environmental microbiology

Environmental Microbiology (2010) **12**(9), 2411–2425 doi:10.1111/j.1462-2920.2010.02215.x

Novel α-ketoglutarate dioxygenase *tfdA*-related genes are found in soil DNA after exposure to phenoxyalkanoic herbicides

M. C. Gazitúa,[1,2] A. W. Slater,[1] F. Melo[1] and B. González[1,2]*

[1]*Departamento de Genética Molecular y Microbiología, Facultad de Ciencias Biológicas, Millennium Nucleus on Plant Functional Genomics, Center for Advanced Studies in Ecology and Biodiversity, Pontificia Universidad Católica de Chile, Santiago, Chile.*

[2]*Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez, Santiago, Chile.*

Summary

Phenoxyalkanoic herbicides such as 2,4-dichlorophenoxyacetate (2,4-D), 2,4-dichlorophenoxybutyrate (2,4-DB) or mecoprop are widely used to control broad-leaf weeds. Several bacteria have been reported to degrade these herbicides using the α-ketoglutarate-dependent, 2,4-dichlorophenoxyacetate dioxygenase encoded by the *tfdA* gene, as the enzyme catalysing the first step in the catabolic pathway. The effects of exposure to different phenoxyalkanoic herbicides in the soil bacterial community and in the *tfdA* genes diversity were assessed using an agricultural soil exposed to these anthropogenic compounds. Total community bacterial DNA was analysed by terminal restriction fragment length polymorphism of the 16S rRNA and the *tfdA* gene markers, and detection and cloning of *tfdA* gene related sequences, using PCR primer pairs. After up to 4 months of herbicide exposure, significant changes in the bacterial community structure were detected in soil microcosms treated with mecoprop, 2,4-DB and a mixture of both plus 2,4-D. An impressive variety of novel *tfdA* gene related sequences were found in these soil microcosms, which cluster in new *tfdA* gene related sequence groups, unequally abundant depending on the specific herbicide used in soil treatment. Structural analysis of the putative protein products showed small but significant amino acid differences. These *tfdA* gene sequence variants are, probably, required for degradation of natural substrate(s) structurally related to these herbicides and their presence explains self-remediation of soils exposed to phenoxyalkanoic herbicides.

Received 18 November, 2009; accepted 18 February, 2010. *For correspondence. E-mail bernardo.gonzalez@uai.cl; Tel. (+56) 2331 1619; Fax (+56) 2331 1906.

Introduction

Phenoxyalkanoic (PA) herbicides are synthetic plant growth regulators that have been widely used to control broad leaf weeds since their introduction in the 40 s and 50 s (Ahrens, 1994). They usually are phenoxyacetic-, propionic- or butyric acids, with chloro and/or methyl residues in the aromatic ring (Fig. S1). Despite their foliar application, a significant amount of these herbicides reaches the ground and interacts with the soil microbiota where they can be degraded by bacteria (Fulthorpe *et al.*, 1996; Kamagata *et al.*, 1997; Kohler, 1999; Müller *et al.*, 1999) and fungi (Vroumsia *et al.*, 2005). Degradation of 2,4-dichlorophenoxyacetic acid (2,4-D) has been widely studied as a model for microbial decomposition of PA herbicides. The first step in the degradation of 2,4-D is performed by an α-ketoglutarate-dependent 2,4-D dioxygenase, encoded by the *tfdA* gene in β- and γ-*Proteobacteria* (Fukumori and Hausinger, 1993a; Suwa *et al.*, 1996; Vedler *et al.*, 2000), and by the *tfdA*α gene in α-*Proteobacteria* (Itoh *et al.*, 2002). It has been also described that some α-*Proteobacteria* use a 2,4-D monooxygenase, encoded by the *cadAB* genes, to degrade PA herbicides (Kitagawa *et al.*, 2002; Itoh *et al.*, 2004).

The archetypal TfdA enzyme, initially described in *Cupriavidus necator* JMP134 (previously named as *Ralstonia eutropha* JMP134), uses iron and catalyses the introduction of one oxygen atom into the ether bond of the PA substrate, and the other oxygen atom in the α-ketoacid co-substrate (Fukumori and Hausinger, 1993a; Fukumori and Hausinger, 1993b). This enzyme belongs to the group II of the α-ketoglutarate dependent dioxygenases (Eichhorn *et al.*, 1997), showing a clear homology to the *Escherichia coli* taurine/α-ketoglutarate dioxygenase (TauD). Based on this homology and site-directed mutagenesis studies, the conserved amino acid residues of TfdA protein that participate in binding to Fe^{2+}, α-ketoglutarate and 2,4-D have been defined (Hegg *et al.*,

1999; Hogan *et al*., 2000; Dunning Hotopp and Hausinger, 2002; Elkins *et al*., 2002).

It has been reported that TfdA enzymes are also involved in the first-step cleavage of other PA herbicides (Fig. S1), such as phenoxyacetic acid (PAA), 2-methyl-4-chlorophenoxyacetic acid (MCPA), 2,4,5-trichlorophenoxyacetic acid (2,4,5-T) and 2-(2,4-dichlorophenoxy) propionic acid (dichlorprop) (Fukumori and Hausinger, 1993a; Itoh *et al*., 2002). Bacterial strains that carry *tfdA*- or *tfdA*-related genes degrade other PA compounds such as 2-methylphenoxyacetic acid (2MPA), 2 (2-methyl-4-chlorophenoxy)-propionic acid (mecoprop), 2,4-dichlorophenoxybutyric acid (2,4-DB) and 2 (2-methyl-4-chlorophenoxy)-butyric acid (Pieper *et al*., 1988; Nickel *et al*., 1997; Smejkal *et al*., 2001). SdpA and RdpA are dioxygenases related to the TfdA enzyme, involved in the degradation of chiral PA compounds such as dichlorprop and mecoprop (Zipper *et al*., 1996; Müller *et al*., 1999; Müller *et al*., 2006). These enzymes use the (*S*)-, or the (*R*)-isomer as substrate, respectively, with much more affinity than TfdA protein (Nickel *et al*., 1997). These two enzymes show only 25% of amino acid identity between them, and less than 35% of amino acid identity with TfdA protein. However, SdpA and RdpA share the consensus sequence H*X*(D/E)$X_{138\text{-}207}$HX_{10}R/K, showing clear homology to the group II α-ketoglutarate dependent dioxygenases, as TfdA and TauD proteins (Schleinitz *et al*., 2004).

Complete *tfdA* gene sequences have been obtained from isolated bacterial strains able to degrade PA herbicides (see as examples, Streber *et al*. 1987; Suwa *et al*., 1996; Vedler *et al*., 2000; Itoh *et al*., 2002; Poh *et al*., 2002; Hoffmann *et al*., 2003), while numerous partial *tfdA* gene sequences have been found in total genomic DNAs by PCR amplification using degenerated primers (Shaw and Burns, 2004; Bælum *et al*., 2006; Gonod *et al*., 2006; Zakaria *et al*., 2007; Bælum *et al*., 2008). The analyses of the sequences have defined the existence of four groups of *tfdA* gene related sequences in *Proteobacteria*: *tfdA* gene classes I, II and III (Mcgowan *et al*., 1998), and *tfdA*α genes (Itoh *et al*., 2000). The reasons for this emerging picture of gene sequence variants are presently unknown.

In contrast with the number of studies dealing with PA degradation in bacterial isolates, very few studies have focused on the effects of PA compounds on microbial communities and, more specifically, on bacteria encoding these PA metabolizing dioxygenases (Gonod *et al*., 2006; Macur *et al*., 2007; Zakaria *et al*., 2007; Bælum *et al*., 2008). Using a culture-independent approach, it has been demonstrated that continuous exposure to *R*- or racemic mecoprop affects the soil bacterial community composition (Zakaria *et al*., 2007). Another study has shown that exposure to different amounts of 2,4-D also modified the bacterial community structure with the highest concentration treatment resulting in a significant decrease in bacterial diversity and an enrichment of bacterial populations related to the *Burkholderia* genus (Macur *et al*., 2007).

Using culture-independent, total genomic DNA and bioinformatic approaches, this work reports the isolation of an impressive number of novel *tfdA* gene related sequences from a single soil exposed to different PA herbicides. A new group of TfdA-related dioxygenases was defined, based on structural features of these novel sequences and the changes in the structure of the soil bacterial community and the *tfdA* gene related sequences pool upon exposure to different PA compounds. Thus, this study increases our knowledge on diversity of TfdA-related dioxygenases in microbial metabolism of PA herbicides and would help design of bioremediation strategies to clean environments polluted with these anthropogenic compounds.

Results

Total genomic DNA detection of tfdA gene related sequences in a soil exposed to different PA herbicides

To look for the presence of *tfdA* gene related sequences, a PCR detection procedure using a *tfdA* gene sequence primer pair (the so-called short PCR product) was used to scan total genomic DNA obtained from soil microcosms exposed for 4 months to different PA herbicide treatments. PCR products of the expected length (360 bp) were obtained for all treatments (data not shown), including the non-incubated soil and soil exposed to benzoic acid, or water, used as controls to track the specific effect of PA herbicides. *tfdA* gene sequence PCR products from the non-incubated soil and those from the treatment with 2,4-D and the mixture of three herbicides were cloned and subjected to RFLP analysis with the MspI endonuclease, in order to have a preliminary idea of the number of different sequences from each soil treatment. One hundred and eighty five clones were analysed (50, 75 and 60 clones from the non-incubated soil, and the soils treated with 2,4-D or the mixture of herbicides respectively) and 37 distinct MspI RFLP patterns were detected. Fifty-three clones, representing 37 unique patterns, were sequenced and their sequences were compared with the non-redundant protein sequence database at the NCBI, with the blastp software (Altschul *et al*., 1997), revealing the presence of 46 sequences with significant identity to reported α-ketoglutarate dioxygenases. These sequences have 33–58% amino acid sequence identity with the canonical TfdA enzyme from *C. necator* JMP134, and possess characteristic residues from the active site of this enzyme (data not shown). None of these sequences matched previously reported *tfdA* gene sequences. The

new TfdA protein related sequences were optimally aligned against TauD protein from *E. coli* (data not shown). The predicted three-dimensional mapping of these TfdA protein related sequences revealed that the short PCR product targets part of the active zone, but does not cover the part of the protein that may be responsive to changes in the structure of the PA substrate.

To obtain longer TfdA protein related sequences, a new set of degenerated forward primers was specifically designed to extend the amplified sequence towards the amino terminus and thus include residues involved in substrate binding. Longer PCR products for *tfdA* gene related sequences were again found in the non-incubated soil, and in soil microcosms treated with PAA, 2MPA, 2,4-D, mecoprop or 2,4-DB. PCR products of the expected length (~630 bp) were cloned and analysed by RFLP. Two hundred and twenty-two clones were analysed and 77 distinct HaeIII RFLP patterns were detected. The most frequent RFLP patterns and some found only in specific treatments (59 clones, in total) were sequenced and their sequences were aligned against the non-redundant nucleotide database using blastn program (Altschul *et al*., 1990). These sequences were translated and aligned against the TfdA/other dioxygenases MSA, as described in *Experimental procedures*. Phylogenetic trees were built based on this final MSA using maximum likelihood optimization criteria. The phylogram is shown in Fig. 1. Eleven of the new *tfdA* gene related sequences grouped with previously reported proteobacterial TfdA protein sequences and 23 clustered close to the RdpA, SdpA or TauD oxygenases. A second group of 23 sequences appeared as the sister group of the previously reported proteobacterial TfdA protein sequences, indicating that they are part of a new family of *tfdA* gene related sequences. All sequences within this group have a very short branch length because they are closely related (above 95% sequence identity). Finally, two highly divergent sequences clustered apart of TauD and TfdA related dioxygenases (clones 4–11 and 4–35).

Based on the known TfdA sequences, a new score scheme (the T score) was designed and used to assist the classification of the new sequences (see *Experimental procedures*). The T score ranges from 0 to 1, where a value of 1 would correspond to a hypothetical TfdA protein that has at each position the most frequently observed amino acid in all known TfdAs. It should be noted that this score is based only on the known TfdA proteins and the observed frequency of occurrence of their amino acids at each position. Therefore, this score is different than the sequence alignment score used to build the dendrogram and then it should be more robust for classifying new TfdA protein related sequences. Based on the T score, three groups of TfdA protein related sequences were clearly distinguished (Fig. 1), which were named as follows: (i) TfdA protein related dioxygenases: those proteins with *T* values higher than 0.7, which include all the known α-, and β-proteobacterial TfdA proteins; (ii) TauD protein related sequences: those proteins that have *T* values lower than 0.6, which are closer to RdpA, SdpA and TauD dioxygenases; and finally (iii) PA dioxygenases, a proposed name for those protein sequences with *T* values between 0.6 and 0.7, which although appears as the sister group of the TfdA protein related dioxygenases in the phylogram (Fig. 1), their branch has not a strong statistical support (i.e. bootstraping value smaller than 50), and it is also close to the TauD protein related dioxygenase branch.

Presence of functionally important residues in the new TfdA protein related sequences

To assess the presence of functionally important residues in these new *tfdA* protein related sequences found in soil, six functional categories were defined (Table 1), based on previously described information for TauD enzyme (O'Brien *et al*., 2003). These categories were: Fe^{2+} binding; α-ketoglutarate binding; substrate binding; substrate locking; substrate burial or accessibility; and dimer formation (see *Experimental procedures*). The relative positions of the residues in each category within the alignment were identified and selected, generating sub-alignments that were used for the generation of sequence logos (Fig. 2). Two residues involved in Fe^{2+} binding (H99, and D101 in *E. coli* TauD protein), and one involved in α-ketoglutarate binding (T126 in *E. coli* TauD protein) were found in all sequences analysed here: the previously reported TfdA protein related sequences, the TfdA protein related sequences reported here, the PA dioxygenases and the TauD protein related sequences. It should be mentioned that two important functional residues (H255, involved in iron binding and R266, participating in α-ketoglutarate binding) are not shown because they are out of the region amplified by the longer PCR primer pair. The sequence logos exhibit several conserved residues among the known TfdA protein related dioxygenases, most of which are not observed in the TauD protein related dioxygenases (Fig. 2). For example, for the substrate binding region there are three serine residues (at positions 3, 7 and 9 in the sequence logo), one asparagine (at position 6), one phenylalanine (at position 8) and one histidine (at position 11). This pattern of conservation is clearly not present in the TauD protein related dioxygenases, although part of the sequences within this group has some of these residues in common (i.e. the serine residue at fourth position in the sequence logo of the TauD protein related sequences group). Interestingly, the group of PA dioxygenases shows a somehow similar pattern to that observed in the known TfdA protein related sequences,

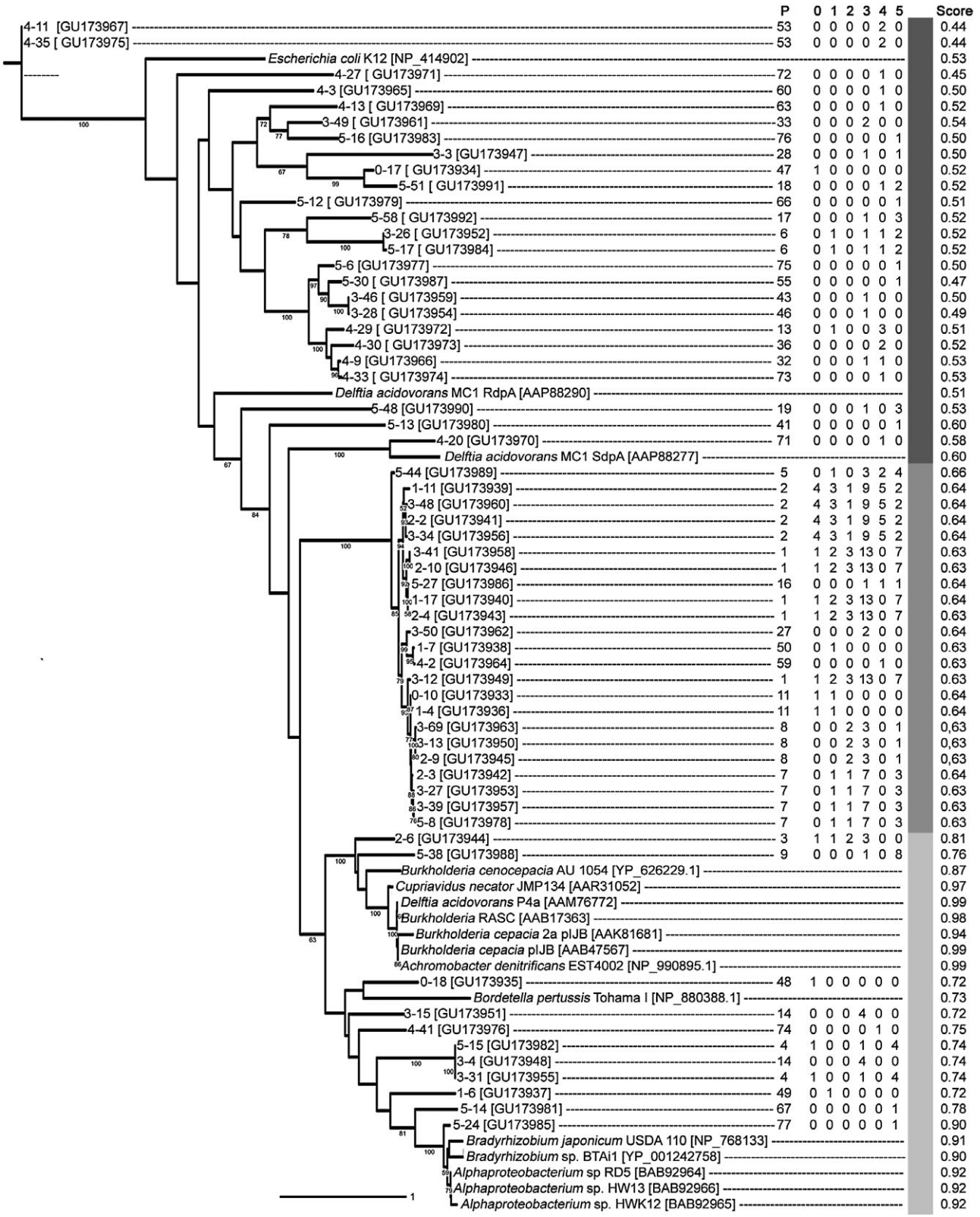
P 0 1 2 3 4 5 Score
4-11 [GU173967] 53 0 0 0 0 2 0 0.44
4-35 [GU173975] 53 0 0 0 0 2 0 0.44
Escherichia coli K12 [NP_414902] 0.53
4-27 [GU173971] 72 0 0 0 0 1 0 0.45
4-3 [GU173965] 60 0 0 0 0 1 0 0.50
4-13 [GU173969] 63 0 0 0 0 1 0 0.52
3-49 [GU173961] 33 0 0 0 2 0 0 0.54
5-16 [GU173983] 76 0 0 0 0 0 1 0.50
3-3 [GU173947] 28 0 0 0 1 0 1 0.50
0-17 [GU173934] 47 1 0 0 0 0 0 0.52
5-51 [GU173991] 18 0 0 0 0 1 2 0.52
5-12 [GU173979] 66 0 0 0 0 0 1 0.51
5-58 [GU173992] 17 0 0 0 1 0 3 0.52
3-26 [GU173952] 6 0 1 0 1 1 2 0.52
5-17 [GU173984] 6 0 1 0 1 1 2 0.52
5-6 [GU173977] 75 0 0 0 0 0 1 0.50
5-30 [GU173987] 55 0 0 0 0 0 1 0.47
3-46 [GU173959] 43 0 0 0 1 0 0 0.50
3-28 [GU173954] 46 0 0 0 1 0 0 0.49
4-29 [GU173972] 13 0 1 0 0 3 0 0.51
4-30 [GU173973] 36 0 0 0 0 2 0 0.52
4-9 [GU173966] 32 0 0 0 1 1 0 0.53
4-33 [GU173974] 73 0 0 0 0 1 0 0.53
Delftia acidovorans MC1 RdpA [AAP88290] 0.51
5-48 [GU173990] 19 0 0 0 1 0 3 0.53
5-13 [GU173980] 41 0 0 0 0 0 1 0.60
4-20 [GU173970] 71 0 0 0 0 1 0 0.58
Delftia acidovorans MC1 SdpA [AAP88277] 0.60
5-44 [GU173989] 5 0 1 0 3 2 4 0.66
1-11 [GU173939] 2 4 3 1 9 5 2 0.64
3-48 [GU173960] 2 4 3 1 9 5 2 0.64
2-2 [GU173941] 2 4 3 1 9 5 2 0.64
3-34 [GU173956] 2 4 3 1 9 5 2 0.64
3-41 [GU173958] 1 1 2 3 13 0 7 0.63
2-10 [GU173946] 1 1 2 3 13 0 7 0.63
5-27 [GU173986] 16 0 0 0 1 1 1 0.64
1-17 [GU173940] 1 1 2 3 13 0 7 0.64
2-4 [GU173943] 1 1 2 3 13 0 7 0.63
3-50 [GU173962] 27 0 0 0 2 0 0 0.64
1-7 [GU173938] 50 0 1 0 0 0 0 0.63
4-2 [GU173964] 59 0 0 0 0 1 0 0.63
3-12 [GU173949] 1 1 2 3 13 0 7 0.63
0-10 [GU173933] 11 1 1 0 0 0 0 0.64
1-4 [GU173936] 11 1 1 0 0 0 0 0.64
3-69 [GU173963] 8 0 0 2 3 0 1 0,63
3-13 [GU173950] 8 0 0 2 3 0 1 0,63
2-9 [GU173945] 8 0 0 2 3 0 1 0,63
2-3 [GU173942] 7 0 1 1 7 0 3 0.64
3-27 [GU173953] 7 0 1 1 7 0 3 0.63
3-39 [GU173957] 7 0 1 1 7 0 3 0.63
5-8 [GU173978] 7 0 1 1 7 0 3 0.63
2-6 [GU173944] 3 1 1 2 3 0 0 0.81
5-38 [GU173988] 9 0 0 0 1 0 8 0.76
Burkholderia cenocepacia AU 1054 [YP_626229.1] 0.87
Cupriavidus necator JMP134 [AAR31052] 0.97
Delftia acidovorans P4a [AAM76772] 0.99
Burkholderia RASC [AAB17363] 0.98
Burkholderia cepacia 2a pIJB [AAK81681] 0.94
Burkholderia cepacia pIJB [AAB47567] 0.99
Achromobacter denitrificans EST4002 [NP_990895.1] 0.99
0-18 [GU173935] 48 1 0 0 0 0 0 0.72
Bordetella pertussis Tohama I [NP_880388.1] 0.73
3-15 [GU173951] 14 0 0 0 4 0 0 0.72
4-41 [GU173976] 74 0 0 0 0 1 0 0.75
5-15 [GU173982] 4 1 0 0 1 0 4 0.74
3-4 [GU173948] 14 0 0 0 4 0 0 0.74
3-31 [GU173955] 4 1 0 0 1 0 4 0.74
1-6 [GU173937] 49 0 1 0 0 0 0 0.72
5-14 [GU173981] 67 0 0 0 0 0 1 0.78
5-24 [GU173985] 77 0 0 0 0 0 1 0.90
Bradyrhizobium japonicum USDA 110 [NP_768133] 0.91
Bradyrhizobium sp. BTAi1 [YP_001242758] 0.90
Alphaproteobacterium sp RD5 [BAB92964] 0.92
Alphaproteobacterium sp. HW13 [BAB92966] 0.92
Alphaproteobacterium sp. HWK12 [BAB92965] 0.92
1

Fig. 1. Phylogram of *TfdA* protein sequences. The phylogenetic tree was built using maximum likelihood optimization criteria with the JTT amino acid substitution model. Only bootstrapping values higher than 50 are shown. Known TfdA proteins from α-proteobacteria, β-proteobacteria and other a-ketoglutarate dioxygenases are indicated. The new TfdA protein sequences are described by two numbers separated by a dash, with the first one corresponding to the treatment, as follows: non-incubated soil (0); PAA (1); 2MPA (2); 2,4-D (3), mecoprop (4) and 2,4-DB (5). The vertical rectangle indicates the classification of the sequences according to their T score: TfdA protein related sequences (light grey), TauD protein related sequences (dark grey) and PA dioxygenases (grey). T score values are shown at the right of the vertical rectangle while the RFLP pattern (P) and the clonal abundance found in each treatment is shown at its left side.

but with some differences. For example, the serine at third position described above is now replaced by a threonine in the PA dioxygenase group, which is an amino acid with similar physicochemical properties. Similarly, the observed NSFS pattern at positions sixth to ninth in the logo is now replaced by SKFS pattern in the PA dioxygenase group. The TauD protein related group of sequences exhibits a more diverse repertoire of amino acids, but with the phenylalanine residue typically replaced by residues with aromatic groups (tyrosine or tryptophan), which suggests a critical role of an aromatic functional group at this position in all these dioxygenases. The location of the substrate binding residues is illustrated in the three-dimensional structure model of TfdA protein from *C. necator*, along with their position in the TfdA protein primary sequence (Fig. 3). These residues belong to five independent regions in the protein, which are non-local at the sequence level, but converge in the three-dimensional space surrounding the substrate. The regions related to substrate accessibility show little conservation between the different groups. Only a glutamic acid at the first position of the sequence logo appears to be conserved in the four groups, because the serine and histidine around positions tenth and fifteenth in the sequence logo (positions 207 and 271 in the alignment), are involved in substrate binding (i.e. some residues considered in the region of substrate accessibility are included in the substrate binding region). As it was described above, the serine and histidine residues are conserved in the TfdA protein related sequences. Therefore, this glutamic acid at position 84 in the alignment may also be important for the function of these enzymes, as it was pointed out above for the aromatic group at position 144. The unique group that exhibits high conservation of the amino acids involved in substrate accessibility is the PA dioxygenase group (Fig. 2). This is because this group is highly homogeneous and by far the less diverse overall, as it can be seen from the phylogram (Fig. 1), since the 23 sequences belonging to this group share more than 95% sequence identity between them. Therefore, conservation of amino acids in this specific group of sequences should be carefully interpreted. The three-dimensional mapping of the residues belonging to the substrate accessibility region showed that these residues correspond to non-local regions at the sequence level that form three independent loops that surround the entrance of the substrate to the active site of the enzyme (Fig. 3).

The analysis of the conservation of residues involved in dimer formation (not shown) suggests that some leucine residues are critical, because they are strongly conserved in all three groups. The residues involved in dimer formation belong to two alpha helices that are located behind the active site in the structure (Fig. 3), and thus should not be interacting directly with the substrate, cofactor or products. This suggests that dimer formation may be important for the function or the regulation of the activity of these enzymes.

Distribution and abundance of TfdA protein related, PA dioxygenase and TauD protein related sequences in soils treated with different PA herbicides

The number of retrieved clones/sequences was low and could not be used to explain the *tfdA* genes diversity

Table 1. Functionally important residues in taurine dioxygenase (TauD) from *E. coli*.

Feature	TauD amino acid[a]
Fe^{2+} binding	H99, D101, H255
α-Ketoglutarate binding	T126, R266
Substrate binding[b]	H70, Y73, I83, D94, N95, N97, T100, V102, F104, S158, F159, F206, R270
Substrate locking	Y73, D94, F159, F206
Substrate burial or accessibility	68-HIHPVY-73
	155-FRKSF-159
	203-NEGF-206
Dimer formation 1	139-SVPFRQLLS-147
Dimer formation 2	214-SEKESEALLSFLFAHI-229

a. Residue numbering corresponds to that observed in the crystal structure of TauD protein (PDB code 1OS7) from *E. coli* (O'Brien *et al*., 2003).
b. These residues were selected because they have at least one atom closer than 5.0 Å to any other atom of the substrate taurine. Although residues H99 and D101 fulfil the distance restraints mentioned above, they were not considered in the substrate-binding category because they participate directly in the binding to iron.

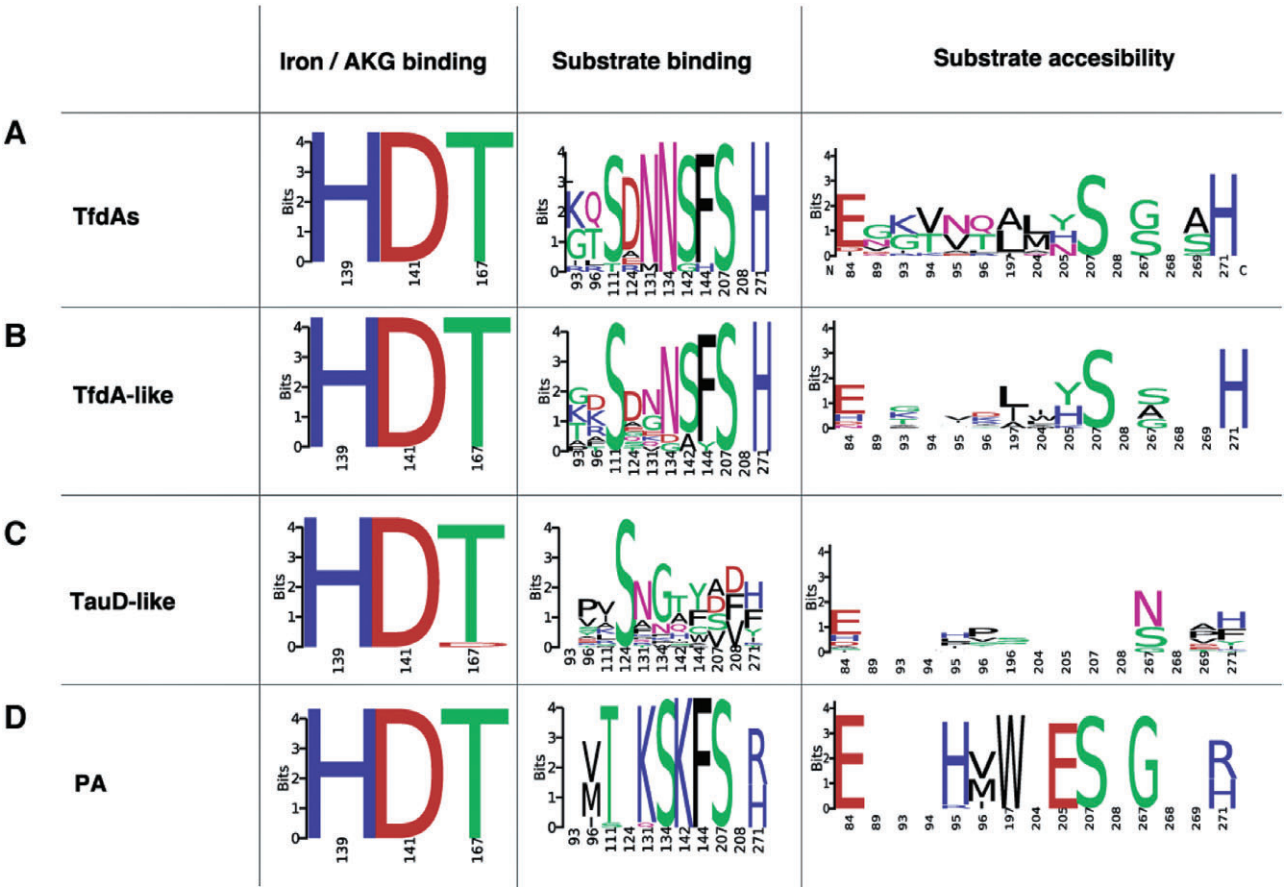

Fig. 2. Conservation of functionally important residues. Sequence logos for known (A) TfdA proteins, (B) TfdA protein related sequences, (C) TauD protein related sequences and (D) PA dioxygenases were calculated from some selected positions in the multiple sequence alignment that correspond to different functional categories, as deduced from what is known in TauD protein from *E. coli*. Numbers in the *x*-axis of the logos correspond to the sequential position number in the multiple sequence alignment. AKG, α-ketoglutarate.

found in this soil. However, the effect of specific PA treatments in the distribution and abundance of the three groups of new sequences retrieved from soil can still be analysed, with caution. Three treatments (0, 1 and 2; non-incubated soil, and exposed to PA and 2-MPA respectively) did not include chlorinated compounds as the other three (3, 4 and 5; exposed to 2,4-D, mecoprop and 2,4-DB respectively). The PA dioxygenase group and the TfdA protein related sequences group contained at least one clone from each treatment, whereas the TauD protein related sequences group was clearly biased to clones coming from treatments with each of the three chlorinated derivatives (Fig. 1). Since the low number of clones/sequences analysed, it should be kept in mind that the absence of any particular group of sequences does not necessarily means that this group is not present. The analyses carried out with short PCR products provided essentially the same trends (data not shown). The *tfdA* related sequences found in shorter fragments clustered in the three groups of dioxygenase sequences defined with longer amplicons (Fig. 1). In addition, no perfect matches or amino acid sequence identities higher than 99% were found after comparison of short with long *tfdA* sequences, further supporting the high level of diversity of these soil sequences.

RFLP analysis of 222 clones was carried out to determine the relative abundance of sequences belonging to the three groups of dioxygenases in the different soil treatments (Table 2). RFLP patterns belonging to the PA

Table 2. Relative abundance (%) of clones belonging to the three groups of TfdA-related dioxygenases.

Treatment	TfdA	PA	TauD
Non-incubated soil	3 (30.0%)	6 (60.0%)	1 (10.0%)
PAA	2 (15.4%)	9 (69.2%)	2 (15.4%)
2MPA	2 (22.2%)	7 (77.8%)	0 (0%)
2,4-D	10 (17.5%)	38 (66.7%)	9 (15.8%)
Mecoprop	1 (4.0%)	9 (36.0%)	15 (60.0%)
2,4-DB	14 (29.2%)	18 (37.5%)	16 (33.3%)

One hundred fifty two of the 222 clones were classified according to their restriction fragment length polymorphism pattern using the HaeIII endonuclease.

Fig. 3. Three-dimensional protein structure model of TfdA protein. Upper panel: The 3D model from *C. necator* TfdA protein (A) is shown along with the coordinates of taurine (green), α-ketoglutarate (yellow) and iron (red), modelled as block residues from the template structure of TauD protein from *E. coli* (PDB code 1OS7). The substrate taurine was modelled only to illustrate its location in the structure. Residues involved in Fe^{2+} binding are shown in blue and those involved in α-ketoglutarate binding are shown in pink. The horizontal bar at the bottom represents the primary protein sequence with the location of key residues for iron and α-ketoglutarate binding shown as vertical bars. The sequence region corresponding to the long-PCR product is shown in light brown. The region that was not amplified by PCR is shown in grey. Two key residues for Fe^{2+} and α-ketoglutarate binding that were not present in the region amplified by PCR are also shown as grey vertical bars. Bottom panel: Only the region of the TfdA protein model corresponding to the long-PCR product is shown. Key residues involved in Fe^{2+} and α-ketoglutarate binding are shown in blue colour, taurine is shown in green, α-ketoglutarate in yellow and iron as a red ball. Key residues involved in substrate binding (B), substrate accessibility (C) and dimer formation (D) are shown in red colour along with their location in the primary protein sequence.

dioxygenase group sequences were more abundant in the non-incubated soil and those soil microcosms exposed to the non-chlorinated herbicides. Although the same trend was also observed with the soil treated with 2,4-D, the relative abundance of PA dioxygenase group sequences in soil treated with 2,4-DB and mecoprop was lower. RFLP patterns belonging to TauD protein related sequences were especially abundant in soil exposed to mecoprop (Table 2), and in 2,4-DB. About 1/5-1/3 of the sequences in each treatment belonged to the TfdA protein related sequences group with the exception, again, of the soil exposed to mecoprop.

Changes in the soil bacterial community structure upon exposure to PA herbicides

To address the effect of the different soil treatments in the bacterial community structure, a terminal restriction fragment length polymorphism (T-RFLP) analysis of the total genomic DNA was carried out using PCR primer pairs targeting *tfdA* gene related sequences (the short sequences), to track functional changes, and 16S rRNA gene sequences, to follow changes in the taxonomic groups of the bacterial community structure. Gross changes in the conventional ecological indexes for richness (S), and diversity (H) were only detected in the 16S rRNA T-RFLP profiles from soils exposed to mecoprop, and the mixture containing mecoprop, 2,4-DB and 2,4-D (Table S2), but no important changes were found for the *tfdA* gene sequence T-RFLP profiles. T-RFLP profiles obtained with primer pairs tracking long *tfdA* gene sequences, carried out with selected soil microcosms showed essentially the same trends that of T-RFLP profiles with short sequences (not shown). This indicates that more subtle effects would occur upon PA exposure. The comparison among the T-RFLP profiles, based on Bray-Curtis similarity coefficients, was visualized through NMDS and cluster analyses. One-way analysis of similarity (ANOSIM) was carried out to examine the statistical significance of grouping. The NMDS visualization is shown in Fig. 4. After 4 months of incubation, the 16S rRNA T-RFLP profiles show that the structure of the bacterial communities from soil microcosms exposed to mecoprop, the herbicide mixture, and benzoate, substantially differs from all the other treatments (Fig. 4A). The same trend was observed after 2 months of incubation (data not shown). The corresponding cluster analysis (Fig. S2) demonstrated that, at the 50% similarity level, the 16S rRNA T-RFLP profiles formed four distinct clusters: the first group, with high similarity (> 70%), includes T-RFLP profiles from the non-incubated soil, the water control, and the treatments with PAA, 2,4-D or 2MPA. The second cluster corresponds to T-RFLP profiles from soils treated with 2,4-DB; the third to those treated with mecoprop and the mixture of 2,4-D + mecoprop + 2,4-DB; and the fourth to the soils treated with benzoate. The similarities within the last three clusters ranged between 65% and 85%.

The NMDS analyses of the *tfdA* gene sequences T-RFLP profiles showed significant differences for the bacterial community composition in soils treated with mecoprop, 2,4-DB, and the PA mixture (Fig. 4B). The cluster analysis of the *tfdA* gene sequences T-RFLP profiles (Fig. S2) showed that the profiles from soils treated with the mixture cluster apart from those treated with mecoprop, and that both are different from all the other soil treatment profiles. Within this last group, the T-RFLP profiles from soils treated with 2,4-DB were separated from the rest, at the 62% similarity level.

Discussion

This work reports the detection in soil total genomic DNA of sequences related to the *tfdA* gene that encodes the α-ketoglutarate dependent dioxygenase involved in the first step of the degradation of PA herbicides. An impressive variety of new *tfdA* gene related sequence variants was found. The sequence analyses of these new gene sequence variants showed small but distinguishable amino acid changes, allowing the definition of a new group of these dioxygenase sequences. The continuous exposure of soil microcosms to different PA herbicides produced significant changes in the taxonomic (16S rRNA gene marker) and functional (*tfdA* gene marker) structure of the soil bacterial community.

The finding of new *tfdA* gene related sequences in environmental DNA is not absolutely unexpected since such sequences have been reported in bacteria isolated from soils (Fulthorpe *et al*., 1996; Vallaeys *et al*., 1996; Kamagata *et al*., 1997; Itoh *et al*., 2002, 2004), aquatic ecosystems (de Lipthay *et al*., 2002), as well as total genomic DNA samples (Ka *et al*., 1994). The striking aspect of the findings reported here is that these novel *tfdA* genes related sequences are detected in only one soil. This soil has been subjected to standard farming practices with annual weed control based on the use of *s*-triazine herbicides such as simazine and, quite occasionally, with MCPA herbicide to control simazine resistant weeds (Manzano *et al*., 2007). The presence of these numerous *tfdA* gene related sequences may explain the rapid turnover of 2,4-D in this soil (Manzano *et al*., 2007). Their detection in non-incubated soil and in soil microcosms exposed to benzoic acid or water indicates that these *tfdA* genes related sequences are abundant enough to be detected in the soil total genomic DNA without a need for enrichment of the bacteria harbouring gene sequences putatively involved in degradation of specific PA herbicides.

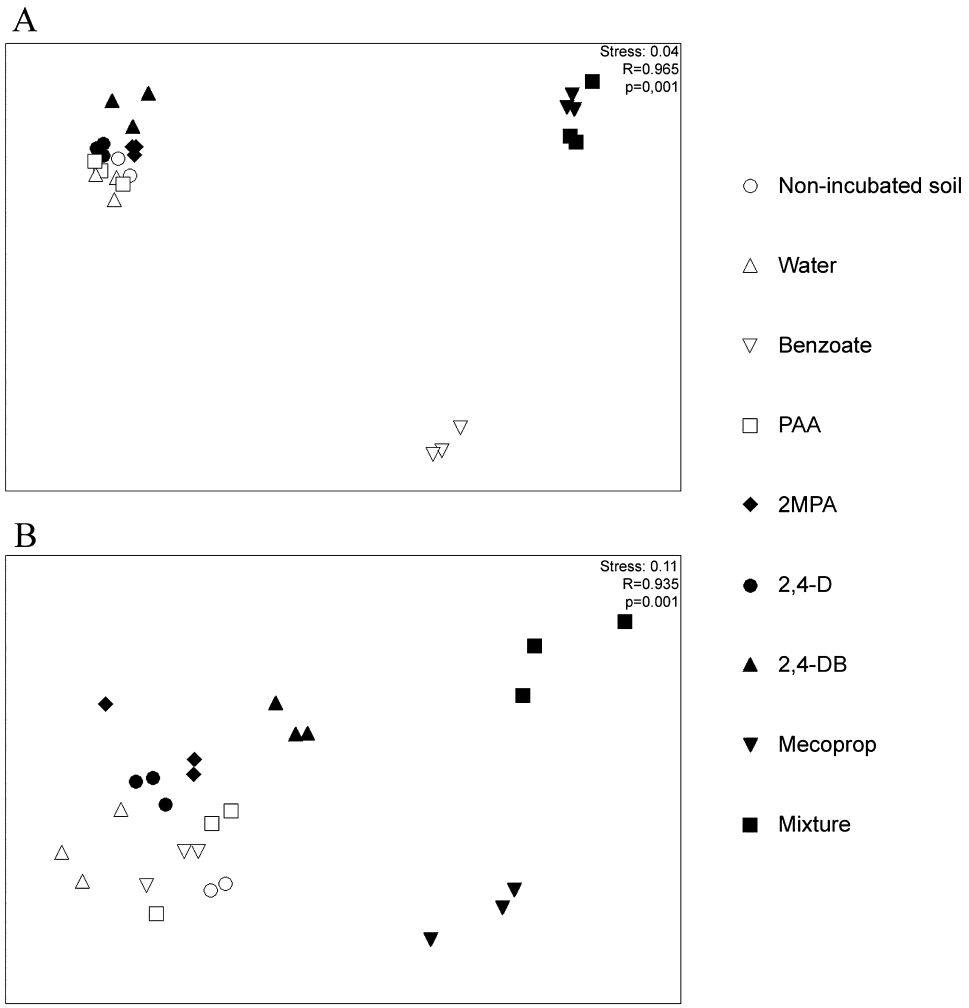

Fig. 4. Non-metric multidimensional scaling (NMDS) analysis of the 16S rRNA (A) and *tfdA* gene related sequences (B) T-RFLP profiles for total genomic DNA from soil microcosms continuously exposed for 4 months to PAA (□), 2MPA (◆), 2,4-D (●), 2,4-DB (▲), Mecoprop (▼), 2,4-D + Mecoprop + 2,4-DB (■), water (△) or benzoate (▽). The non-incubated soil is represented by (○). Each symbol corresponds to a single T-RFLP profile. Stress values for NMDS analysis, global *R* values and significance for ANOSIM among treatments are indicated. T-RFLP data obtained with the endonuclease MspI and HaeIII for 16S rRNA and *tfdA* gene related sequences were used for this analysis respectively. Use of the endonuclease HhaI gave essentially the same results, in both cases.

The abundance and diversity of *tfdA* gene related sequences in soils may be explained by the role of α-ketoglutarate dioxygenases in catabolism of natural, more common, compounds. The enzyme from *C. necator* JMP134 is able to use cinnamic acid derivatives as substrates (Dunning Hotopp and Hausinger, 2001), which are components of plant exudates. Indoleacetic acid derivatives (Marriott *et al*., 2000) and aryl-ether compounds released during degradation of lignin have been also proposed as natural substrates of TfdA enzyme (Hogan *et al*., 1997). The analysis of these new sequences showed several interesting features. The general trends were observed with both the shorter and the larger PCR products. The latter includes most of the functionally relevant sequences of the TfdA protein (Fig. 3). The more relevant result of this analysis is the definition of a new group of α-ketoglutarate dioxygenases (Fig. 1). In addition to the group of proteobacterial α-ketoglutarate dioxygenases and the heterogeneous group of sequences more closely related to RdpA, SdpA and TauD dioxygenases, a third group was found. This group, comprising only novel sequences, is characterized by intermediates *T* values that indicate similarity levels equally distant from the other two groups. We have named this novel group as PA dioxygenases since their sequences were found in soils exposed to both chlorinated and non-chlorinated PA herbicides. However, biochemical and genetics studies are clearly required to define if these putative dioxygenases are involved in metabolism of PA compounds.

The *tfdA* gene related sequences that appear only distantly related to the TfdA and TauD proteins in this soil, i.e. those from clones 4–11 and 4–35 (Fig. 1), may corre-

spond to α-ketoglutarate dioxygenase performing other oxygenative reactions. Xanthine dioxygenase, involved in hydroxylation of xanthine (Cultrone *et al*., 2005), the oxidative alkylsulfatase from *Pseudomonas putida* (Kahnert and Kertesz, 2000), or the sulfonate dioxygenase from *Saccharomyces cerevisiae* (Hogan *et al*., 1999), are examples of these dioxygenases. However, these proteins only share some structural/functional features of α-ketoglutarate dioxygenases and have very low identity at the amino acid level with the sequences reported here.

The sequences of the three groups of dioxygenases (Figs 2 and 3), possess the three key residues (H99, D101, T126 in *E. coli* TauD protein), two for iron binding and one for α-ketoglutarate binding, strongly suggesting that most of these sequences encode functional α-ketoglutarate dioxygenases that perform dioxygenolytic cleavage of PA derivatives. Based on the conservation of substrate binding residues, it seems that the PA dioxygenases group is more similar to TfdA protein related sequences than to TauD protein related dioxygenases, thus suggesting that these enzymes could use similar substrates.

The functionality of these *tfdA* gene related sequences is also suggested by other kind of evidence. The functional structure of the soil bacterial community, i.e. the number of *tfdA* gene related sequences (functional richness), and their relative abundance, is significantly affected by the specific herbicide compound used during enrichment (Fig. 4). The shifts in the taxonomic structure of the soil bacterial community, based on tracking of the 16S rRNA gene sequences, parallel those of the functional structure, indicating that the bacterial groups that possess the *tfdA* gene related sequences are also those that react to the presence of the herbicide. Indeed, more direct approaches tracking the expression of particular *tfdA* gene related sequences are required to evaluate the functionality of these sequences, as it has been demonstrated for *tfdA* gene from *C. necator* during degradation of MCPA in an agricultural soil (Nicolaisen *et al*., 2008). Changes in the relative abundance of variants of the *tfdA* gene sequences have been reported in soils exposed to MCPA herbicide using PCR primer pairs targeting the class III proteobacterial *tfdA* genes (Bælum *et al*., 2006). More recently, the use of primer pairs that distinguish among subgroups of proteobacterial TfdA protein sequences allowed establishment of some differences in the contribution of specific *tfdA* gene sequences to degradation of MCPA or 2,4-D in soils (Bælum *et al*., 2008; Bælum and Jacobsen, 2009).

Some soil treatments caused more pronounced effects than others, in both the taxonomic and functional structure of the soil bacterial community. The larger effects were observed when mecoprop, the only chiral herbicide tested, was present either alone or in the mixture. Mecoprop is metabolized by the distinct α-ketoglutarate dioxygenases, RdpA and SdpA, which are more closely related to TauD protein from *E. coli* than to the TfdA protein of 2,4-D degrading bacteria such as *C. necator* JMP134 (Fig. 1). Using SdpA, RdpA and TfdA probes in pure cultures, it has been suggested that the *rdpA* gene is involved in R-mecoprop degradation while *tfdA* and *sdpA* genes are involved in the degradation of both R- and S-mecoprop (Zakaria *et al*., 2007). Consistently, a significant increase of sequences of the TauD protein related group was found in soil microcosms exposed to mecoprop. In contrast, PA dioxygenase sequences were more abundant in the treatments with non-chlorinated compounds.

It cannot be ruled out that the exposure to specific PA herbicides may enrich the soil community DNA in catabolic gene sequences other than *tfdA* related genes. As it is mentioned above a gene cluster encoding a 2,4-monooxygenase able to cleave PA derivatives has been reported (Kitagawa *et al*., 2002; Itoh *et al*., 2004). The primer pairs used in this work are unable to track *cad* genes and other, not yet described gene sequences, putatively involved in PA herbicides degradation.

On the other hand, the significant change in the soil bacterial community in soil exposed to benzoate can be explained by the toxicity of this compound and their catabolic intermediates (Lambert *et al*., 1997; Cao and Loh, 2008). As expected, this effect is rather general than specifically related to *tfdA* gene bearing bacteria and, consistently, gross changes in bacterial community structure were also seen with the 16S rRNA marker (Fig. 4, Fig. S2).

The results reported here provide further support for the amazing catabolic abilities of microorganisms toward PA herbicides. In addition, give clues to explain the relatively fast turnover of these herbicides in soils and the natural bioremediation potential that can be observed in pristine and polluted environments (see as example, Fulthorpe *et al*., 1996; Gonod *et al*., 2006; Manzano *et al*., 2007).

Experimental procedures

Soil experiments

Soil microcosms were prepared with a 28-year-old avocado (*Persea americana*) plantation soil located in central Chile (32°50′S, 71°13′W), which had been periodically treated with simazine, an herbicide not related to PA compounds, to control annual weeds, and occasionally with MCPA to control simazine resistant weeds. Rhizospheric (but not the rizhoplane) soil (3 kg) was taken from the first 20 cm of depth, at a distance of 0.5 m from the trunk of three trees, then sieved and air-dried. The original soil (also referred as non-incubated soil in experimental conditions) was stored in a dark, 4°C room. Microcosms were prepared in 20 ml glass flasks containing 10 g of soil, incubated for up to 4 months at

22 ± 5°C. Each microcosm (three replicates for each condition) was amended every two weeks with 6 μmol (approximately 120 p.p.m.) of PA herbicides (PAA, 2-MPA, 2,4-D, mecoprop, 2,4-DB, or a mixture of the last three compounds), which is close to the levels used for weed control (usually 30–50 p.p.m.); benzoic acid to test for effects of a non-PA aromatic compound, or distilled water, as a control for incubation effects.

Soil DNA isolation

Soil total genomic DNA was isolated from 1 g soil samples at the beginning of the experiment and after 4 months using the 'Ultra Clean Soil DNA Kit' (MoBio Laboratories), following the instructions of the manufacturers, and modifications reported elsewhere (Manzano *et al*., 2007).

Terminal restriction fragment length polymorphism analysis

The T-RFLP technique was used to analyse the bacterial community structure and the diversity of the *tfdA* gene related sequences in each soil treatment. For the first case, oligonucleotide primer pairs designed for detection of bacterial 16S rRNA genes 8F (5′-AGA GTT TGA TCC TGG CTC AG-3′) (Lane, 1991) and 1392R (5′-ACG GGC GGT GTG TAC-3′) (Lane *et al*., 1985) were used. Primer 8F was labelled at the 5′ end with the NED fluorochrome. PCR reactions were carried out in a Perkin-Elmer 2400 thermal cycler using the reaction conditions described elsewhere (Morán *et al*., 2008). PCR product sizes (1.4 kb) were visualized after electrophoresis on 1% agarose gels prepared in TAE 1× buffer (40 mM Tris Acetate, 2 mM EDTA, pH 8) at 100 mV for 45 min and subsequent staining in ethidium bromide for 10 min. For the *tfdA* gene sequences analysis, *tfdA* gene sequence fragments of approximately 360 bp (and also those of longer sizes, see below) were amplified using degenerated primer pairs derived from conserved regions of the *tfdA* genes (Vallaeys *et al*., 1996; Itoh *et al*., 2002): forward primer 5′-AC(C/G) GAG TTC TG(C/T) GA(C/T) ATG-3′ and reverse primer 5′-(A/G)AC GCA GCG (G/A)TT (G/A)TC CCA-3′, the latter labelled with 6-FAM fluorochrome at the 5′ end. Each PCR reaction was performed in a total volume of 20 μL, containing 2 μL of 10× PCR buffer (200 mM Tris-HCl pH 8, 500 mM KCl), 1.2 mM $MgCl_2$, 0.2 mM of each primer, 0.2 mM of each deoxynucleotide triphosphate, 4% of dimethylsulfoxide, 50 ng μl^{-1} bovine seroalbumin, ~10–50 ng of soil DNA template and 1 U of *Taq* polymerase (Invitrogen, Brazil). PCR conditions were 94°C for 5 min, 35 cycles of 94°C for 1 min, 56°C of annealing temperature for 1 min and 72°C for 1 min, and a final extension step of 72°C for 7 min. PCR products were visualized as described for the 16S rRNA gene. PCR products were digested with endonucleases MspI or HhaI for 16S rRNA, and HaeIII or HhaI for *tfdA* gene sequences. Reactions were performed with 20 U of enzyme in a total volume of 20 μl at 37°C for 3 h, and the digested bacterial PCR products were purified by ethanol precipitation (Morán *et al*., 2008). The DNA fragments were separated by capillary electrophoresis using a Perkin-Elmer ABI Prism 310 sequencer. The fragment sizes were determined using the internal standard ROX 500 as reference. T-RFLP data handling was carried out as described elsewhere (Morán *et al*., 2008). Raw data sets consisted of peaks reflecting the size in bp of T-RFs and the area of each peak, measured in fluorescence units. T-RFs representing less than 0.5% of the total area and those with length sizes < 50 and > 500 bp were not included in the analyses. Technical duplicates gave essentially the same patterns and these data were averaged. Patterns from biological triplicates were treated separately. Data were standardized by calculating the area of each peak as a percentage of the total area. Profiles were plotted as peak area (relative abundance) against fragment size. Literature recommendations to minimize the drift between true and observed T-RFs were followed (Kaplan and Kitts, 2003). Richness (*R*) was calculated as the total T-RFs in each sample. Diversity (H′) was calculated by the Shannon–Weaver index with the formula $H' = \Sigma p_i * \mathrm{Ln}\ (p_i)$ (Margalef, 1958), where *p* is the proportion of an individual T-RF area relative to the sum of all T-RFs areas for each sample. Evenness (*E*) values were calculated with the formula $E = H'/\mathrm{Ln}\ (R)$ (Margalef, 1958). The T-RFs areas were analysed with the multivariate statistical software Primer 5 (Primer-E Ltd; Plymouth, UK) as described before (Morán *et al*., 2008; Pavissich *et al*., 2010). The statistical analysis, including non-metric multidimensional scaling (NMDS) analyses to group data according to their similarity, and one-way ANOSIM to examine the statistical significance of grouping, was performed as described elsewhere (Pavissich *et al*., 2010). The output statistic for NMDS, *R*, which indicates the magnitude of difference among group samples, takes a value of 0 if there is no separation of the bacterial community structure due to the factor analysed, and 1 if total separation takes place. Values of $R \geq 0.5$ account for high differences, whereas between 0.25 and 0.5 differences are considered as moderate (Morán *et al*., 2008; Pavissich *et al*., 2010).

Detection and cloning of tfdA gene related sequences

Partial *tfdA* gene related sequences of approximately 360 bp were amplified using the primer pairs described above. In order to obtain longer *tfdA* gene related sequences (~630 bp), a new set of modified forward primers was specifically designed to extend the sequence towards the amino terminus, based on conserved regions of the complete *tfdA* gene sequences from β-*Proteobacteria* (Table S1). The new set of forward primers was: 5′-CAG CA(G/A) ATC GCC TT(T/C) GCG-3′. These oligonucleotides have the following predicted features that would make them suitable for PCR: melting temperature in the range 60–65°C (Panjkovich *et al*., 2005), energy of intra-molecular interactions lower than −4 kcal mol^{-1} (Markham and Zuker, 2005), total number of contiguous inter-molecular interactions smaller than 6, total number of contiguous inter-molecular interactions at the 3′ end smaller than 5, and total number of non-contiguous inter-molecular interactions smaller than 9 bp. The reverse primer and the PCR conditions were the same indicated above.

The amplified short *tfdA* PCR products from the non-incubated soil, and from soils treated with 2,4-D and the mixture of herbicides, and the long *tfdA* PCR products from the non-incubated soil and from the treatments with PAA, 2MPA, 2,4-D, mecoprop or 2,4-DB treatments, were directly

ligated into pGEM-T Easy Vector and transformed into *E. coli* DH5α competent cells. The short *tfdA* gene related sequences clone library was analysed by RFLP with the enzyme MspI, while HaeIII was used for RFLP analysis of the long *tfdA* gene related sequences. Unique clones were sequenced and deposited in GenBank under the accession number GU173880-GU173932 (short sequences) and GU173933-GU173992 (long sequences).

Phylogenetic analysis

Thirteen *tfdA* gene related sequences, mostly of them with biochemically and or genetically supported function (Table S1), the RdpA and SdpA dioxygenase gene sequences from *D. acidovorans*, and taurine dioxygenase (TauD) from *E. coli*, were selected from the GenBank database (Benson *et al*., 2008). A progressive block alignment procedure was carried out with ClustalW (Li, 2003). TfdA protein related sequences were divided in two groups based on taxonomic criteria: α-and β-proteobacterial sequences. Each group was independently aligned using the BLOSUM62 matrix, gap opening and extension penalties of 1.0 and 0.1 respectively. The resulting multiple sequence alignments (MSAs), were aligned as blocks based on the matrix and gap penalties indicated above, to produce a TfdA MSA. With such procedure, an alignment of two alignments is generated and the original MSAs are not internally modified. This has the advantage of preserving common and specific features that would be relevant within each group. Then, the SdpA, RdpA and TauD protein sequences were individually aligned, using ClustalW with default gap penalties, against the TfdA MSA to produce a TfdA/other dioxygenases MSA. Then, each one of the 59 clone sequences reported here was individually aligned against the TfdA/other oxygenases MSA, using default ClustalW parameters, producing the TfdA/other oxygenases/clones MSA. This MSA was manually inspected with Jalview alignment editor (Waterhouse *et al*., 2009), and the first 47 and last 51 positions in the alignment were deleted. This editing procedure was carried out to exclude those protein regions at the N- and C-terminus that were not present in the clones. This edited MSA, containing a total of 75 sequences, was used to perform the phylogenetic and sequence analysis reported here.

A phylogenetic tree was built using maximum likelihood optimization criteria, with the PhyML package (Guindon *et al*., 2005). The JTT amino acid substitution model was used as previously described (Jones *et al*., 1992). Bootstrap values were calculated using one hundred replicates. Phylogenetic trees images were produced with the iTOL web server tool (Letunic and Bork, 2007) available at: http://itol.embl.de/.

Classification of new TfdA protein related sequences

To evaluate the degree of relationship between each new *tfdA* gene related sequence and those from known TfdA proteins, we created a global scoring scheme (the T score) that weights the contribution of each amino acid in the new sequence depending on its observed abundance in the same alignment position in the known TfdA proteins. Therefore, the maximum score value would be achieved by a new hypothetical TfdA protein sequence having at each position in the alignment the most frequently observed amino acid in the known TfdA proteins. Briefly, from the previously reported TfdA protein related sequences included in the alignment, the frequency of occurrence of each amino acid at each relative position in the alignment (positions 1–261 in the final MSA) was calculated. Each absolute frequency value was then divided by the total number of sequences contributing at that position (i.e. excluding gaps), thus obtaining a relative frequency value for the amino acid X at position Y in the alignment. After applying this procedure, a matrix of 261 columns (i.e. alignment positions) and 21 rows (i.e. 20 standard amino acids and a gap) was produced. This matrix was finally used to calculate the score (*T*) of each new sequence, using the following equation:

$$T = \frac{\sum_{i=1}^{261} M_{i,j}}{\sum_{i=1}^{261} \max M_{i,k}} \quad \forall k \in [1..21]$$

where *M* is the 261×21 matrix containing the relative frequencies, *i* is the alignment position and *j* is the amino acid of the sequence being scored found at the *i* position in the alignment.

Comparative structure modelling of C. necator TfdA protein

A comparative model of the TfdA protein sequence from *C. necator* (YP_025400) was built with MODELLER software (Sali and Blundell, 1993), using as a template the reported X-ray structure of TauD protein (PDB code 1OS7) from *E. coli* (O'Brien *et al*., 2003). This template structure shares 29% sequence identity with the TfdA protein sequence from *C. necator*. A pairwise alignment of the sequences was generated using ClustalW software with default gap opening and extension penalty values and BLOSUM62 matrix. Taurine, α-ketoglutarate and iron heteroatom groups present in the template structure were also included and modelled as block residues. The three-dimensional modelling was carried out with MODELLER software.

Definition of relevant residues in TfdA protein

Six categories of functionally relevant amino acids were defined in the TfdA protein three-dimensional model, based on their correspondence to those amino acids already described as important in TauD protein from *E. coli* (O'Brien *et al*., 2003) (Table 1). These functional categories included: Fe^{2+} binding, α-ketoglutarate binding, substrate binding, substrate locking, substrate burial or accessibility, and dimer formation. The relative positions of all these residues in the final MSA were determined and the corresponding subalignments of each functional category extracted. Then, sequence logos for each category were calculated using the WebLogo tool (http://weblogo.berkeley.edu/) (Crooks *et al*., 2004).

Supplementary data

Several files, in a variety of formats, containing the high-resolution phylogram, the multiple sequence alignment, DNA

sequences, protein sequences and the *C. necator* TfdA protein three-dimensional structure model are available from the following website: http://protein.bio.puc.cl/sup-mat.html.

Acknowledgements

This work has been funded by the 'Millennium Nucleus in Microbial Ecology and Environmental Microbiology and Biotechnology' Grant P/04-007-F, the 'Plant Functional Genomics' Grant P06-009F and the FONDECYT Grant 1080158. This study is also part of the research program FONDAP 1501-0001 funded by CONICYT to the Center for Advanced Studies in Ecology and Biodiversity Program 7. Additional support from grant PBCT RED-12, and grant ICA4-CT-2002-10011 (ACCESS) contract of the European Union is acknowledged.

References

Ahrens, W.H. (1994) *Herbicide Handbook (WSSA)*, 7th edn. Champaign, IL, USA: Weed Science Society of America.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215:** 403–410.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389–3402.

Bælum, J., and Jacobsen, C.S. (2009) TaqMan probe-based Real-Time PCR assay for detection and discrimination of class I, II, and III *tfdA* genes in soils treated with phenoxy acid herbicides. *Appl Environ Microbiol* **75:** 2969–2972.

Bælum, J., Henriksen, T., Hansen, H.C.B., and Jacobsen, C.S. (2006) Degradation of 4-chloro-2-methylphenoxyacetic acid in top- and subsoil is quantitatively linked to the class III *tfdA* gene. *Appl Environ Microbiol* **72:** 1476–1486.

Bælum, J., Nicolaisen, M.H., Holben, W.E., Strobel, B.W., Sørensen, J., and Jacobsen, C.S. (2008) Direct analysis of *tfdA* gene expression by indigenous bacteria in phenoxy acid amended agricultural soil. *ISME J* **2:** 677–687.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Wheeler, D.L. (2008) GenBank. *Nucleic Acids Res* **36:** D25–D30.

Cao, B., and Loh, K.-C. (2008) Catabolic pathways and cellular responses of *Pseudomonas putida* P8 during growth on benzoate with a proteomics approach. *Biotechnol Bioeng* **101:** 1297–1312.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res* **14:** 1188–1190.

Cultrone, A., Scazzocchio, C., Rochet, M., Montero-Morán, G., Drevet, C., and Fernández-Martín, R. (2005) Convergent evolution of hydroxylation mechanisms in the fungal kingdom: molybdenum cofactor-independent hydroxylation of xanthine via alpha-ketoglutarate-dependent dioxygenases. *Mol Microbiol* **57:** 276–290.

Dunning Hotopp, J.C., and Hausinger, R.P. (2001) Alternative substrates of 2,4-dichlorophenoxyacetate/a-ketoglutarate dioxygenase. *J Mol Catal B Enzym* **15:** 155–162.

Dunning Hotopp, J.C., and Hausinger, R.P. (2002) Probing the 2,4-dichlorophenoxyacetate/alpha-ketoglutarate dioxygenase substrate-binding site by site-directed mutagenesis and mechanism-based inactivation. *Biochemistry* **41:** 9787–9794.

Eichhorn, E., van der Ploeg, J.R., Kertesz, M.A., and Leisinger, T. (1997) Characterization of alpha-ketoglutarate-dependent taurine dioxygenase from *Escherichia coli*. *J Biol Chem* **272:** 23031–23036.

Elkins, J.M., Ryle, M.J., Clifton, I.J., Dunning Hotopp, J.C., Lloyd, J.S., Burzlaff, N.I., *et al*. (2002) X-ray crystal structure of *Escherichia coli* taurine/alpha-ketoglutarate dioxygenase complexed to ferrous iron and substrates. *Biochemistry* **41:** 5185–5192.

Fukumori, F., and Hausinger, R.P. (1993a) Purification and characterization of 2,4-dichlorophenoxyacetate/α-ketoglutarate dioxygenases. *J Biol Chem* **268:** 24311–24317.

Fukumori, F., and Hausinger, R.P. (1993b) *Alcaligenes eutrophus* JMP134 '2,4-dichlorophenoxyacetate monooxygenase' is an α-ketoglutarate-dependent dioxygenase. *J Bacteriol* **175:** 2083–2086.

Fulthorpe, R.R., Rhodes, A.N., and Tiedje, J.M. (1996) Pristine soils mineralize 3-chlorobenzoate and 2,4-dichlorophenoxyacetate via different microbial populations. *Appl Environ Microbiol* **62:** 1159–1166.

Gonod, L.V., Martin-Laurent, F., and Chenu, C. (2006) 2,4-D impact on bacterial communities, and the activity and genetic potential of 2,4-D degrading communities in soil. *FEMS Microbiol Ecol* **58:** 529–537.

Guindon, S., Lethiec, F., Duroux, P., and Gascuel, O. (2005) PHYML Online - a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res* **33:** W557–W559.

Hegg, E.L., Whiting, A.K., Saari, R.E., McCracken, J., Hausinger, R.P., and Que, L. (1999) Herbicide-degrading alpha-keto acid-dependent enzyme TfdA: Metal coordination environment and mechanistic insights. *Biochemistry* **38:** 16714–16726.

Hoffmann, D., Kleinsteuber, S., Müller, R.H., and Babel, W. (2003) A transposon encoding the complete 2,4-dichlorophenoxyacetic acid degradation pathway in the alkalitolerant strain *Delftia acidovorans* P4a. *Microbiology* **149:** 2545–2556.

Hogan, D.A., Auchtung, T.A., and Hausinger, R.P. (1999) Cloning and characterization of a sulfonate/alpha-ketoglutarate dioxygenase from *Saccharomyces cerevisiae*. *J Bacteriol* **181:** 5876–5879.

Hogan, D.A., Buckley, D.H., Nakatsu, C.H., Schmidt, T.M., and Hausinger, R.P. (1997) Distribution of the *tfdA* gene in soil bacteria that do not degrade 2,4 dichlorophenoxyacetic acid, 2,4-D. *Microb Ecol* **34:** 90–96.

Hogan, D.A., Smith, S.R., Saari, E.A., McCracken, J., and Hausinger, R.P. (2000) Site-directed mutagenesis of 2,4-dichlorophenoxyacetic acid/alpha-ketoglutarate dioxygenase - Identification of residues involved in metallocenter formation and substrate binding. *J Biol Chem* **275:** 12400–12409.

Itoh, K., Kanda, R., Momoda, Y., Sumita, Y., Kamagata, Y., Suyama, K., *et al*. (2000) Presence of 2,4-D-catabolizing bacteria in a Japanese arable soil that belong to BANA

(*Bradyrhizobium-Agromonas-Nitrobacter-Afipia*) cluster in α-*Proteobacteria*. *Microbes Environ* **15:** 113–117.

Itoh, K., Kanda, R., Sumita, Y., Kim, H., Kamagata, Y., Suyama, K., *et al*. (2002) *tfdA*-like genes in 2,4-dichlorophenoxyacetic acid-degrading bacteria belonging to the *Bradyrhizobium-Agromonas-Nitrobacter-Afipia* cluster in α-*Proteobacteria*. *Appl Environ Microbiol* **68:** 3449–3454.

Itoh, K., Tashiro, Y., Uobe, K., Kamagata, Y., Suyama, K., and Yamamoto, H. (2004) Root nodule *Bradyrhizobium* spp. harbour *tfdA*α and *cadA*, homologous with genes encoding 2,4-dichlorophenoxyacetic acid-degrading proteins. *Appl Environ Microbiol* **70:** 2110–2118.

Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* **8:** 275–282.

Ka, J.O., Holben, W.E., and Tiedje, J.M. (1994) Use of gene probes to aid in recovery and identification of functionally dominant 2,4-dichlorophenoxyacetic acid-degrading populations in soil. *Appl Environ Microbiol* **60:** 1116–1120.

Kahnert, A., and Kertesz, M.A. (2000) Characterisation of a sulfur-regulated oxygenative alkylsulfatase from *Pseudomonas putida*. *J Biol Chem* **257:** 31661–31667.

Kamagata, Y., Fulthorpe, R., Tamura, K., Takami, H., Forney, L.J., and Tiedje, J.M. (1997) Pristine environments harbour a new group of oligotrophic 2,4-dichlorophenoxyacetic acid-degrading bacteria. *Appl Environ Microbiol* **63:** 2266–2272.

Kaplan, C.W., and Kitts, C.L. (2003) Variation between observed and true terminal restriction fragment length is dependent on true TRF length and purine content. *J Microbiol Methods* **54:** 121–125.

Kitagawa, W., Takami, S., Miyauchi, K., Masai, E., Kamagata, Y., Tiedje, J.M., *et al*. (2002) Novel 2,4-dichlorophenoxyacetic acid degradation genes from oligotrophic *Bradyrhizobium* sp. strain HW13 isolated from a pristine environment. *J Bacteriol* **184:** 509–518.

Kohler, H.P.E. (1999) *Sphingomonas herbicidovorans* MH: a versatile phenoxyalkanoic acid herbicide degrader. *J Ind Microbiol Biotechnol* **23:** 336–340.

Lambert, L.A., Abshire, K., Blankenhorn, D., and Slonczewski, J.L. (1997) Proteins induced in *Escherichia coli* by benzoic acid. *J Bacteriol* **179:** 7595–7599.

Lane, D.J. (1991) *16S/23S rRNA Sequencing. Nucleic Acid Techniques in Bacterial Systematics*. New York, USA: John Wiley and Sons.

Lane, D.J., Pace, B., Olsen, G.J., Stahl, D.A., Sogin, M.L., and Pace, N.R. (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc Natl Acad Sci USA* **82:** 6955–6959.

Letunic, I., and Bork, P. (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23:** 27–28.

Li, K.B. (2003) ClustalW-MPI: ClustalW analysis using distributed and parallel computing. *Bioinformatics* **19:** 1585–1586.

de Lipthay, J.R., Aamand, J., and Barkay, T. (2002) Expression of *tfdA* genes in aquatic microbial communities during acclimation to 2,4-dichlorophenoxyacetic acid. *FEMS Microbiol Ecol* **40:** 205–214.

Macur, R.E., Wheeler, J.T., Burr, M.D., and Inskeep, W.P. (2007) Impacts of 2,4-D application on soil microbial community structure and on populations associated with 2,4-D degradation. *Microbiol Res* **162:** 37–45.

Manzano, M., Morán, A.C., Tesser, B., and González, B. (2007) Role of eukaryotic microbiota in soil survival and catabolic performance of the 2,4-D herbicide degrading bacteria *Cupriavidus necator* JMP134. *Anton Leeuwen Int J Gen Mol Microbiol* **91:** 115–126.

Margalef, R. (1958) Information theory in ecology. *Gen Syst* **3:** 36–71.

Markham, N.R., and Zuker, M. (2005) DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res* **33** (Web Server issue)**:** W577–W581.

Marriott, M.W., Smejkal, C.W., and Lappin-Scott, H.M. (2000) Biodegradation of mixtures of chlorophenoxyalkanoic acid herbicides by *Alcaligenes denitrificans*. *J Ind Microbiol Biotechnol* **25:** 1–5.

McGowan, C., Fulthorpe, R., Wright, A., and Tiedje, J.M. (1998) Evidence for interspecies gene transfer in the evolution of 2,4-dichlorophenxyacetic acid degraders. *Appl Environ Microbiol* **64:** 4089–4092.

Morán, A.C., Hengst, M.B., De la Iglesia, R., Andrade, S., Correa, J.A., and González, B. (2008) Changes in bacterial community structure associated with coastal copper enrichment. *Environ Toxicol Chem* **27:** 2239–2245.

Müller, T.A., Fleischmann, T., van der Meer, J.R., and Kohler, H.P.E. (2006) Purification and characterization of two enantioselective α-ketoglutarate-dependent dioxygenases, RdpA and SdpA, from *Sphingomonas herbicidovorans* MH. *Appl Environ Microbiol* **72:** 4853–4861.

Müller, R.H., Jorks, S., Kleinsteuber, S., and Babel, W. (1999) *Comamonas acidovorans* strain MC1: a new isolate capable of degrading the chiral herbicides dichlorprop and mecoprop and the herbicides 2,4-D and MCPA. *Microbiol Res* **154:** 241–246.

Nickel, K., Suter, M.J.F., and Kohler, H.P.E. (1997) Involvement of two α-ketoglutarate-dependent dioxygenases in enantioselective degradation of (*R*)- and (*S*)-mecoprop by *Sphingomonas herbicidovorans* MH. *J Bacteriol* **179:** 6674–6679.

Nicolaisen, M.H., Baelum, J., Jacobsen, C.S., and Sorensen, J. (2008) Transcription dynamics of the functional *tfdA* gene during MCPA herbicide degradation by *Cupriavidus necator* AEO106 (pRO101) in agricultural soil. *Environ Microbiol* **10:** 571–579.

O'Brien, J.R., Schuller, D.J., Yang, V.S., Dillard, B.D., and Lanzilotta, W.N. (2003) Substrate-induced conformational changes in *Escherichia coli* taurine/alpha-ketoglutarate dioxygenase and insight into the oligomeric structure. *Biochemistry* **42:** 5547–5554.

Panjkovich, A., Norambuena, T., and Melo, F. (2005) dnaMATE: a consensus melting temperature prediction server for short DNA sequences. *Nucleic Acids Res* **33:** 570–572.

Pavissich, J.P., Silva, M., and González, B. (2010) Sulfate reduction, molecular diversity, and copper amendment effects in bacterial communities enriched from sediments exposed to copper mining residues. *Environ Toxicol Chem* **29:** 256–264.

Pieper, D.H., Reineke, W., Engesser, K.H., and Knackmuss, H.J. (1988) Metabolism of 2,4-dichlorophenoxyacetic acid, 4-chloro-2-methylphenoxyacetic acid and 2-methylphenoxyacetic acid by *Alcaligenes eutrophus* JMP134. *Arch Microbiol* **150:** 95–102.

Poh, R.P.-C., Smith, A.R.W., and Bruce, I.J. (2002) Complete characterization of Tn*5530* from *Burkholderia cepacia* strain 2a (pIJB1) and studies of 2,4-dichlorophenoxyacetate uptake by the organism. *Plasmid* **48:** 1–12.

Sali, A., and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234:** 779–815.

Schleinitz, K.M., Kleinsteuber, S., Vallaeys, T., and Babel, W. (2004) Localization and characterization of two novel genes encoding stereospecific dioxygenases catalyzing 2(2,4-dichlorophenoxy)propionate cleavage in *Delftia acidovorans* MC1. *Appl Environ Microbiol* **70:** 5357–5365.

Shaw, L., and Burns, R.G. (2004) Enhanced mineralization of [U-14C]2,4-dichlorophenoxyacetic acid in soil from the rhizosphere of *Trifolium pretense*. *Appl Environ Microbiol* **70:** 4766–4774.

Smejkal, C.W., Vallaeys, T., Burton, S.K., and Lappin-Scott, H.M. (2001) Substrate specificity of chlorophenoxyalkanoic acid-degrading bacteria is not dependent upon phylogenetically related *tfdA* gene types. *Biol Fertil Soils* **33:** 507–513.

Streber, W.R., Timmis, KN., and Zenk, M.H. (1987) Analysis, cloning, and high-level expression of 2,4-dichlorophenoxyacetate monooxygenase gene *tfdA* of *Alcaligenes eutrophus* JMP134. *J Bacteriol* **169:** 2950–2955.

Suwa, Y., Wright, A.D., Fukumori, F., Nummy, K.A., Hausinger, R.P., Holben, W.E., and Forney, L.J. (1996) Characterization of a chromosomally encoded 2,4-dichlorophenoxyacetate/ a-ketoglutarate dioxygenase from *Burkholderia* sp. strain RASC. *Appl Environ Microbiol* **62:** 2464–2469.

Vallaeys, T., Fulthorpe, R., Wright, A.M., and 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 Microbiol Ecol* **20:** 163–172.

Vedler, E., Kõiv, V., and Heinaru, A. (2000) Analysis of the 2,4-dichlorophenoxyacetic acid-degradative plasmid pEST4011 of *Achromobacter xylosoxidans* subsp. *denitrificans* strain EST4002. *Gene* **255:** 281–288.

Vroumsia, T., Steiman, R., Seigle-Murandi, F., and Benoit-Guyod, J.L. (2005) Fungal bioconversion of 2,4-dichlorophenoxyacetic acid (2,4-D) and 2,4-dichlorophenol (2,4-DCP). *Chemosphere* **60:** 1471–1480.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009) Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25:** 1189–1191.

Zakaria, D., Lappin-Scott, H., Burton, S., and Whitby, C. (2007) Bacterial diversity in soil enrichment cultures amended with 2-(2-methyl-4-chlorophenoxy)propionic acid (mecoprop). *Environ Microbiol* **9:** 2575–2587.

Zipper, C., Nickel, K., Angst W., and Kohler, H.P.E. (1996) Complete microbial degradation of both enantiomers of the chiral herbicide mecoprop [(RS)-2-(4-chloro-2-methylphenoxy)propionic acid] in an enantioselective manner by *Sphingomonas herbicidovorans* sp. *Appl Environ Microbiol* **62:** 4318–4322.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Chemical structure of phenoxyalkanoic herbicides. Compounds used in this study are indicated with asterisk.
Fig. S2. Cluster analysis of 16S rRNA (A) and *tfdA* gene sequences (B) T-RFLP profiles.
Table S1. TfdA protein sequences used in this study.
Table S2. Total richness (S), diversity (H) and evenness (E) values obtained from the T-RFLP profiles of the 16S rRNA and the *tfdA* gene related sequences.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.