria isolated from 2,4-D-treated field soils. *Applied and Environmental Microbiology*, *60* (4), 1106-1106.
- Kellogg, S. T., Chatterjee, D. K., & Chakrabarty, A. M. (1981). Plasmid-assisted molecular breeding: new technique for enhanced biodegradation of persistent toxic chemicals. *Science*, *214* (4525), 1133-1135. doi:10.1126/science.7302584
- Kijima, K., Mita, H., Kawakami, M., & Amada, K. (2018). Role of CadC and CadD in the 2,4-dichlorophenoxyacetic acid oxygenase system of *Sphingomonas agrestis* 58-1. *Journal of Bioscience and Bioengineering*, *125* (6), 649-653. doi:10.1016/j.jbiosc.2018.01.003
- Kitagawa, W., Takami, S., Miyauchi, K., Masai, E., Kamagata, Y., Tiedje, J. M., & Fukuda, M. (2002). Novel 2,4-dichlorophenoxyacetic acid degradation genes from oligotrophic *Bradyrhizobium* sp. strain HW13 isolated from a pristine environment. *Journal of Bacteriology*, *184* (2), 509-518. doi:10.1128/jb.184.2.509-518.2002
- Kleinstaub, S., Müller, R. H., & Babel, W. (2001). Expression of the 2,4-D degradative pathway of pJP4 in an alkaliphilic, moderately halophilic soda lake isolate, *Halomonas* sp. EF43. *Extremophiles*, *5* (6), 375-384. doi:10.1007/s007920100202
- Köhler, H. P. E. (1999). *Sphingomonas herbicidovorans* MH: a versatile phenoxyalkanoic acid herbicide degrader. *Journal of Industrial Microbiology and Biotechnology*, *23* (4-5), 336-340. doi:10.1038/sj/jim/2900751

- Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33 (7), 1870-1874. doi:10.1093/molbev/msw054
- Kumari, R., Subudhi, S., Suar, M., Dhingra, G., Raina, V., Dogra, C., . . . Lal, R. (2002). Cloning and characterization of lin genes responsible for the degradation of Hexachlorocyclohexane isomers by *Sphingomonas paucimobilis* strain B90. *Applied and Environmental Microbiology*, 68 (12), 6021-6028. doi:10.1128/aem.68.12.6021-6028.2002
- Laemmli, C. M., Leveau, J. H., Zehnder, A. J., & van der Meer, J. R. (2000). Characterization of a second tfd gene cluster for chlorophenol and chlorocatechol metabolism on plasmid pJP4 in *Ralstonia eutropha* JMP134(pJP4). *Journal of Bacteriology*, 182 (15), 4165-4172. doi:10.1128/jb.182.15.4165-4172.2000
- Liu, T., & Chapman, P. J. (1984). Purification and properties of a plasmid-encoded 2,4-dichlorophenol hydroxylase. *FEBS Letters*, 173 (2), 314-318. doi:10.1016/0014-5793(84)80797-8
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., & Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids research*, 42 (D1), D490-495. doi:10.1093/nar/gkt1178
- Mampel, J., Ruff, J., Junker, F., & Cook, A. M. (1999). The oxygenase component of the 2-aminobenzenesulfonate dioxygenase system from *Alcaligenes* sp. strain O-1 The GenBank accession number for the sequence reported in this paper is AF109074. *Microbiology*, 145 (11), 3255-3264. doi:https://doi.org/10.1099/00221287-145-11-3255
- Marron-Montiel, E., Ruiz-Ordaz, N., Rubio-Granados, C., Juarez-Ramirez, C., & Galindez-Mayer, C. J. (2006). 2,4-D-degrading bacterial consortium: Isolation, kinetic characterization in batch and continuous culture and application for bioaugmenting an activated sludge microbial community. *Process Biochemistry*, 41 (7), 1521-1528. doi:https://doi.org/10.1016/j.procbio.2006.02.012
- McFall, S. M., Chugani, S. A., & Chakrabarty, A. M. (1998). Transcriptional activation of the catechol and chlorocatechol operons: variations on a theme. *Gene*, 223 (1-2), 257-267. doi:10.1016/s0378-1119(98)00366-7
- Muller, T. A., Byrde, S. M., Werlen, C., van der Meer, J. R., & Kohler, H.-P. E. (2004). Genetic analysis of phenoxyalkanoic acid degradation in *Sphingomonas herbicidovorans* MH. *Applied and Environmental Microbiology*, 70 (10), 6066. doi:10.1128/AEM.70.10.6066-6075.2004
- Nielsen, T. K., Rasmussen, M., Demaneche, S., Cecillon, S., Vogel, T. M., & Hansen, L. H. (2017). Evolution of Sphingomonad gene clusters related to pesticide catabolism revealed by genome sequence and mobilomics of *Sphingobium herbicidovorans* MH. *Genome Biology and Evolution*, 9 (9), 2477-2490. doi:10.1093/gbe/evx185
- Nielsen, T. K., Xu, Z., Gozdereliler, E., Aamand, J., Hansen, L. H., & Sorensen, S. R. (2013). Novel insight into the genetic context of the cadAB genes from a 4-chloro-2-methylphenoxyacetic acid-degrading *Sphingomonas*. *PLOS One*, 8 (12), e83346-e83346. doi:10.1371/journal.pone.0083346
- Nojiri, H., Maeda, K., Sekiguchi, H., Urata, M., Shintani, M., Yoshida, T., . . . Omori, T. (2002). Organization and transcriptional characterization of catechol degradation genes involved in carbazole degradation by *Pseudomonas resinovorans* strain CA10. *Bioscience, Biotechnology, and Biochemistry*, 66 (4), 897-901. doi:10.1271/bbb.66.897
- Perkins, E. J., Gordon, M. P., Caceres, O., & Lurquin, P. F. (1990). Organization and sequence analysis of the 2,4-dichlorophenol hydroxylase and dichlorocatechol oxidative operons of plasmid pJP4. *Journal of Bacteriology*, 172 (5), 2351-2359. doi:10.1128/jb.172.5.2351-2359.1990
- Rejczak, T., & Tuzimski, T. (2015). A review of recent developments and trends in the QuEChERS sample preparation approach. *Open Chemistry*, 13 (1), 980-1010. doi:10.1515/chem-2015-0109

- Rice, J. F., Menn, F. M., Hay, A. G., Sanseverino, J., & Sayler, G. S. (2005). Natural selection for 2,4,5-trichlorophenoxyacetic acid mineralizing bacteria in agent orange contaminated soil. *Biodegradation*, 16 (6), 501-512. doi:10.1007/s10532-004-6186-8
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30 (14), 2068-2069. doi:10.1093/bioinformatics/btu153
- Serbent, M. P., Rebelo, A. M., Pinheiro, A., Giongo, A., & Tavares, L. B. B. (2019). Biological agents for 2,4-dichlorophenoxyacetic acid herbicide degradation. *Applied Microbiology and Biotechnology*, 103 (13), 5065-5078. doi:10.1007/s00253-019-09838-4
- Smejkal, C. W., Vallaes, T., Burton, S. K., & Lappin-Scott, H. M. (2001). A rapid method to screen degradation ability in chlorophenoxyalkanoic acid herbicide-degrading bacteria. *Letters in Applied Microbiology*, 32 (4), 273-277. doi:10.1046/j.1472-765x.2001.00900.x
- Stellman, J. M., Stellman, S. D., Christian, R., Weber, T., & Tomasallo, C. (2003). The extent and patterns of usage of Agent Orange and other herbicides in Vietnam. *Nature*, 422 (6933), 681-687. doi:10.1038/nature01537
- Streber, W. R., Timmis, K. N., & Zenk, M. H. (1987). Analysis, cloning, and high-level expression of 2,4-dichlorophenoxyacetate monooxygenase gene *tfdA* of *Alcaligenes eutrophus* JMP134. *Journal of Bacteriology*, 169 (7), 2950-2955. doi:10.1128/jb.169.7.2950-2955.1987
- Takenaka, S., Koshiya, J., Okugawa, S., Takata, A., Murakami, S., & Aoki, K. (2011). Fe-superoxide dismutase and 2-hydroxy-1,4-benzoquinone reductase preclude the auto-oxidation step in 4-aminophenol metabolism by *Burkholderia* sp. strain AK-5. *Biodegradation*, 22 (1), 1-11. doi:10.1007/s10532-010-9369-5
- Tao, Y., Fishman, A., Bentley, W. E., & Wood, T. K. (2004). 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. *Applied and Environmental Microbiology*, 70 (7), 3814-3820. doi:10.1128/aem.70.7.3814-3820.2004
- Tatusov, R. L., Galperin, M. Y., Natale, D. A., & Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*, 28 (1), 33-36. doi:10.1093/nar/28.1.33
- Thiel, M., Kaschabek, S. R., Groning, J., Mau, M., & Schlomann, M. (2005). Two unusual chlorocatechol catabolic gene clusters in *Sphingomonas* sp. TFD44. *Archives of Microbiology*, 183 (2), 80-94. doi:10.1007/s00203-004-0748-3
- Tonso, N. L., Matheson, V. G., & Holben, W. E. (1995). Polyphasic characterization of a suite of bacterial isolates capable of degrading 2,4-D. *Microbial Ecology*, 30 (1), 3-24. doi:10.1007/bf00184510
- Vallaes, T., Fulthorpe, R. R., Wright, A. M., & Soulas, G. (1996). The metabolic pathway of 2,4-dichlorophenoxyacetic acid degradation involves different families of *tfdA* and *tfdB* genes according to PCR-RFLP analysis. *FEMS Microbiology Ecology*. doi:10.1016/0168-6496(96)00027-X
- van der Zaan, B. M., Saia, F. T., Stams, A. J., Plugge, C. M., de Vos, W. M., Smidt, H., . . . Geritse, J. (2012). Anaerobic benzene degradation under denitrifying conditions: Peptococcaceae as dominant benzene degraders and evidence for a syntrophic process. *Environmental Microbiology*, 14 (5), 1171-1181. doi:10.1111/j.1462-2920.2012.02697.x
- Vedler, E., Vahter, M., & Heinaru, A. (2004). The completely sequenced plasmid pEST4011 contains a novel IncP1 backbone and a catabolic transposon harboring *tfd* genes for 2,4-dichlorophenoxyacetic acid degradation. *Journal of Bacteriology*, 186 (21), 7161-7174. doi:10.1128/jb.186.21.7161-7174.2004
- Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Brucoleri, R., . . . Medema, M. H. (2015). antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic*

*Acids Research*, 43 (W1), W237-243. doi:10.1093/nar/gkv437

Weisburg, W. G., Barns, S. M., Pelletier, D. A., & Lane, D. J. (1991). 16S ribosomal DNA amplification for phylogenetic study. *Journal of Bacteriology*, 173 (2), 697-703. doi:10.1128/jb.173.2.697-703.1991

Westing, A. H. (1984). Herbicides in War: Past and Present in "Herbicides in war: the long-term ecological and human consequence".

Yang, Z., Xu, X., Dai, M., Wang, L., Shi, X., & Guo, R. (2017). Rapid degradation of 2,4-dichlorophenoxyacetic acid facilitated by acetate under methanogenic condition. *Bioresource Technology*, 232, 146-151. doi:10.1016/j.biortech.2017.01.069

Zaborina, O., Daubaras, D. L., Zago, A., Xun, L., Saido, K., Klem, T., . . . Chakrabarty, A. M. (1998). Novel pathway for conversion of chlorohydroxyquinol to maleylacetate in *Burkholderia cepacia* AC1100. *Journal of Bacteriology*, 180 (17), 4667-4675.

Zharikova, N. V., Iasakov, T. R., Zhurenko, E. Y., Korobov, V. V., & Markusheva, T. V. (2018). Bacterial genes of 2,4-dichlorophenoxyacetic acid degradation encoding  $\alpha$ -ketoglutarate-dependent dioxygenase activity. *Biology Bulletin Reviews*, 8, 155-167.

## Data Accessibility

The NCBI-annotated genomes of BT1 2.2 (PRJNA586721), BT1 5.2 (PRJNA586745), BT1 9.2 (PRJNA586752), and BT1 10.2 (PRJNA587043) that support the findings of this study were deposited in GenBank with their Bioproject numbers listed in between brackets. Links to these data, initially for review only, are listed below. After acceptance of the manuscript, the data will become publicly available along with an assigned DOI number.

*Pseudomonas aeruginosa* strain:BT1 2.2: <https://dataview.ncbi.nlm.nih.gov/object/PRJNA586721?reviewer=q7ko9k52t4a6u>

*Sphingomonas histidinilytica* strain:BT1 5.2: <https://dataview.ncbi.nlm.nih.gov/object/PRJNA586745?reviewer=i6f5dcie0oc>

*Bordetella petrii* strain:BT 1 9.2: <https://dataview.ncbi.nlm.nih.gov/object/PRJNA586752?reviewer=anls0i9ag2l5ehv9f3q7cl>

*Achromobacter xylosoxidans* strain:BT1 10.2: <https://dataview.ncbi.nlm.nih.gov/object/PRJNA587043?reviewer=ca8ottdc5c>

## Author Contributions

Nguyen Thi Lan Anh designed and performed the research, analysed the data, prepared figures and tables, authored or reviewed drafts of the paper, approved the final draft.

Thi Cam Ha Dang delivered soil material, reviewed drafts of the paper, approved the final draft.

Jacco Koekkoek determined the concentration of the herbicides, approved the final draft.

Martin Braster assisted in the laboratory experiments, reviewed the draft, approved the final draft.

John R. Parsons gave suggestions, reviewed the draft, approved the final draft.

Tjalf de Boer analyzed and annotated the genome of the isolates, reviewed the draft, approved the final draft.

Abraham Brouwer reviewed drafts of the paper, approved the final draft.

Rob J.M. van Spanning designed the experiments, analysed the data, prepared figures and tables, authored or reviewed drafts of the paper, approved the final draft.

## Tables and Figures

### Tables

Table 1: Genome assembly data of the 4 bacterial isolates.

Species type	Closest homologous species	%	Contigs	Total length (Mbps)	N50	Nr of predicted genes
BT1 2.2	<i>Pseudomonas aeruginosa</i>	88	99	6.930	144405	6369
BT1 5.2	<i>Sphingomonas histidinilytica</i>	71	73	6.049	169073	5797
BT1 9.2	<i>Bordetella petrii</i>	60	68	6.098	141283	5577
BT1 10.2	<i>Achromobacter xylosoxidans</i>	75	60	6.098	208222	5732

Percentages indicate an overall percentage of identity between the isolate and closest homologous species. Total length, N50 and number of contigs are determined after removal of contigs smaller than 5 kbps. The N50 is defined as the minimum contig length needed to cover 50% of the genome. It means, half of the genome sequence is in contigs larger than or equal to the N50 contig size.

Table 2: List of bacterial isolates derived from the first and second enrichments (denoted by ‘1’ and ‘2’ as last digit in the isolate names, respectively) of T1 soil communities on medium with herbicides.

S/N	Isolates	Closest related species	Class
1	BT1 1.1	<i>Rhodanobacter denitrificans</i>	$\gamma$ -Proteobacteria
2	BT1 2.1	<i>Arthrobacter oxydans</i>	Actinobacteria
3	BT1 3.1	<i>Pseudomonas nitroreducens</i>	$\gamma$ -Proteobacteria
4	BT1 5.1	<i>Achromobacter</i> sp.	$\beta$ -Proteobacteria
5	BT1 6.1	<i>Sphingomonas</i> sp.	$\alpha$ -Proteobacteria
6	BT1 7.1	<i>Bosea</i> sp.	$\alpha$ -Proteobacteria
7	BT1 10.1	<i>Pedobacter tournemirensis</i>	Sphingobacteriia
8	BT1 1.2	<i>Sphingopyxis ginsengisoli</i>	$\alpha$ -Proteobacteria
9	BT1 2.2	<i>Pseudomonas aeruginosa</i>	$\gamma$ -Proteobacteria
10	BT1 5.2	<i>Sphingomonas histidinilytica</i>	$\alpha$ -Proteobacteria
11	BT1 8.2	<i>Staphylococcus</i> sp.	Bacilli
12	BT1 9.2	<i>Bordetella petrii</i>	$\beta$ -Proteobacteria
13	BT110.2	<i>Achromobacter xylosoxidans</i>	$\beta$ -Proteobacteria
14	BT1 11.2	<i>Ancylobacter dichloromethanicus</i>	$\alpha$ -Proteobacteria

Figures

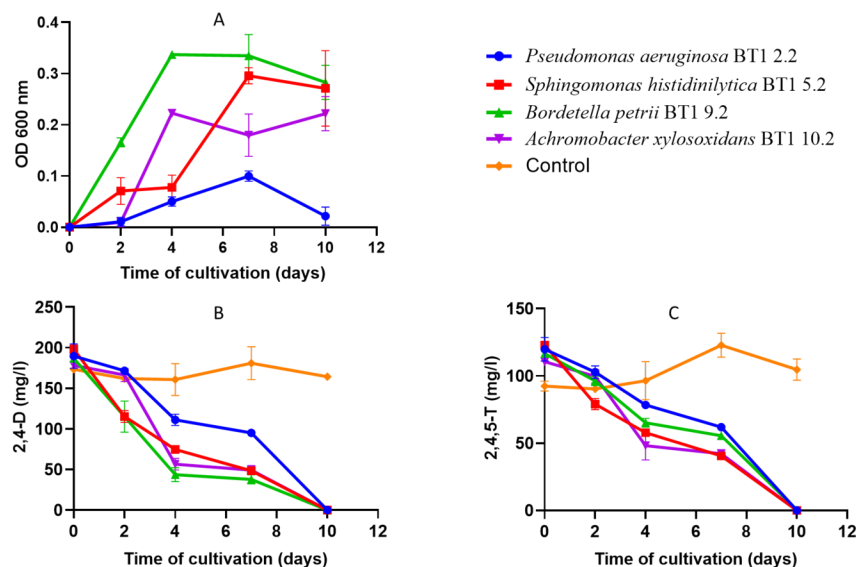


Figure 1. Bacterial growth (A), degradation of 2,4-D (B) and degradation of 2,4,5-T (C) in mineral salt medium with a mixture of 2,4-D and 2,4,5-T inoculated with either of the 4 isolated species listed. The optical density at the start of each growth curve was subtracted from those further on in these curves. Symbols are the means of three replicates. Error bars represent the standard deviation which was within 5% of the mean.

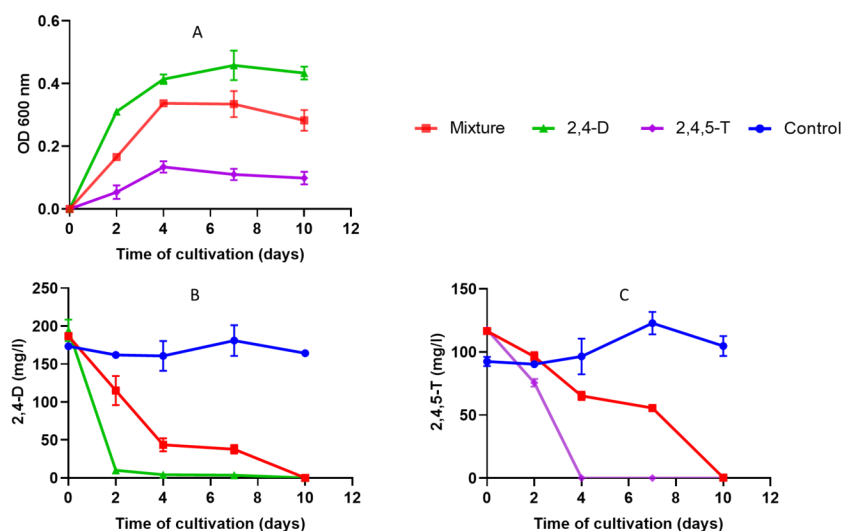


Figure 2. Bacterial growth (A), degradation of 2,4-D (B) and degradation of 2,4,5-T (C) in mineral salt medium with either a mixture of 2,4-D and 2,4,5-T or with only 2,4-D or 2,4,5-T as single herbicide inoculated with *B. petrii* BT1 9.2. Data points are the means of three replicates. The optical density at the start of each growth curve was subtracted from those further on in these curves. Error bars represent the standard deviation which was within 5% of the mean.



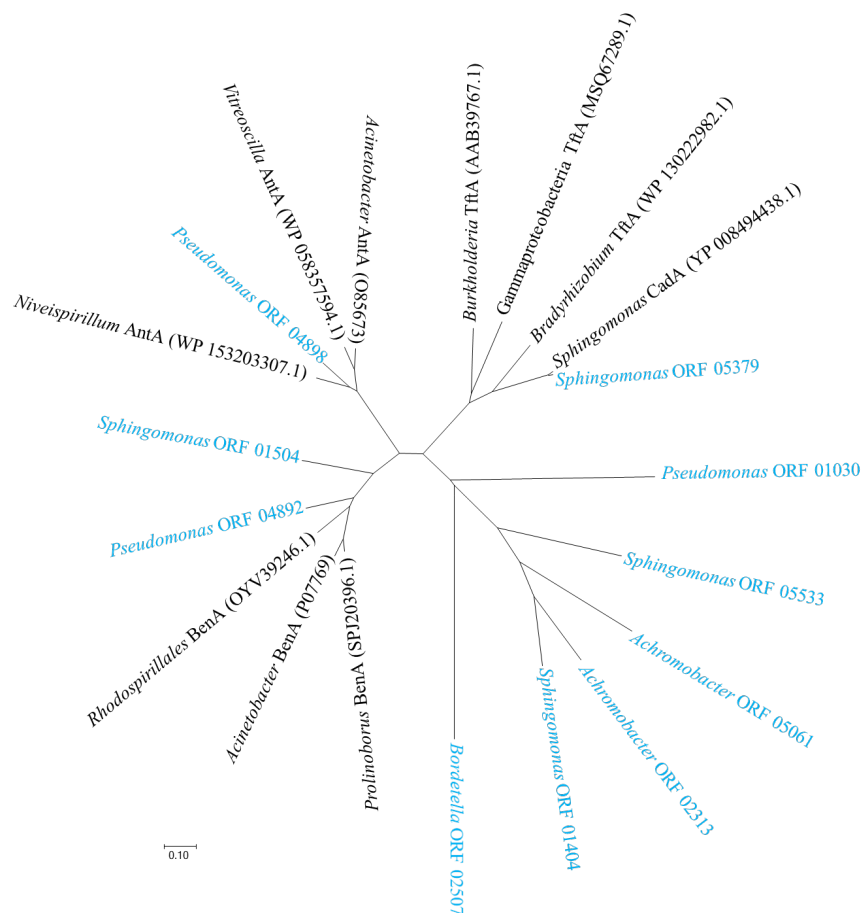


Figure 3. Dendrogram showing the relatedness of relevant TftA homologues sequences of the 4 isolated species (names in light blue) along with reference protein sequences from KEGG and NCBI databases (names in black). Homologues include CdaA, BenA and AntA sequences from the KEGG database along with their best matching counterparts in  $\alpha$ -,  $\beta$ - and  $\gamma$ -proteobacteria. The dendrogram was obtained by the neighbour-joining method using MEGA 7 based on sequence alignments calculated by CLUSTAL W using the default options. *Bradyrhizobium* TftA (WP\_130222982.1): *Bradyrhizobium* sp. Leo121, *Burkholderia* TftA (AAB39767.1): *Burkholderia cepacia*, *Gammaproteobacteria* TftA (MSQ67289.1): *Gammaproteobacteria* bacterium, *Rhodospirillales* BenA (OYV39246.1): *Rhodospirillales* bacterium 20-64-7, *Prolinoborus* BenA (SPJ20396.1): *Prolinoborus fasciculus*, *Acinetobacter* BenA (P07769): *Acinetobacter calcoaceticus*, *Niveispirillum* AntA (WP\_153203307.1): *Niveispirillum* sp. SYP-B3756, *Vitreoscilla* AntA (WP\_058357594.1): *Vitreoscilla massiliensis*, *Acinetobacter* AntA (O85673): *Acinetobacter baylyi* ADP1, *Sphingomonas* CdaA (YP\_008494438.1): *Sphingomonas* sp. ERG5 and isolates *Pseudomonas* ORF 1030, ORF 4892, ORF 4898: *P. aeruginosa* BT1 2.2, *Sphingomonas* ORF 5533, ORF 1504, ORF 5379, ORF 1404: *S. histidinilytica* BT1 5.2, *Bordetella* ORF 2507: *B. petrii* BT1 9.2, *Achromobacter* ORF 2313, ORF 5061: *A. xylosoxidans* BT1 10.2.

