Abundance of Novel and Diverse tfdA-Like Genes, Encoding Putative Phenoxyalkanoic Acid Herbicide-Degrading Dioxygenases, in Soil‡†

Adrienne Zaprassis,†‡ Ya-Jun Liu,1,2‡ Shuang-Jiang Liu,2 Harold L. Drake,1 and Marcus A. Horn1*

Department of Ecological Microbiology, University of Bayreuth, Dr.-Hans-Frisch-Straße 1-3, 95440 Bayreuth, Germany;1 and State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, People’s Republic of China2

Received 21 July 2009/Accepted 20 October 2009

Phenoxyalkanoic acid (PAA) herbicides are widely used in agriculture. Biotic degradation of such herbicides occurs in soils and is initiated by α-ketoglutarate- and Fe(II)-dependent dioxygenases encoded by tfdA-like genes (i.e., tfdA and tfdAα). Novel primers and quantitative kinetic PCR (qPCR) assays were developed to analyze the diversity and abundance of tfdA-like genes in soil. Five primer sets targeting tfdA-like genes were designed and evaluated. Primer sets 3 to 5 specifically amplified tfdA-like genes from soil, and a total of 437 sequences were retrieved. Coverages of gene libraries were 62 to 100%, up to 122 genotypes were detected, and up to 389 genotypes were predicted to occur in the gene libraries as indicated by the richness estimator Chao1. Phylogenetic analysis of in silico-translated tfdA-like genes indicated that soil tfdA-like genes were related to those of group 2 and 3 Bradyrhizobium spp., Sphingomonas spp., and uncultured soil bacteria. Soil-derived tfdA-like genes were assigned to 11 clusters, 4 of which were composed of novel sequences from this study, indicating that soil harbors novel and diverse tfdA-like genes. Correlation analysis of 16S rRNA and tfdA-like gene similarity indicated that any two bacteria with D > 20% of group 2 tfdA-like gene-derived protein sequences belong to different species. Thus, data indicate that the soil analyzed harbors at least 48 novel bacterial species containing group 2 tfdA-like genes. Novel qPCR assays were established to quantify such new tfdA-like genes. Copy numbers of tfdA-like genes were 1.0 × 106 to 65 × 108 per gram (dry weight) soil in four different soils, indicating that hitherto-unknown, diverse tfdA-like genes are abundant in soils.

Phenoxyalkanoic acid (PAA) herbicides such as MCPA (4-chloro-2-methyl-phenoxyacetic acid) and 2,4-D (2,4-dichlorophenoxyacetic acid) are widely used to control broad-leaf weeds in agricultural as well as nonagricultural areas (19, 77). Degradation occurs primarily under oxic conditions in soil, and microorganisms play a key role in the degradation of such herbicides (62, 64). Although relatively rapidly degraded biotically and phytodegradably, herbicides harbor tfdA-like genes that are 76% identical to class I tfdA genes. Class III tfdA genes are closely related to those of Cupriavidus necator JMP134 (formerly Ralstonia eutrophus). Class II tfdA genes consist of those of Burkholderia sp. strain RASC and a few strains that are 66% identical to class I tfdA genes. Class III tfdA genes are 77% identical to class I and 80% identical to class II tfdA genes and linked to MCPA degradation in soil (3). The second group consists of alphaproteobacteria, which are closely related to Bradyrhizobium spp. with tfdAα genes having 60% identity to tfdA of group 1 (18, 29, 34). The third group also harbors the tfdAα genes and consists of Sphingomonas spp., within the alphaproteobacteria (30).

Diverse PAA herbicide degraders of all three groups were identified in soil by cultivation-dependent studies (32, 34, 41, 78). Besides CadAB, TfdA and certain TfdAα proteins catalyze the conversion of PAA herbicides (29, 30, 35). All groups of tfdA-like genes are potentially linked to the degradation of PAA herbicides, although alternative primary functions of group 2 and 3 TfdAs have been proposed (30, 35). However, recent cultivation-independent studies focused on 16S rRNA genes or solely on group 1 tfdA sequences in soil (e.g., references 3–5, 13, and 41). Whether group 2 and 3 tfdA-like genes are also quantitatively linked to the degradation of PAA herbicides in soils is unknown. Thus, tools to target a broad range of tfdA-like genes are needed to resolve such an issue. Primers

‡ Contributed equally to this study.
†† Published ahead of print on 30 October 2009.

* Corresponding author. Mailing address: Department of Ecological Microbiology, Dr.-Hans-Frisch-Str. 1-3, University of Bayreuth, 95440 Bayreuth, Germany. Phone: (49) (0)921-555620. Fax: (49) (0)921-555790. E-mail: marcus.horn@uni-bayreuth.de.
† Supplemental material for this article may be found at http://aem.asm.org/.

Copyright © 2010, American Society for Microbiology. All Rights Reserved.
TABLE 1. Soil characteristics and abundances of tfdA-like genes

<table>
<thead>
<tr>
<th>Site</th>
<th>pH</th>
<th>Clay (%)</th>
<th>Silt (%)</th>
<th>Sand (%)</th>
<th>C/N ratio</th>
<th>Reference(s)</th>
<th>16S rRNA gene copy no. (× 10^6 g⁻¹)</th>
<th>tfdA-like gene copy no. (× 10^6 g⁻¹)</th>
<th>tfdA-like gene copy no. per 16S rRNA gene (× 10⁻³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agricultural field</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Scheyern)</td>
<td>5.8</td>
<td>22</td>
<td>36</td>
<td>42</td>
<td>10</td>
<td>48</td>
<td>2.5 ± 0.1</td>
<td>12.9 ± 2.0</td>
<td>64.9 ± 0.7</td>
</tr>
<tr>
<td>Spruce forests</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solling</td>
<td>3.3</td>
<td>28</td>
<td>58</td>
<td>14</td>
<td>7</td>
<td></td>
<td>1.4 ± 0.1</td>
<td>2.0 ± 0.7</td>
<td>41.5 ± 2.7</td>
</tr>
<tr>
<td>Steigerwald</td>
<td>3.8</td>
<td>18</td>
<td>24</td>
<td>59</td>
<td>19</td>
<td>8, 42</td>
<td>3.0 ± 0.3</td>
<td>4.7 ± 0.8</td>
<td>40.8 ± 13.6</td>
</tr>
<tr>
<td>Unterlaess</td>
<td>3.1</td>
<td>3</td>
<td>23</td>
<td>74</td>
<td>8</td>
<td>37</td>
<td>3.5 ± 0.9</td>
<td>3.3 ± 1.4</td>
<td>9.8 ± 3.4</td>
</tr>
</tbody>
</table>

* Dry weight of soil.  
* Gene copy numbers were determined with primer set 3 (Table 2).  
* Gene copy numbers were determined with primer set 4 (Table 2).  
* Gene copy numbers were determined with primer set 5 (Table 2).  

used to assess the diversity of tfdA-like sequences used in previous studies were based on the alignment of approximately 50% or less of available sequences to date (3, 20, 29, 32, 39, 47, 58, 73). Primers specifically targeting all major groups of tfdA-like genes to assess and quantify a broad diversity of potential PAA degraders in soil are unavailable. Thus, the objectives of this study were (i) to develop primers specific for all three groups of tfdA-like genes, (ii) to establish quantitative kinetic PCR (qPCR) assays based on such primers for different soil samples, and (iii) to assess the diversity and abundance of tfdA-like genes in soil.

MATERIALS AND METHODS

Soil samples and bacterial strain. Samples from four different agricultural or forest soils were used (Table 1). The agricultural soil was collected at the Klostergut Scheyern experimental farm (48°30′00″N, 11°20′07″E) and was treated with MCPA in 2002. The other three soil samples were obtained from different spruce forests in Germany, which were not treated with PAA herbicides (Table 1). All soil samples were obtained from a 0- to 10-cm depth and were stored in the dark on ice during transfer. Each soil sample was homogenized, sieved (2 mm), and stored at −80°C prior to use. Cupriavidus necator JMP134 (DSM 4058; formerly Ralstonia eutropha) was cultured according to the recommendations of the DSMZ (Braunschweig, Germany) and used for testing of primer specificity (Table 1).

Nucleic acid extraction. DNA was extracted from pelleted cells of 1 ml of Cupriavidus necator JMP134 culture (approximately 10⁸ cells) or 0.5 g (wet weight) of soil stored in the dark on ice during transfer. Each soil sample was homogenized, sieved (2 mm), and stored at −80°C prior to use. Cupriavidus necator JMP134 (DSM 4058; formerly Ralstonia eutropha) was cultured according to the recommendations of the DSMZ (Braunschweig, Germany) and used for testing of primer specificity (Table 1).

Primer specificity was evaluated in silico using GenBank’s BLAST tool (1), and PCR conditions were optimized by variation of Mg²⁺ and annealing temperature (Table 2). 16S rRNA genes were amplified with published primers (Table 2). Each amplification reaction was performed in a total volume of 25 μL, and the reaction mixture contained 1× PCR buffer (50 mM KCl, 10 mM Tris-HCl, 1.5 mM MgCl₂) supplemented with 1.0 to 2.5 mM MgCl₂, 1× TaqMaster PCR enhancer (5′-Prime), 60 ng/μL bovine serum albumin (BSA), 200 μM deoxynucleoside triphosphate (dNTP) mix (5′-Prime), 0.2 to 1 μM of each primer (Biomers, Ulm, Germany), 1 μL of template DNA (approximating 50 ng of DNA), and 0.5 U Taq DNA polymerase (5′-Prime). BSA was included in the reaction mixture to reduce PCR inhibition from humic substances coextracted from DNA from soil (38). The PCR was performed with a T-Gradient cycler (Biomat, Göttingen, Germany) with 10 min of initial denaturation at 95°C, followed by 45 cycles, each consisting of denaturation at 95°C for 1 min, annealing at primer-dependent temperatures (Table 2) for 1 min, and elongation at 72°C for 1 min. The final elongation was at 72°C for 5 min.

Quantification of genes in soil. qPCRs were performed with an iQ5 Real-Time qPCR cycler (Bio-Rad, Munich, Germany). All qPCRs were set up in duplicate. Negative controls with sterilized water instead of DNA template were included in every PCR setup. Standard curves were set up by serially diluting M13univec PCR products of a pGem-T vector with the appropriate insert from 10⁵ to 10³ target gene copies μL⁻¹ for every primer set. qPCRs were performed in 20-μL reaction mixtures that were composed of iQ SYBR green Supermix (Bio-Rad) and 10 mM forward and 10 mM reverse primers, 0.5 μL template DNA, and sterilized deionized water. Initial denaturation was at 95°C for 5 min, followed by 45 cycles of denaturation at 95°C for 40 s, annealing at temperatures appropriate for each primer set (Table 2) for 30 s, and elongation at 72°C for 15 s when fluorescence signal was recorded (i.e., a 3-step protocol was utilized). The final PCR elongation step was at 72°C for 5 min. qPCR with primer set 4 (Table 2) yielded a specific product with a melting temperature (Tm) of approximately 91°C and an unspecific product with a Tm of approximately 79°C (see Fig. S1A in the supplemental material). Agarose gel electrophoresis indicated that the latter product was primer dimers. The lower limit of quantification was 10⁴ gene copies μL⁻¹ of DNA extract (see Fig. S1C). Two more steps at 85°C (i.e., 6 s for preheating and 15 s for signal recording) were added after elongation to melt primer dimers prior to signal recording (i.e., a 5-step protocol was utilized). After application of the 5-step protocol, primer dimers were not detected during melting curve analysis (see Fig. S1B). Melting curve analyses were performed from 70 to 95°C with increments of 0.2°C per cycle. 16S rRNA gene copy numbers were determined concomitantly for all environmental samples (Table 2) in order to quantify tfdA-like genes in soil relative to total bacterial 16S rRNA genes.

Inhibition of qPCR by environmental DNA extracts. Humic acids and other organic compounds coextracted from soil with nucleic acids can inhibit PCR significantly (33, 72, 75, 76). Dilution of DNA extracts and correction of threshold cycle (Ct) values are two strategies for overcoming such inhibition. For testing the inhibitory potential of environmental DNA extracts, such extracts were treated with DNase I (Fermentas, St. Leon-Rot, Germany) at 37°C for 4 h according to the manufacturer’s protocol to obtain DNA-free soil extracts. Equal

TABLE 1. Soil characteristics and abundances of tfdA-like genes
TABLE 2. PCR primers and conditions evaluated in this study

<table>
<thead>
<tr>
<th>Primer set(^a)</th>
<th>Sequence (5(^\prime)–3(^\prime))(^b)</th>
<th>Experimental</th>
<th>Target</th>
<th>Approx product length (bp)</th>
<th>PCR product of agricultural field soil(^c)</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Annealing temp (°C)</td>
<td>Mg(^2+) (mM)</td>
<td></td>
<td>Native</td>
<td>JMP134 spiked</td>
</tr>
<tr>
<td>1</td>
<td>TfdA21F(\text{a}) TfdA836R</td>
<td>63.0</td>
<td>3.0</td>
<td>tfd4 (group 1)</td>
<td>815</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>++</td>
<td>This work</td>
</tr>
<tr>
<td>2</td>
<td>TfdA334F(\text{c}) TfdA608R</td>
<td>60.0</td>
<td>3.0</td>
<td>tfd4 (group 1)</td>
<td>260</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>++</td>
<td>This work</td>
</tr>
<tr>
<td>3</td>
<td>TfdA52F(\text{c}) TfdA408R</td>
<td>70.0 (65.0)(^d)</td>
<td>4.0 (3.0)(^d)</td>
<td>tfd4a (group 2)</td>
<td>360</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>++</td>
<td>This work</td>
</tr>
<tr>
<td>4</td>
<td>TfdA421F</td>
<td>63.6 (68.0)(^d)</td>
<td>2.5 (3.0)(^d)</td>
<td>tfd4a (groups 2 and 3)</td>
<td>360</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>TfdA779R</td>
<td></td>
<td></td>
<td></td>
<td>Modified from references 29 and 73</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>TfdA421dF</td>
<td>53.0</td>
<td>3.0</td>
<td>tfd4-like (groups 1 to 3)</td>
<td>360</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>TfdA778vK</td>
<td></td>
<td></td>
<td></td>
<td>Modified from references 29 and 73</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>TfdA421Fv</td>
<td>59.0</td>
<td>3.0</td>
<td>tfd4-like (group 1)</td>
<td>360</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>TfdA783R</td>
<td></td>
<td></td>
<td></td>
<td>Modified from reference 73</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Eub341F(\text{c}) Eub534R</td>
<td>55.7</td>
<td>3.0</td>
<td>16S rRNA</td>
<td>190</td>
<td>++</td>
</tr>
</tbody>
</table>

\(^a\) Numbers in primer names indicate 5\(^\prime\) position of binding sites relative to the reference tfd4 sequence of Cupriavidus necator JMP134 (AY365053) (primer sets 1, 2, 4, 5, and 6) and of Alphaproteobacteria RDS-5-C2 (AB074490) (primer set 3). F, forward; R, reverse.

\(^b\) S, C/G; Y, C/T; V, A/C/G; W, A/T; M, A/C; K, G/T; R, A/G.

\(^c\) As indicated by agarose gel electrophoresis. + +, bright band; ++, weak band; -, no band of correct size.

\(^d\) Numbers in parentheses refer to qPCR chemicals (see Materials and Methods).

FIG. 1. Effect of inhibitors present in soil DNA extracts on standard curves generated during qPCR with primer set 4. Solid lines indicate standards spiked with various concentrations of DNA-free soil extract (\(\bullet\), nondiluted; \(\Delta\), 10-fold diluted; \(\bigcirc\), 50-fold diluted); the dashed line indicates the nonspiked standard curve. Equations of linear regressions and \(R^2\) values of every curve are given.
hein, Germany), which was used for transformation of Escherichia coli JM109 competent cells (Promega, Madison, WI) according to the manufacturer’s protocol. For screening of the clone libraries from (q)PCR products, the fragments were directly amplified from 1 µl of resuspended clones and electrophoresed on agarose gels (1%). The gene libraries were screened by restriction fragment length polymorphism (RFLP). PCR products of clones generated with primer set 3 (Table 2) were digested with 1 U of MspI and TaqAl (New England Biolabs, Ipswich, MA) in separate reactions, and those generated with primer sets 4 and 5 (Table 2) were likewise digested with 1 U of MspI and TaqAl (New England BioLabs) in separate reactions for 3 h at 37°C (TaqAl at 65°C). The fragments were separated on agarose gels (3%). One to 10 representative clones of each RFLP pattern per library were selected for sequencing. Representative PCR products were purified (PCR cleanup plates; Millipore, Schwalbach, Germany) according to the manufacturer’s protocol and commercially sequenced (Macrogen, South Korea). Sequence comparisons and database searches were carried out by using the BLAST service at the NCBI website (http://blast.ncbi.nlm.nih.gov/).

Phylogenetic analyses. Phylogenetic analyses were performed with the ARB software package and Mega4 software (http://www.megasoftware.net/) (69). tfdA-like genes were retrieved from NCBI, translated in silico, and prealigned with the ClustalW algorithm, and the alignment was refined manually. The DNA was aligned according to the aligned proteins. Regions of primer binding sites were excluded from further sequence analysis. Phylogenetic trees were calculated based on amino acid sequences using the neighbor-joining algorithm (55). Tree topologies of neighbor-joining trees were verified by bootstrap analysis with 1,000 replicates.

Prealigned 16S rRNA gene sequences were retrieved from SILVA (http://www.arb-silva.de/) (52). Distance matrices for the correlation of pairwise similarities of tfdA-like and 16S rRNA gene sequences were generated with the ARB software package (40).

Analysis of tfdA-like genotype diversity. Sequences that had the same RFLP patterns as those of sequenced tfdA-like gene fragments were represented by the fully sequenced fragments. Distance matrices were generated from aligned amino acid sequences. DOTUR (http://schloss.micro.umass.edu/software/index.html) (57) was applied to define genotypes based on the amino acid sequence dissimilarity of proteins derived from translated tfdA-like sequences by the furthest-neighbor method. Sequences with maximal 3% or 20% dissimilarity (i.e., distance D = 3 or 20%) were thereby assigned to one genotype. Additional information on DOTUR is available at the publisher’s website (http://schloss.micro.umass.edu/software/index.html). Coverage (C) is the number of the detected genotypes relative to their expected total number in a gene library and was calculated as C = (1 - n × N⁻¹) × 100, where n is the number of genotypes that occurred only once and N is the number of clones screened (21). The diversity of genotypes represented in gene libraries of tfdA-like genes was analyzed by rarefaction analysis (28). Shannon diversity index, and the genotype richness estimator Chaol (24). All calculations were done with DOTUR (57). Rarefaction curves were produced using the Analytic Rarefaction software (version 1.2; http://www.uga.edu/strata/software/index.html).

Nucleotide sequence accession numbers. Sequences obtained in this study were deposited in the EMBL nucleotide sequences database (http://www.ebi.ac.uk) under accession numbers FN376574 to FN376845.

RESULTS

Novel primers for amplification of tfdA-like genes. Eight conserved regions of tfdA-like genes were used to design consensus primers (data not shown). Five different degenerated primer sets were designed (1 to 5, Table 2) to amplify (i) almost full-length fragments of the tfdA gene group 1, (ii) a partial gene fragment of tfdA gene group 1, (iii) a partial gene fragment of tfdA gene group 2, (iv) a partial gene fragment of tfdA-like genes covering group 2 genes, and (v) a partial gene fragment of tfdA-like genes targeting group 1 and 2 genes. Nevertheless, primer sets 4 and 5 amplified group 3 tfdA-like genes (Table 2; Fig. 2). In silico analyses indicated that primers of sets 1 and 2 had 1 to 4 mismatches to group 2 tfdA-like genes. Primer set 3 targeted a subset of group 2 tfdA-like genes and had more than 4 mismatches to group 1 tfdA. Primer set 4 covered all known group 2 tfdA-like genes and had one mismatch to group 1 tfdA. Primer set 5 had no mismatches to any known tfdA-like gene. In silico analyses of primer sets 4 and 5 were not performed with group 3 tfdA-like fragments since the fragments publicly available to date were truncated. PCR products generated with primer set 6 from Scheyern soil contained only low concentrations of targets with the expected size (Table 2). However, all primer sets that were tested readily amplified fragments of the expected sizes (Table 2) from Scheyern soil that was amended with 10⁴, 10⁵, and 10⁶ cells of C. necator JMP134 (harboring group 1 tfdA) per gram of soil, indicating that all primer sets detected target genes in soil (data not shown). The proportion of group 1 tfdA sequences per gene library (25 to 29 sequences per library) generated from Scheyern soil with primer set 5 approximated 15, 76, and 100% for soil amended with 10⁴, 10⁵, and 10⁶ cells of C. necator JMP134 per gram of soil, respectively, indicating that the detection limit of group 1 tfdA in Scheyern soil with primer set 5 is below 10⁴ copies per gram of soil (assuming one copy of group 1 tfdA per cell). Gene libraries generated from nonspiked Scheyern soil with primer sets 4 and 5 contained group 2- and 3-related tfdA-like genes. Primer set 3 yielded likewise specific products which were most closely related to group 2 tfdA-like genes (see Fig. S2 in the supplemental material). The results showed that the designed primers allowed a specific targeting of tfdA-like genes from soil.

Detected diversity of tfdA-like genes. A total number of 437 tfdA-positive clones (201, 119, and 117 clones for primer sets 3, 4, and 5, respectively) were screened by RFLP, and 210 representative clones were sequenced. Coverages approximated 93 and 100% at D = 3% and D = 20%, respectively, for the gene library generated with primer set 3 targeting a fraction of group 2 tfdA-like genes (Table 3; see also Fig. S2 and S3 in the supplemental material). Coverages approximated 62 and 91% at D = 3% and D = 20%, respectively, for the combined gene libraries generated with primers 4 and 5 targeting tfdA-like genes of groups 1 to 3 (Table 3; Fig. 2 and 3), indicating that the gene libraries were sufficiently sampled.

The gene library targeting group 2 tfdA-like genes (primer set 3) contained 34 and 6 genotypes at D = 3% and D = 20%, respectively (Table 3). The estimated numbers of genotypes present in the gene library were 47 and 6, respectively; Shannon’s diversity indices were 2.5 and 0.5, respectively (Table 3). All sequences were related to tfdA-like genes of cultured Alphaproteobacteria (Bradyrhizobium spp.) or tfdA-like genes re-
tried from other soils (see Fig. S2 in the supplemental material). Group 2 tfdA-like genes retrieved from Scheyern agricultural soil were assigned to 2 phylogenetic clusters (see Fig. S2). Cluster 2 contained sequences from Scheyern soil only, indicating that phylogenetically novel tfdA-like genes occur in Scheyern agricultural soil. The maximal phylogenetic distance of cluster 1 sequences was 23.3%, and the cluster contained Bradyrhizobium-related sequences. Cluster 1 sequences generated with primer set 3 (see Fig. S2) were represented by cluster 2a in gene libraries generated with primers 4 and 5 (maximal phylogenetic distance, 23.7% [Fig. 2]).

Primer sets 4 and 5 amplified similar fragments and detected similar groups of tfdA-like genes (Fig. 2). Thus, the gene libraries generated with such primers were combined for further analysis and contained 122 and 48 genotypes at $D \leq 3\%$ and $D \leq 20\%$, respectively (Table 3). The estimated numbers of genotypes present in the gene library were 389 and 78, respectively; Shannon’s diversity indices were 4.3 and 3.2, respectively (Table 3). All sequences of the gene library generated with primer set 3 (see Fig. S2) have consequently an even higher probability of belonging to different species. Group 2 sequences that share $\leq 82\%$ tfdA-like gene similarity (equivalent to $\leq 84\%$ protein similarity of translated tfdA-like genes) have a very high probability of belonging to different species. Group 2 sequences that share $\leq 77\%$ tfdA-like gene similarity (equivalent to $\leq 80\%$ protein similarity of translated tfdA-like genes) have consequently an even higher probability of belonging to different species.

**Quantitation and recovery of tfdA-like genes.** Primer sets 3 to 5 were used for quantification of tfdA-like genes in soil by

**Comparison of tfdA-like and 16S rRNA gene similarity.** The comparative phylogeny of tfdA-like and 16S rRNA genes of bacteria harboring tfdA-like genes was assessed to determine if tfdA-like genes could be used for the identification of such organisms. Pairwise similarities of tfdA-like and 16S rRNA gene sequences ranged from 76.3 to 100% and 75.5 to 100% for group 1, respectively. For group 2 sequences, pairwise similarities of tfdA-like and 16S rRNA gene sequences ranged from 80.4 to 100% and 97.4 to 100%, respectively. Such data indicate that 16S rRNA genes are more conserved than tfdA-like genes (Fig. 4). Group 1 organisms that shared $< 77\%$ tfdA-like gene similarity (or 80% protein similarity of translated tfdA-like genes) always shared $< 97\%$ 16S rRNA gene similarity (a conservative threshold similarity for assigning two organisms to different species [65, 66]) (Fig. 4). The scattering of data points for group 1 organisms (Fig. 4) indicated horizontal gene transfer of group 1 tfdA, making it difficult to predict 16S rRNA gene similarities and consequently new species based on group 1 tfdA-like gene similarities. However, there was a correlation between 16S rRNA and group 2 tfdA-like gene similarity, indicating little or no horizontal gene transfer. Approximately 90% of any group 2 organisms that shared $> 97\%$ 16S rRNA sequence similarity shared $> 82\%$ tfdA-like gene similarity (equivalent to 84% protein similarity of translated tfdA-like genes). Thus, group 2 organisms that share $\leq 82\%$ tfdA-like gene similarity (equivalent to $\leq 84\%$ protein similarity of translated tfdA-like genes) have a very high probability of belonging to different species. Group 2 sequences that share $\leq 77\%$ tfdA-like gene similarity (equivalent to $\leq 80\%$ protein similarity of translated tfdA-like genes) have consequently an even higher probability of belonging to different species.

**TABLE 3. Summary of data from amino acid analyses of in silico- translated tfdA-like gene sequence fragments retrieved from Scheyern agricultural soil**

<table>
<thead>
<tr>
<th>Primer set(s)</th>
<th>Target</th>
<th>$n$</th>
<th>$D^a$ (%)</th>
<th>$S^b$</th>
<th>$C^c$ (%)</th>
<th>Chao1$^d$</th>
<th>$H^e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 tfdA-like (group 2)</td>
<td>201</td>
<td>3</td>
<td>34</td>
<td>92.5</td>
<td>47</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>4 and 5 tfdA-like (groups 1 to 3)</td>
<td>236</td>
<td>3</td>
<td>122</td>
<td>61.9</td>
<td>389</td>
<td>4.3</td>
<td></td>
</tr>
</tbody>
</table>

$^a$ See Table 2.
$^b$ $n$, number of clones screened.
$^c$ $D$, maximal amino acid sequence dissimilarity used for the assignment of sequences to genotypes (see Materials and Methods).
$^d$ $S$, number of genotypes.
$^e$ $C$, coverage; see Materials and Methods for details.
$^f$ Richness estimator; indicative of estimated number of genotypes per gene library.
$^g$ $H'$, Shannon index.

![FIG. 3. Rarefaction curves of cloned tfdA-like gene fragments for an evolutionary distance of 3% (A) or 20% (B). tfdA-like sequences were retrieved from Scheyern soil with primer sets 4 and 5 (○) and primer set 3 (●). Dotted and solid lines indicate 95% confidence intervals.](http://aem.asm.org/)
Novel and diverse tfdA-like genes were detected in soil by primers developed in this study to match a broad diversity of tfdA-like genes (Fig. 2; Table 3; see also Fig. S2 in the supplemental material). All tfdA-like genes were phylogenetically distinct from tauD (Fig. 2; see also Fig. S2), which encodes an α-KG-dependent taurine dioxygenase with approximately 30% sequence identity to tfdA-like genes (74), indicating that the novel primers discriminate well against tauD. tfdA-like gene fragments recovered from an agricultural soil that was treated with MCPA in 2002 were assigned to 11 clusters and were indicative of up to 122 distinct genotypes (at D < 3 [Table 3]) mainly related to alphaproteobacterial tfdA-like genes or two hitherto-unknown groups (Fig. 2; see also Fig. S2). Other studies of tfdA-like gene sequence diversity in soil without prior enrichment and/or isolation are still rather scarce and detect primarily group 1 tfdA-like genes indicative of Beta- and Gammaproteobacteria (3–5, 20). However, enrichments and analysis of isolates from soils revealed the presence of alphaproteobacterial PAA degraders in soils (29, 34) and are consistent with the detection of alphaproteobacterial tfdA-like genes in this study. Alphaproteobacteria of the genus Sphingomonas numerically dominate PAA degraders in certain soils (32), and those of the genera Sphingomonas and Bradyrhizobium are often detected by isolation (11, 27, 34, 41, 51). Since highly diverse tfdA-like genes indicative of Alphaproteobacteria were detected in this study, Alphaproteobacteria might represent a widely distributed, more diverse group of PAA degraders in soil than previously thought.

Bradyrhizobium-related PAA degraders of the Alphaproteobacteria are slow-growing oligotrophs (doubling times approximate 20 h), which are sensitive to high concentrations of organic nutrients (34, 59). Such organisms were also isolated from pristine soils that never received PAA treatment (34). Group 2 tfdA-like gene products are capable of PAA conversion (29), indicating that those genes are potentially involved in PAA conversion in soil, despite the occurrence of alternative potential PAA-degradative genes like cadAB among certain Bradyrhizobia (30, 35). Group 2 tfdA-like gene products display a higher activity for nonchlorinated PAA than for chlorinated ones (29). Group 1 tfdA-like gene products are capable of utilizing phenoxyacetic acid and cinnamic acid derivatives which occur in soil and are produced by plants (references 26 and 61 and references therein). Cinnamic acid-related flavonoids enhanced PAA utilization by a Bradyrhizobium sp. strain, indicating that such naturally occurring compounds

**FIG. 4.** Correlation of 16S rRNA gene similarity with tfdA-like sequence similarity of pure cultures. Group 1 (●) and 2 (○) sequences are indicated. The solid line indicates 16S rRNA gene threshold similarity values for species delineation, and the dotted line indicates threshold similarity values below which sequences are indicative of novel species with high probability. The dashed line represents the 90% quantile of pairwise sequence comparisons with a 16S rRNA similarity of >97%.
might stimulate the expression of group 2 tfdA-like genes (58, 59). Since group 2 tfdA-like gene products are capable of converting PAA (29) and tfdA-like genes were abundant in the soil analyzed (Table 1), tfdA-like genes might have potential environmental relevance for (cometabolic) PAA degradation in soil. The primers developed in this study will facilitate a thorough evaluation of such a hypothesis in future studies.

The analysis of structural genes like tfdA might well be used for diversity studies of process-associated bacteria, but evidence for species-level diversity from such data might be obtained only if horizontal gene transfer of the structural gene in question is scarce. 16S rRNA gene similarities correlated with tfdA-like gene similarities for group 2 sequences but not for group 1 sequences (Fig. 4), indicating that horizontal gene transfer played a major role during evolution of group 1 rather than group 2 tfdA-like-gene-harboring organisms. Several lines of evidence support such a conclusion: (i) plasmid-encoded group 1 tfdA is common among PAA-degrading bacteria of the Betaproteobacteria (e.g., references 2, 14, and 18), (ii) such an encoded PAA-degrading function is transmissible (34, 46), and (iii) the degree of congruency between 16S rRNA and group 1 tfdA-like-gene-based phylogenetic trees is low (46). In contrast, the PAA-degrading function of group 2 and 3 tfdA-like gene-carrying organisms is not or only barely transmissible (34), the degree of congruency between 16S rDNA and group 2 tfdA-like-gene-based phylogenetic trees is high (30), and the GC contents of group 3 tfdA-like and housekeeping genes in Sphingomonas spp. are highly similar (30).

The virtual absence of horizontal gene transfer for group 2 and 3 tfdA-like genes allows for an estimation of species-level diversity based on group 2 and 3 tfdA-like gene analysis. Correlation analysis of tfdA-like and 16S rRNA gene similarity thus indicates that the number of tfdA-like genotypes that share less than 80% of translated protein sequences is a minimal estimate for species-level diversity of tfdA-like-gene-carrying bacteria (i.e., the true species-level diversity might be much higher). The calculated threshold similarities are well in agreement with those obtained for other structural genes (e.g., for narG, nosZ, dsrAB, and amoA gene fragments, they were 70 to 90% [25, 36, 37, 50]). The detected tfdA-like genes in this study consequently suggest that at least 48 distinct species of tfdA-like-gene-carrying bacteria occurred in agricultural soil (Table 3).

Such diverse tfdA-like genes were abundant in agricultural and pristine forest soils as indicated by up to $10^7$ copy numbers per gram (dry weight) soil (Table 1). tfdA-like genes accounted for up to 3% of bacterial 16S rDNA genes. Copy numbers obtained in other studies with group 1 tfdA-like-gene-specific qPCR assays vary from $10^2$ to $10^7$ per gram (dry weight) soil prior to PAA incubation (3, 5, 20), indicating that group 2 and 3 genes outnumber group 1 genes in certain soils and highlighting a hitherto unknown PAA degradation potential in soils.

All qPCR assays in this study included a correction for inhibitory effects of environmental DNA extracts on the DNA polymerase used. Inhibition of the DNA polymerase during qPCR results in an underestimation of gene copy numbers (43). Such inhibitors include humic acids derived from plant material (53, 76). There are many approaches to overcome inhibition of qPCR, such as increase of Mg$^{2+}$ concentration, addition of BSA, or dilution of template DNA (e.g., references 3, 23, and 38). DNA extraction and purification methods were also improved to reduce the amount of coextracted inhibitors (22, 75). However, all of these approaches failed in our study to avoid inhibition (Fig. 1 and data not shown). Thus, gene copy numbers were corrected for the inhibition caused by environmental DNA extracts. Since internal positive-control fragments are more suitable for TaqMan probe-based assays (12, 44), environmental DNA was spiked with an approximately 10-fold excess of the target gene to identify and correct for inhibition during qPCR. The quality of DNA extracts obtained with the same method from soils varies (63), necessitating the evaluation of inhibitory effects for every qPCR assay to minimize artificial variations in gene copy numbers.

Novel primers developed in this study enabled the detection of highly diverse group 2 and 3 tfdA-like gene sequences in soil (Fig. 2; see also Fig. S2 in the supplemental material). Such genes outnumber group 1 tfdA-like genes by several orders of magnitude in certain soils without prior enrichment on PAA (Table 1) (3, 4), indicating that PAA degradation is not the primary function of those genes. Inducers of gene expression of group 2 and 3 tfdA-like genes in soil and pure cultures need to be identified to shed light on the potential function of such genes. The present study provides a basis for such investigations.

ACKNOWLEDGMENTS

We are grateful to Stephan Schulz, Michael Schloter, Jean Charles Munch, Daniela Degelmann, and Steffen Kolb for provision of soil samples.

Support for this study was provided by the Deutsche Forschungsgemeinschaft (grant HO4042/1-1, DFG Priority Program 1315 “Biogeochemical interfaces in soil”), by the University of Bayreuth, and by a grant from the German Academic Exchange Service (DAAD) to Y.-J. Liu.

REFERENCES


Pesticide remediation in soils and water. John Wiley & Sons, Chichester, United Kingdom.


