Appl Microbiol Biotechnol (2017) 101:797–808
DOI 10.1007/s00253-016-7934-5

ENVIRONMENTAL BIOTECHNOLOGY

Identification of novel toluene monooxygenase genes in a hydrocarbon-polluted sediment using sequence- and function-based screening of metagenomic libraries

E. Bouhajja[1] · M. McGuire[2] · M. R. Liles[2] · G. Bataille[3] · S. N. Agathos[1,4] · I. F. George[5]

Received: 5 July 2016 / Revised: 6 October 2016 / Accepted: 12 October 2016 / Published online: 26 October 2016

Abstract The microbial potential for toluene degradation within sediments from a tar oil-contaminated site in Flingern, Germany, was assessed using a metagenomic approach. High molecular weight environmental DNA from contaminated sediments was extracted, purified, and cloned into fosmid and BAC vectors and transformed into *Escherichia coli*. The fosmid library was screened by hybridization with a PCR amplicon of the α-subunit of the toluene 4-monooxygenase gene to identify genes and pathways encoding toluene degradation. Fourteen clones were recovered from the fosmid library, among which 13 were highly divergent from known *tmoA* genes and several had the closest relatives among *Acinetobacter* species. The BAC library was transferred to the heterologous hosts *Cupriavidus metallidurans* (phylum Proteobacteria) and *Edaphobacter aggregans* (phylum Acidobacteria). The resulting libraries were screened for expression of toluene degradation in the non-degradative hosts. From expression in *C. metallidurans*, three novel toluene monooxygenase-encoding operons were identified that were located on IncP1 plasmids. The *E. aggregans*-hosted BAC library led to the isolation of a cloned genetic locus putatively derived from an Acidobacteria taxon that contained genes involved in aerobic and anaerobic toluene degradation. These data suggest the important role of plasmids in the spread of toluene degradative capacity and indicate putative novel *tmoA* genes present in this hydrocarbon-polluted environment.

Keywords Biodegradation · Toluene 4-monooxygenase · Metagenomics · Catabolic gene diversity · Expression in heterologous hosts · Macroarray

Electronic supplementary material The online version of this article (doi:10.1007/s00253-016-7934-5) contains supplementary material, which is available to authorized users.

✉ I. F. George
igeorge@ulb.ac.be

[1] Earth and Life Institute, Laboratoire de Génie Biologique, Université catholique de Louvain, Place Croix du Sud 2, boite L7.05.19, 1348 Louvain-la-Neuve, Belgium

[2] Department of Biological Sciences, Auburn University, 101 Rouse Life Science Building, Auburn, Alabama 36849, USA

[3] Earth and Life Institute, Biodiversity Research Centre, Université catholique de Louvain, Place Croix du Sud 4-5, Bte L.7.07.04, 1348 Louvain-la-Neuve, Belgium

[4] School of Life Sciences and Biotechnology, Yachay Tech University, San Miguel de Urcuquí, Ecuador

[5] Laboratoire d'Ecologie des Systèmes Aquatiques, Université libre de Bruxelles, Campus de la Plaine CP 221, Boulevard du Triomphe, 1050 Brussels, Belgium

Introduction

Benzene, toluene, ethylbenzene, and xylene (BTEX) are volatile monoaromatic compounds which collectively constitute one of the major families of pollutants in groundwater. They are discharged mainly from leaking underground gasoline tanks in gas stations and from petroleum and natural gas production sites (Gross et al. 2013). Toluene degradation has been widely used to model aerobic degradation of BTEX compounds in contaminated environments. A detailed view of these degradation pathways, the key enzymes, and their respective encoding genes are available in the biocatalysis/biodegradation database (Gao et al. 2010). Toluene 4-monooxygenase (T4-MO) is a multicomponent enzyme involved in the hydroxylation of the aromatic ring of toluene and its conversion to hydroxytoluene. It was purified and characterized for the first time in *Pseudomonas mendocina* KR1 that can use toluene as a sole carbon source (Whited and Gibson 1991). The *tmoA* gene encodes the α-subunit of

the hydroxylase component of the T4-MO complex. *tmoA* has been widely used as a marker to evaluate bacterial community dynamics during bioremediation of BTEX-polluted soils and aquifers (Hendrickx et al. 2006; Lee et al. 2011; Okunishi et al. 2012; Rakoczy et al. 2011).

Metagenomic approaches facilitate the culture-independent discovery of gene sequences from poorly characterized or totally unknown microorganisms (Brennerova et al. 2009; Kimura et al. 2010; Rondon et al. 2000; Sierra-Garcia et al. 2014). In particular, functional screening of metagenomic libraries has the potential to discover novel genes associated with an activity since prior knowledge of genetic sequences is not required (Gillespie et al. 2002; Leis et al. 2013; Singleton et al. 2012; Suenaga et al. 2007). The success of this method depends on the ability of environmental DNA to be expressed by a heterologous host. On the other hand, sequence-based screening of metagenomic libraries allows the exploration of gene diversity through previously described sequences (Leis et al. 2013). These two paths to explore catabolic diversity within metagenomic libraries are complementary and can provide valuable information about encoded microbiome functions.

In Flingern, Germany, a former gasworks site in operation for over 30 years contains sediments and groundwater that are highly contaminated with tar oil compounds and a plume of BTEX and polyaromatic hydrocarbons (order of tens of mg ml^{-1}). Toluene is the dominant contaminant and electron donor in the plume. Molecular methods were applied to characterize the distribution of specific toluene degraders and to localize the ongoing degradation process and assess natural attenuation potential (Anneser et al. 2010; Larentis et al. 2013; Winderl et al. 2008). Conditions in the contaminant plume are mainly anoxic (Winderl et al. 2008). In the lower plume fringe, the amount of toluene degraders was correlated with zones of increased anaerobic degradation as revealed by quantification of 16S ribosomal RNA (rRNA) bacterial genes and benzylsuccinate synthase genes (Winderl et al. 2008). Interestingly, Larentis et al. (2013) found that the number of *tmoA* genes reached a maximum in the heavily reduced core of the plume and did not increase in the more oxidized upper fringe of the plume. In order to understand this unexpected distribution of as-yet unidentified aerobic degraders, we proposed to apply a metagenomic library-based approach. We aimed to assess the microbial potential for toluene aerobic degradation in the sediment, using both sequence-based and function-based methods.

Materials and methods

Sediment collection

A sediment sample was collected in February 2006 from a quaternary sandy aquifer at a former gasworks site located in Flingern, Germany. During operation and breakdown of the plant, tar oil compounds had been released in the soil underground and a plume of monoaromatic and polyaromatic hydrocarbons developed over 200 × 40 m (horizontal dimensions). This plume consisted mainly of BTEX at a concentration of 20 mg l^{-1} (Anneser et al. 2010). The sediment sample was collected from 6.7 to 7.1 m below the surface in the anoxic zone of the plume. It contained 8.8×10^6 cells g^{-1} of wet weight sediment (Anneser et al. 2010). The sample was frozen on dry ice and stored at −20 °C for subsequent DNA extraction.

DNA isolation and purification

An improved DNA extraction method was developed to recover high molecular weight (HMW) DNA from the studied sediment and of sufficient purity for cloning. This protocol overcame the low cell density within the sediment, which was inferior to the minimum generally required for metagenomic library construction (~10^{11} cells; van Elsas et al. 2008). Detailed steps of DNA isolation and purification are provided in Supplementary Material, Materials and Methods S1. Briefly, DNA was extracted from 10 g of sediment using a modified version of the protocol of Zhou and colleagues (1996). The recovered HMW crude DNA was desalted and concentrated using a centrifugal concentrator Vivaspin 500 Membrane 10,000 molecular weight cutoff PES, Sartorius Stedim Biotech, Germany, according to the manufacturer's instructions. For further purification, remaining humic substances were removed by Q Sepharose (Jackson et al. 1997; Mettel et al. 2010) following the protocol described by (Sharma et al. 2007).

Construction and genetic screening of the fosmid library

Construction of fosmid library

The purified HMW DNA was used to prepare a fosmid metagenomic library to be used for sequence-based screening. The fosmid pCC1FOS was chosen due to its ease of use in constructing metagenomic libraries from different environments and especially from hydrocarbon-polluted soils. Indeed, novel aromatic hydrocarbon degradation genes encoding extradiol dioxygenases, monooxygenases, and degradation pathways were discovered and characterized from fosmid libraries (Brennerova et al. 2009; Kimura et al. 2010; Sierra-Garcia et al. 2014). DNA was blunt-end ligated into a linearized and a dephosphorylated vector and transformed into *Escherichia coli* EPI300-T1R. The library was produced using the CopyControl™ Fosmid Library Production Kit from Epicentre containing the ready-to-use fosmid pCC1FOS according to the manufacturer's instructions. The library was organized into two macroarrays for screening by hybridization for *tmoA* genes. A total of 55,296 clones were robotically

picked in duplicate onto two nylon membranes at LibraGen (France) and organized in 384-well microplates for further storage at −80 °C. Every membrane contained the equivalent of 72,384-well microplates in duplicate. According to the position on the gridded membrane, each corresponding clone can be retrieved in the cryopreserved library at −80 °C. The average insert size was 40 kb per clone based on randomly selected fosmid clone restriction digests resolved via agarose gel electrophoresis (data not shown). Therefore, the fosmid library corresponded to approximately 530 *E. coli* genome equivalents.

Library screening by colony blot hybridization

The HMW DNA that was used for the construction of the library was also used as template to generate a *tmoA* PCR amplicon probe labeled with digoxigenin-11-dUTP (DIG-11-dUTP) using the PCR DIG Synthesis Kit (Roche, Switzerland) according to the manufacturer's instructions. Primers were TMOA-F 5′ CGAAACCGGCTT(C/T)ACCAA(C/T)ATG 3′ and TMOA-R 5′ ACCGGGATATTT(C/T)TCTTC(C/G)AGCCA 3′ (Hendrickx et al. 2005). They targeted a very diverse pool of *tmoA* genes within the studied sediment as shown by Larentis et al. (2013). The reaction conditions were as follows: denaturation at 94 °C for 2 min; 30 cycles of 94 °C for 30 s, 60 °C for 30 s, and 72 °C for 40 s; and a final 7-min cycle for elongation at 72 °C. These primers target the α-subunit of the hydroxylase component of the multicomponent toluene monooxygenases from Proteobacteria. The macroarrays were screened by colony blot hybridization at 55 °C with medium hybridization stringency. If the stringency was too high, only hits highly similar to the *tmoA* probe sequences would have been retrieved and the opportunity to obtain novel sequences would have been reduced. The detection of hybridization was performed using the digoxigenin detection system (Roche, Switzerland). Positive clones were detected with the chemiluminescent substrate CDP-Star (disodium 2-chloro-5-(4-methoxyspiro[1,2-dioxetane-3,2′-(5-chlorotricyclo[3.3.1.13.7]decan])-4-yl]-1-phenyl phosphate) for alkaline phosphatase. After hybridization, macroarrays were exposed to X-ray films for 40 min. X-ray films were developed with an OPTIMAX X-ray film processor. Clones were considered positive when they exhibited chemiluminescence in duplicate.

Verification of hybridization-positive clones

Fosmid clone DNAs were extracted from hybridization-positive (hyb+) clones and fixed on a nylon membrane. A second round of hybridization was conducted using the same probe mixture under the same conditions in order to confirm the positive signal observed on the macroarrays. Moreover, hyb+ clones were tested by PCR for amplification of the *tmoA* with the primer set TMOA-F/TMOA-R. Approximately 20 ng of recombinant fosmid DNA was added to a PCR mix containing 1× Taq polymerase Green Master Mix (Promega, Madison, USA) and 1 μM of each primer. The reaction conditions were as follows: denaturation at 94 °C for 5 min; 30 cycles of 94 °C for 30 s, 60 °C for 30 s, and 72 °C for 1 min; and a final 10-min cycle for elongation at 72 °C. Lastly, all hyb+ clones were subjected to restriction digestion with *BamHI* enzyme (Promega) to predict the diversity of inserts and to estimate the insert size of each clone.

Construction and functional screening of BAC libraries

The purified HMW DNA was cloned first into the bacterial artificial chromosome pBAC-SBO vector (Lucigen Corporation, Middleton, WI) in *E. coli* EC100 to ensure long-term storage of DNA in a well-known and easy-to-handle host. Then, two alternative bacterial hosts from the phyla Acidobacteria (*Edaphobacter aggregans* DSM 19364) and Proteobacteria (*Cupriavidus metallidurans* CH34 DSM 2839) were explored for functional metagenomic screening of the library.

BAC library in E. coli EC100

Generally speaking, BTEX-degrading enzymes are encoded by 10- to 15-kb operons (Bosch et al. 1999; Lau et al. 1994). For the first time, we used the inducible-copy bacterial artificial chromosome pBAC-SBO vector to clone HMW DNA from a sediment sample polluted with hydrocarbons. This vector was kindly provided by Dr. Ronald Godiska from the Lucigen Corporation (Middleton, WI). It allows preferential cloning of large fragments (up to 200 kb in size) and transfer of libraries from the traditional host *E. coli* to other hosts (Wexler and Johnston 2010). The pBAC-SBO backbone contains an origin of replication for a single-copy-state control, a *repE* gene, and a *par*A-B-C partition locus. Medium copy state is controled by *oriV* which is active only in the presence of the TrfA replication protein. Expression of *trfA* is under the control of the *ara*C-P promoter whose expression is induced by L-arabinose. It also contains a chloramphenicol (Cm) resistance gene as a selection marker. A similar copy-inducible BAC vector containing the RK2 mini-replicon *oriV* and an arabinose-inducible *trfA* has been previously published by Kakirde et al. (2011). Details of BAC library construction and transfer are in Supplementary Material, Materials and Methods S2. Approximately 0.25 μg of HMW DNA was ligated into 0.5 μg of pBAC-SBO using the Fast-Link™ DNA Ligation Kit (Epicentre, Madison, USA) according to the manufacturer's recommendations. Ready-to-use electrocompetent *E. coli* EC100 cells (Epicentre) were electroporated with 2 μl of ligation product. A BAC library containing 83.6 million clones was recovered on selective LB agar (Difco, France) plates containing 12.5 μg ml^{-1} Cm. An aliquot of the library was induced

with 0.01 % (*w*/*v*) of arabinose in order to isolate cloned DNA and to transfer it afterwards into alternative hosts.

Transfer of the BAC library into C. metallidurans CH34 and E. aggregans DSM 19364

C. metallidurans CH34 is able to degrade toluene (Mergeay et al. 1978). In *C. metallidurans* CH34, the toluene 4-monooxygenase operon (*tmo*) is composed of five successive genes: *tmoA*, *tmoB*, *tmoC*, *tmoD*, and *tmoE* (Janssen et al. 2010). The *tmoB*, *tmoC*, and *tmoD* genes were replaced by a trimethoprim (Tm) resistance cassette to suppress toluene degradation and therefore allow screening of BAC clones for functional complementation of toluene degradation. Growth of the resulted mutant CH34 Δ*tmoBCD*::Tm(TmR) (referred to hereafter as mCH34) on solid Tris salt mineral MM284 with toluene as the only carbon source was almost completely inhibited (ultra-small colonies, data not shown). Toluene degradation in liquid MM284 was not completely inhibited; however, it was lower than in the wild-type strain (Table 1). We explain this residual activity by the presence in *C. metallidurans* CH34 of two additional multicomponent monooxygenases (two phenol hydroxylases) (Janssen et al. 2010), which reflect the existence of mixed pathways resulting in wider substrate versatility and degradation efficiency (Notomista et al. 2003).

E. aggregans DSM 19364 is naturally unable to degrade toluene (data not shown); therefore, it was used as a heterologous host for the BAC library without modification. Electrocompetent *C. metallidurans* mCH34 and *E. aggregans* DSM 19364 cells were prepared according to Taghavi et al. (1994) and to a modified protocol of Trevors and Starodub (1990), respectively. Electrocompetent *C. metallidurans* mCH34 and *E. aggregans* DSM 19364 cells were electroporated with 2 μl of the BAC library DNA previously extracted from the host EC100 at 2500 V in 2-mm-gap electrocuvettes and were immediately transferred into 1 ml of pre-warmed LB or R2A (pH 5.5) broth at 30 °C for hosts mCH34 (incubation 1 h) and *E. aggregans* DSM 19364 (incubation 96 h), respectively, before spreading on selective media. The BAC library hosted in *C. metallidurans* mCH34 (referred to hereafter as "BAC mCH34 library") had 1.3×10^6 clones per milliliter on MM284 agar supplemented with 0.2 % sodium gluconate, 100 μg ml^{-1} Tm, and 900 μg ml^{-1} Cm after 2 days of incubation at 30 °C. The BAC library hosted in *E. aggregans* DSM 19364 (referred to hereafter as "BAC DSM 19364 library") had 1.69×10^7 clones per ml on R2A agar (pH 5.5) supplemented with 400 μg ml^{-1} Cm after 10 days of incubation at 30 °C.

Table 1 Toluene degradation by BAC clones expressed in *C. metallidurans*

Sample	Final toluene concentration in the liquid phase after 10 days of incubation (mM)	Decrease in toluene concentration (%) after 10 days of incubation
Blank	0.805 ± 0.004	4.64 ± 0.61
mCH34	0.76 ± 0.004	33.44 ± 0.69
Wild-type CH34	0.53 ± 0.003	78.84 ± 0.73
pBmCH34T-1	0.17 ± 0.001	61.47 ± 0.79
pBmCH34T-2	0.31 ± 0.002	59.06 ± 0.32
pBmCH34T-3	0.329 ± 0.001	4.64 ± 0.61

All values are the average for three replicates with standard deviations for each sample. Toluene degradation values were statistically different. Mann-Whitney test between pairs of values: all p values were <0.005 for the final toluene concentration and for the decrease in toluene concentration

Functional screening of BAC libraries for toluene degradation

An equivalent of 65,000 CFU of the BAC mCH34 library and an equivalent of 845,000 CFU of the BAC DSM 19364 library were spread on MM284 solid medium supplemented with 100 μg ml^{-1} Tm + 900 μg ml^{-1} Cm or 400 μg ml^{-1} Cm, respectively. Toluene was added in the vapor phase. Plates were incubated at 30 °C for 15 days under aerobic conditions inside closed jars to avoid toluene evaporation. A total of three BAC mCH34 clones were able to grow on the selective medium. They harbored fragments of environmental DNA which complemented the host mutation for toluene usage as a sole C source. However, none of the BAC DSM 19364 clones were identified that could grow on a minimal medium (pH 5.5) with toluene as a sole C source. Therefore, the BAC DSM 19364 library was spread on R2A agar (pH 5.5) + 400 μg ml^{-1} Cm and toluene was added in the vapor phase to induce the expression of toluene degradation genes.

Analysis of toluene degradation

Toluene consumption by each of the three BAC mCH34 clones that could use toluene as a sole C source was measured individually in triplicate in 5 ml of culture in MM284 broth ($OD_{600\ nm} \sim 0.1$) under aerobic conditions. The medium was supplemented with 100 μg ml^{-1} Tm, 900 μg ml^{-1} Cm, and 2 μl of pure toluene (~0.8 mM in the aqueous phase) (AnalaR Normapur®, Belgium) in the vapor phase within 15-ml serum bottles sealed with Teflon stoppers. Biodegradation controls were included: medium and toluene alone ("blank") and medium, toluene, and *C. metallidurans* mCH34 strain without cloned DNA ("mCH34"). After 10 days of incubation, a 1-ml aliquot was sampled from each culture and mixed with an equal volume of methanol. Cells were removed by centrifugation for 3 min at 10,000×*g*. Toluene concentration in the aqueous phase was measured with an UHPLC Agilent 1290 Infinity fitted with a C18 ZORBAX Eclipse HAP column

(50 × 2.1 mm, particle size 1.8 μm). Samples were eluted isocratically with water-acetonitrile (1:1) at a flow rate of 0.6 ml min^{-1}, and detection was monitored at 210 nm. Toluene concentration was calculated using external calibration with pure toluene. Evaluation of toluene consumption by clones was based on measurement of toluene concentration that remained in the liquid phase.

Sequencing and data analysis

BAC and fosmid clone DNAs were extracted using a standard alkaline lysis protocol followed by a phenol/chloroform purification. Host genomic DNA was selectively degraded using Plasmid-Safe™ ATP-Dependent DNase (Epicentre). Purified DNAs were quantified by Quant-iT™ PicoGreen® dsDNA Assay Kit (Invitrogen™). Fifty nanograms of each clone was used to prepare indexed paired-end libraries for subsequent cluster generation and DNA sequencing using the Illumina Nextera® DNA Sample Preparation Kit (Illumina, CA, USA). All clones were sequenced in the Genomics and Sequencing Laboratory at Auburn University (Alabama, USA) using a HiSeq 1500 sequencer following the rapid run mode. Reads of 100 bp length were filtered and assembled into contiguous fragments (contigs) using the CLC Genomics Workbench assembly algorithm (CLC bio, MA, USA). Reads were first mapped to vector and host genomes to discard contaminant sequences. Then, remaining unmapped reads were assembled de novo. They led to one or more than one contig per clone (size ranging from 5 to 127.104 kb, average coverage ranging from 30× to 1112×).

Contigs were analyzed for the presence of genes related to toluene resistance. Fully sequenced inserts were subjected to gene prediction and automatic annotation using the Joint Genome Institute (JGI) server. Automatic annotation of each clone was corroborated with manual annotation. This step consisted in comparing JGI-identified ORFs with ORFs found by GLIMMER 3.02 and NCBI's ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). All open reading frames (ORFs) were searched against reference *tmo*A genes using the tBLASTx algorithm. The function of the remaining predicted ORFs was searched against the RefSeq protein database by BLASTp. The list of *tmoA* genes was recovered from the NCBI database and included 36 sequences from cultivated representatives or environmental DNA (referred to as the *tmoA* local database, Supplementary Material, Materials and Methods S3). In BAC mCH34 and fosmid clones where full inserts could not be assembled, the pool of contigs was mined with *tmoA* local database using the tBLASTx algorithm. Only contigs containing sequences that matched with one (or several) *tmoA* genes in the local database were searched for ORFs. A function was then assigned to predicted ORFs in those contigs by BLASTp using the RefSeq protein database. In fosmids, the ORF matching with the *tmoA* local database was called "hybridizing ORF." Unless specified, ORFs were compared to genes in databases based on their deduced amino acid sequences (and not their nucleotide sequences) throughout the "Results" and "Discussion" sections. Taxonomic affiliation of each insert was inferred by calculation and comparison of oligonucleotide frequencies (k-mer) with publicly available sequenced genomes on the open access PhyloPythiaS web server (http://phylopythias.cs.uni-duesseldorf.de/) (Patil et al. 2012).

Nucleotide sequence accession numbers

All sequenced BAC and fosmid clones reported in this study were deposited in GenBank under accession numbers KX244616–KX244633.

Results

Sequence-based screening of the metagenomic library

Identification of positive clones

The fosmid macroarrays contained 53,760 clones and were screened by colony blot hybridization to identify *tmoA* genes encoding the α-subunit of T4-MO. A total of 156 clones (called pFDuss for fosmid clone from Düsseldorf) exhibited great intensity of chemiluminescence after hybridization and were confirmed positive by a second round of hybridization, indicating the presence of a *tmoA* homologous sequence and therefore potentially a toluene degradation pathway. The preliminary test to evaluate the novelty of clones recovered by hybridization was to amplify *tmoA* genes by PCR. Only 34 clones out of 156 gave *tmoA* amplicons (PCR+/Hyb+). These clones were not considered as novel and were not kept for further analysis. Clones in which *tmoA* could not be amplified (i.e., which contained mismatches with the TMOA-F/TMOA-R primer set) were predicted to contain *tmoA* genes divergent from previously described sequences. Therefore, these 122 clones (PCR−/Hyb+) were clustered into groups according to their restriction digestion profiles with the enzyme *BamHI*. We selected 14 clones as representative of each restriction digestion profile for Illumina sequencing.

Sequence analysis of fosmid clones hybridizing with tmoA probe

Full fosmid clone inserts were successfully assembled for clones pFDuss5, pFDuss104, and pFDuss116. Hybridization sites were identified by BLASTn (E value cutoff $<10^{-4}$) using the list of reference *tmoA* sequences as subjects. The contig size, predicted hybridization site, top tBLASTx hit for the hybridization site, percent identity to the top BLAST hit,

and E value were determined for each of these fosmid clones (Table 2), and the complete annotation of each fosmid clone was also conducted (see Supplementary Material, Table S3).

In pFDuss5, the hybridizing ORF pFDuss5_35 was similar (83 % identity, E = 9×10^{-3}) to the unassigned clone DV7tmo19 (Larentis et al. 2013). Calculation of k-mer frequencies suggests that the cloned sequence from pFDuss5 was derived from a close relative of an *Acinetobacter* species. None of the predicted genes, other than pFDuss5_35, was predicted to be directly involved in toluene degradation.

The clone pFDuss116 contained a fragment that shared 95 % of identity for 56 % of the sequence coverage with pFDuss5, with both clones containing very similar encoded functions. In pFDuss116, the hybridizing ORF pFDuss116_16 was predicted to share some similarity (29 % of identity, E = 2×10^{-3}) with the toluene hydroxylase of *Ralstonia pickettii* 12 D. All of the annotated ORFs were significantly similar to those from *Acinetobacter lwoffii*. ORF pFDuss116_16 was adjacent to genetic loci not predicted to be involved in toluene degradation.

In the clone pFDuss104, the hybridizing ORF pFDuss104_28 showed some similarity with a ferritin-like superfamily domain in *Ralstonia eutropha* JMP134 (31 % identity, E = 10^{-11}). Also contained within clone pFDuss104 was an ORF predicted to encode ferredoxin subunits of ring-hydroxylating dioxygenases (Rieske [2Fe-2S] domain) (pFDuss104_21) and an ORF predicted to encode a Fe-S-cluster-containing hydrogenase component (pFDuss104_22). Calculation of the k-mer frequency from sequences of pFDuss104 suggests that the clone was derived from a *Burkholderia* species. The source bacterium is likely to have two *upper* degradation pathways, one involving toluene monooxygenase and the other toluene dioxygenase (Janssen et al. 2010; Olsen et al. 1994).

Despite the high sequencing depth of the Illumina technology, we were unable to assemble de novo full inserts for the remaining 11 clones because of the over-abundance of vector and host reads. Therefore, these clones were re-assembled based on mapping assembly of reads against the pCC1FOS backbone sequence. For each clone, contigs were sorted according to their size (>500 bp) and homology to *tmoA* probe sequences. Selected contigs ranged from 609 to 7321 bp in length. The hybridization site, the top *tmo*A hit, the annotation of each contig, and the fosmid clone it belongs to are detailed in Supplementary Material, Table S4. The clones pFDuss22, pFDuss57, and pFDuss87 had short contigs hybridizing with *tmoA* (size ranging from 561 to 1376 bp). They had low similarity (34 to 36 % identity, E value between 2×10^{-3} and 10^{-9}) with Tmo*A* of *Comamonas* sp. E6, *C. metallidurans* CH34, and *Ralstonia* PHS1, respectively. Calculation of the k-mer frequency for all contigs of pFDuss57, pFDuss87 and pFDuss22 suggested that the two first clones harbored fragments from the Enterobacteriaceae family, while the last clone harbored fragments from the δ-Proteobacteria phylum.

In pFDuss7 and pFDuss4, the hybridizing ORF was predicted to encode a protein similar to the Tbc2A-toluene monooxygenase of *Burkholderia cepacia* JS150 (47 and 31 % identity, E = 3×10^{-4} and 2×10^{-3}, respectively). In pFDuss7, it was adjacent to a locus encoding a protein highly similar to a *p*-hydroxybenzoate 3-monooxygenase in *Pseudomonas syringae*. The latter enzyme has an oxidoreductase activity for electron transfer from NADH to FAD in the toluene monooxygenase system (Entsch and van Berkel 1995). This suggests that the source bacterium may contain at least two different toluene monooxygenases. Calculation of k-mer frequency revealed that pFDuss7 and pFDuss4 inserts were likely originating from a γ-*Proteobacterium* phylum and an *Acinetobacter* sp., respectively. The clones pFDuss40, pFDuss35, pFDuss96, and pFDuss125 had hybridizing ORFs encoding proteins somewhat similar (29 to 43 % identity, E = 1.3×10^{-2} to 3×10^{-3}) to toluene monooxygenase and toluene/phenol hydroxylase found in *Ralstonia*, *Comamonas*, and *Methylibium* species, respectively. Such contigs had a k-mer signature of γ-*Proteobacteria*. Clones pFDuss14 and pFDuss102 had hybridizing sequences that showed similarity (47–50 % identity, E = 3.2×10^{-2} to 7×10^{-3}) with those previously identified by a PCR amplicon study from the same sediment by Larentis et al. (2013). These two clones had k-mer signatures of γ-*Proteobacteria* and were most probably derived from an *Acinetobacter* species.

Table 2 Fosmid clones that contain a *tmoA* gene and had a complete insert sequence

Clone	Largest contig (bp)	Hyb site (start nt–end nt)	Top TmoA tBLASTx hit		
			Hit	ID (%)[a]	E value[a]
pFDuss5	32,235	27,128–27,579	DV7tmo19/JX307454	83	9×10^{-3}
pFDuss104	28,715	23,790–24,771	*R. eutropha* JMP134/AF065891	31	1×10^{-11}
pFDuss116	32,624	13,187–13,993	*R. pickettii* 12 D/CP001645.1	29	2×10^{-3}

Hyb hybridization, *nt* nucleotide, *ID* identity

[a] Identity and E value correspond to alignment of translated ORF against translated subject (Hit)

Functional screening of the BAC library

Functional screening of BAC library in C. metallidurans CH34 host

The screening of 65,000 BAC clones expressed in *C. metallidurans* CH34 for the ability to grow on MM284 agar medium with toluene as a sole C source resulted in the identification of three recombinant clones. The biodegradation of toluene by each of the three BAC clones expressed in *C. metallidurans* CH34 was higher than the wild-type *C. metallidurans* CH34 as shown in Table 1.

The largest assembled contigs from the three sequenced clones, pBmCH34T-1 (50.879 kb, GC = 61.5 %, 65 predicted proteins (PP)), pBmCH34T-2 (26.417 kb, GC = 61.04 %, 33 PP), and pBmCH34T-3 (11.238 kb, GC = 63.1 %, 19 PP), were 100 % identical (E = 0.0) to fragments from backbones of IncP1 plasmids in the GenBank nr/nt database. These include genes encoding proteins involved in plasmid partitioning (*par* operon), replication (*tra* operon), incompatibility (*kor* operon), maintenance (*klc* operon), and mating pair formation during conjugation (*trb* operon). The clone pBmCH34T-1 contained a fragment similar to the Birmingham IncP1-α RK2 plasmid of *Pseudomonas aeruginosa* (Pansegrau et al. 1994). Fragments in clones BmCH34T-2 and BmCH34T-3 were similar to pWEC911 and PB11/PSP21/PB5 (Szczepanowski et al. 2011), respectively. To find contigs carrying potential toluene monooxygenase-encoding genes within the pool of smaller assembled contigs, a tBLASTx analysis was conducted against the local *tmoA* database. The predicted ORFs of positive hits (E < 0.001) were also evaluated for their similarity to *tmo*B, *tmo*C, *tmo*D, *tmo*E, and *tmo*F genes obtained from the KEGG Orthology database. Annotation of contigs containing a potential *tmo*-like cluster was conducted for each of the BmCH34T clones (see Supplementary Material, Table S1). Three contigs were recovered from BmCH34T-1 (5.857 kb), BmCH34T-2 (1.49 kb), and BmCH34T-3 (4.213 kb). A tBLASTx of their predicted ORFs showed the presence of a potential *tmo* cluster (Fig. 1a).

Functional screening of BAC library in E. aggregans DSM 19364 host

The screening of 845,000 BAC clones expressed in *E. aggregans* DSM19364 on R2A with toluene vapor resulted in the identification of one clone, designated as pBEagY that produced a distinct yellow pigmentation. The recombinant BAC was purified and used to transform against wild-type *E. aggregans* DSM19364. The same pigmentation was observed in the newly transformed *E. aggregans* in the presence of toluene, whereas it was absent in the absence of toluene (data not shown).

The BEagY clone insert was completely sequenced. It was 127.104 kb long with a 54.1 % G + C content. In total, 124 ORFs were identified among which 17 % were hypothetical with unknown functions, 57 % were assigned to clusters of orthologous groups (COGs), 56 % were assigned to protein families in the Pfam database, and 19 % were similar to KEGG orthologous (KO)groups (see Supplementary Material, Table S2). Predominant proteins belonged to the replication, recombination, and repair KO group and to intracellular trafficking, secretion, and vesicular transport COGs (conjugal transfer proteins and type IV secretion systems) (see Supplementary Material, Fig. S1). Only one gene

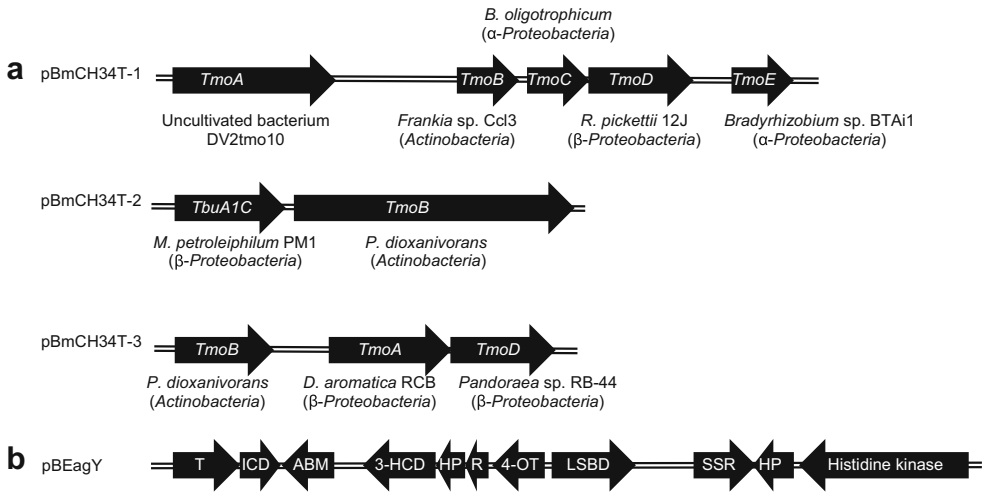

Fig. 1 **a** Organization of genes within *tmoA-like* clusters in pBAC mCH34T clones. *Arrows* indicate the annotation of ORFs, and below are the neighbors with their respective classes in between brackets. **b** Gene organization of the region involved in toluene degradation by pBEagY. *T* Transposase and inactivated derivatives, *ICD* integrase core domain, *ABM* antibiotic biosynthesis monooxygenase, *3-HCD* 3-hydroxyacyl-CoA dehydrogenase, *HP* hypothetical protein, *R* RmlC-like cupin domain, *4-OT* oxalocrotonate tautomerase, *LSBD* LysR substrate binding domain, *SSR* site-specific recombinase

(BEagY_107) was predicted to belong to the xenobiotic degradation KO category.

The majority of annotated genes showed similarity (30 to 90 % identity) with ones from the Acidobacteria taxa, including the species *Acidobacterium capsulatum*, *Candidatus Koribacter versatilis*, *Terriglobus roseus*, and *Granulicella mallensis* and other unclassified *Acidobacteria* taxa (82 genes). Twelve ORFs had their top BLASTp hit within *Proteobacteria* taxa (see Supplementary Material, Table S2). Phylogenetic affiliation based on the identified 13 housekeeping genes, including transcriptional regulators, DNA/RNA helicases, and site-specific recombinases, suggests that the cloned insert in pBEagY is affiliated with an as-yet-uncultured *Acidobacteria* taxon.

The ORF pPBEagY_107 was predicted to encode a protein similar to 4-oxalocrotonate tautomerase (PF01361) of *Pedosphaera parvula* Ellin514 (*Verrucomicrobia*) (70 % identity). This enzyme converts 2-hydroxymuconate to 2-oxo-3-hexenedioate in the *meta*-cleavage pathway of catechol degradation. It is involved in toluene, o-xylene, 3-ethyltoluene, 1,2,4-trimethylbenzene, and dioxin conversion into intermediates of the citric acid cycle. The 4-oxalocrotonate tautomerase locus was associated with PBEagY_108 encoding a LysR transcriptional regulator (Fig. 1b). The predicted protein was similar to a bacterial regulatory helix-turn-helix protein from *Acidobacteriaceae* bacterium KBS 96 (70 % identity). Transcriptional regulators from this family are known to regulate aerobic degradation pathways that act via catechol or protocatechuate (Wang et al. 1995). In the same region, PBEagY_104 was predicted to encode a protein similar to 3-hydroxyacyl-CoA dehydrogenase of *Sphingomonas* sp. PAMC 26621 (76 % identity), an enzyme assumed to be involved in anaerobic toluene degradation (Carmona et al. 2009; Heider et al. 1998).

Discussion

In this study, we developed a *tmoA* gene-targeted metagenomic approach to explore the diversity of genes and microorganisms involved in monooxygenase-mediated toluene degradation in a sample of sediment taken from the former gasworks site in Flingern, Germany. The native microbial community was expected to have great potential to break down the released hydrocarbons and to induce a natural recovery of the ecosystem (Scow and Hicks 2005). The sample was collected in the upper plume fringe, considered as the "hot spot" of biodegradation compared to the plume core (Jobelius et al. 2011; Meckenstock et al. 2015; Winderl et al. 2008).

This study followed an approach based on function- and sequence-based screening of DNA clone libraries combined with next-generation sequencing of positive clones. Compared to shotgun sequencing of microbial community DNA, this approach was expected to provide a broader coverage of genes linked to toluene degradative capacity as well as information on the phylogenetic origin of these sequences and their linked functions (Suenaga 2012). Cloning of large contiguous DNA fragments increases the probability to identify clones that contain entire biodegradative pathways (Liles et al. 2008). In addition, the use of multiple hosts beyond *E. coli* might improve the diversity and the number of hits in metagenomic libraries (Craig et al. 2010; Ekkers et al. 2012), because heterologous expression of genes is more efficient in their phylogenetically related hosts (Angov 2011; Ermolaeva 2001). In light of our results, this "multiple vector, multiple host" strategy opens perspectives for further gene expression experiments and application in biotechnological processes. Note that in this study, the two screening strategies gave very different results. We explain it by different cloning efficiencies of fragments of different lengths and composition according to vector and host.

The fosmid library was subjected to a high-throughput sequence-based screening on a nylon membrane using a *tmoA* gene probe. Among 55,296 clones, 14 clones were positive, among which 13 had very low scores of alignment of their translated sequence against T4-MO amino acid sequences. This significant divergence from known *tmoA* genes was likely explained by the mild hybridization conditions that were chosen, although mild stringency conditions are recommended in the literature to fish for novel gene orthologous (Jacquiod et al. 2014; Parsley et al. 2011). Moreover the frequent lack of *tmo*-related genes in the genomic fragments flanking the hybridized sequences made it unclear to what extent these recovered genes contribute to toluene monooxygenase activity within this sediment microbiome. However, we found in all hybridized sequences a signature characteristic of the interaction of TmoA with the other subunits of the T4-MO complex, which is necessary for its toluene hydroxylation activity. Acheson et al. (2014) have recently depicted the molecular recognition of overlapping binding sites in the T4-MO enzyme. Based on their description, we found out that all hybridizing sequences coded for proteins that had a 2Fe-2S ferredoxin-type iron-sulfur binding region signature. This motif allows the interaction of TmoA with a TmoC subunit which harbors the [2Fe-2S] cluster. Therefore, these detected hybridizing genes may encode functional toluene monooxygenase α-subunits, but future studies are necessary to identify the function of these novel proteins in comparison to previously characterized TmoAs. For example, it is conceivable to clone hybridizing ORFs into the BAC vector and check whether they encode a functional monooxygenase α-subunit that could complement toluene degradation in the knocked-out host mCH34. Finally, the prevalence of *Proteobacteria* sequences surrounding the hybridizing ORFs (see Supplementary Material, Materials and Methods S3) is in line with numerous reports over the last decades

demonstrating the role of (cultivated) *Proteobacteria* in toluene degradation. More intriguing is the prevalence, among *Proteobacteria* sequences, of *Acinetobacter* sequences, which can be related to the abundance of this genus at the Flingern site (14.4 % of 16S rRNA genes in the analyzed sediment sample, data not shown). Sequences from inserts in pFDuss5 and pFDuss116 were closely related to *A. lwoffii*, in which a toluene degradation pathway has not been studied experimentally. It was only proved that *A. lwoffii* K24 can use toluene degradation intermediates such as *p*-hydroxybenzoate, salicylate, and benzoate as sole C sources via protocatechuate 3,4-dioxygenase, catechol 1,2-dioxygenase, and benzoate 1,2-dioxygenase (Kahng et al. 2002; Yoon et al. 2007). Last but not least, our results support the existence of multiple toluene degradation pathways in some environmental bacteria. This was previously reported in a strain assigned to the *Burkholderia* genus by Johnson and Olsen (1997), *Burkholderia* JS150, where three toluene degradation systems had been simultaneously detected (toluene 2-monooxygenase, T4-MO, and dioxygenase).

The functional screening of the mCH3 BAC library gave very few positive hits (three) compared to the genetic screening. However, it is notable that it allowed the recovery of less-abundant taxa encoding toluene degradative capacity. Potential *tmo* clusters were found with putative origin from different genera, e.g., *Frankia*, *Bradyrhizobium*, *Methylibium*, and *Ralstonia*. Such a patchwork of phylogenetically distant genes was previously found in the metagenome of polluted and unpolluted environments (Jacquiod et al. 2014; Sierra-Garcia et al. 2014). In addition, the order of predicted genes in one of the clones (pBmCH34T-3) was different from what was reported in the literature. This illustrates the importance of interspecific recombination within bacteria to gain new metabolic or resistance traits for better adaptation to the environment (Ochman et al. 2000). Although toluene degradation genes have been mostly reported to be present on bacterial chromosomes, some have been found on IncP group plasmids. Actually, the plasmid pTOL of *Pseudomonas* is the only example of a well-studied IncP9 plasmid carrying a toluene catabolic pathway (Sevastsyanovich et al. 2008). In this study, all three biodegradative clones contained sequences identical to IncP1 plasmids. To the best of our knowledge, this is the first time that toluene catabolic pathways were found on IncP1 plasmids, which are therefore most likely important factors in horizontal transfer of toluene resistance genes between microbial community members. Their role in conferring pollutant degradation properties has been suggested in studies dealing with pesticide-polluted environments (Dealtry et al. 2014; Dunon et al. 2013). Another notable result of the functional screening is that it revealed the co-existence, in the cloned fragments, of genes similar to *tmo* that originated from plant-associated bacterial species such as *Frankia* and *Bradyrhizobium* genera. This suggests the importance of genetic recombination and horizontal gene transfer between plant-associated microbes and those living in the sediment. *Bradyrhizobium* species are N_2-fixing symbiotic microbes associated with plants that have been used in phytoremediation of aromatic hydrocarbons, such as *Aeschynomene indica* (Lee et al. 2008). The bacteria-plant interaction in the studied site could play an important role in hydrocarbon detoxification, and some plant organic exudates with structural similarity to toluene may provide a C source that selects for this catabolic capability among plant-associated bacteria.

In addition of *C. metallidurans* CH34, the acidobacterial strain *E. aggregans* DSM19364 was used for the first time as host for a BAC library. *Acidobacteria* are ubiquitous in soils and sediments, underrepresented in strain collections, and only a few studies have explored their genetic versatility (Anderson et al. 2012; Garcia Costas et al. 2012; Kielak et al. 2010; Rawat et al. 2012; Ward et al. 2009). In a recent study, we have demonstrated the transfer and maintenance of broad host range plasmids to *E. aggregans* DSM19364 and thus uncovered the potential of *Acidobacteria* in the development of new cloning and expression systems (Bouhajja et al. 2016). In the present work, we recovered one clone out of 845,000 screened BAC clones that produced a yellow color in the presence of toluene. In the catabolism of aromatic compounds, the generation of a yellow product is a characteristic of *meta*-cleavage of catechol ring by catechol dioxygenases to generate a semialdehyde (Suenaga 2012; Suenaga et al. 2007). Sequencing of the clone showed that genes of predicted *Acidobacteria* taxon origin were involved in aerobic conversion of catechol, a toluene degradation intermediate, but genes involved in anaerobic toluene degradation were also found, which suggests the co-existence of both systems in this facultative anaerobic *Acidobacterium*. A previous study at the Flingern site revealed that *tmoA* genes were more abundant in highly reduced zones (plume core) than in the upper fringe (aerobic zone) (Larentis et al. 2013). This observation suggested that toluene degraders are not exclusively aerobic but could be facultative anaerobic bacteria that are spread everywhere in the plume independently of the depth stratification of oxygen. In light of our results, an Acidobacteria population could be involved in toluene degradation at the Flingern site. Degradation of aromatic hydrocarbons by cultivated Acidobacteria strains has never been investigated in the laboratory, but their presence has been confirmed by culture-independent techniques in polluted environments such as petroleum- and metal-polluted sediments (Allen et al. 2007; Vishnivetskaya et al. 2011). In contrast, other studies have reported their low tolerance to hydrocarbon contamination and their disappearance from polluted sites (George et al. 2009; Saul et al. 2005). Another explanation to the presence of

Acidobacteria genes in the clone cleaving catechol is that these microorganisms are involved in the degradation of other aromatic compounds than toluene, with catechol as a key central intermediate. Such compounds include phenols, flavonoids, and polymer lignin produced by plants (Fuchs et al. 2011).

In summary, the use of both function- and sequence-based approaches to study the encoded toluene degradative capacity of a sediment microbiome resulted in distinct results. The functional screening revealed new aspects of toluene processing in the studied sample. First, IncP1 plasmids may play a role in toluene degradation and horizontal transfer of this function within a sediment microbiome. Second is the detection of an Acidobacterium that could be involved in toluene degradation. Finally, *Acinetobacter* spp. are most likely among the key toluene degraders in the Flingern sediment as they were identified in multiple fosmid clone.

Acknowledgments This work was supported by Marie Curie Initial Training Network (ITN) 'GOODWATER' (Grant Number 212683), funded by the European Union under the 7th Framework Programme (FP7). Isabelle F. George benefited from a grant of "Chargée de Recherches" of the Fund for Scientific Research of Belgium (FRS-FNRS). We are thankful to Dr. Tillmann Lueders and Dr. Michael Larentis (both at Institute of Groundwater Ecology, Helmholtz Zentrum München, Germany), for providing sediment samples and for their help to construct the *tmoA* probe.

Compliance with ethical standards

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Funding This study was funded by the European Union under the 7th Framework Programme (FP7), Marie Curie Initial Training Network (ITN) 'GOODWATER' (Grant Number 212683).

Conflict of interest The authors declare that they have no conflict of interest.

References

Acheson JF, Bailey LJ, Elsen NL, Fox BG (2014) Structural basis for biomolecular recognition in overlapping binding sites in a diiron enzyme system. Nat Commun 5. doi:10.1038/ncomms6009

Allen JP, Atekwana EA, Atekwana EA, Duris JW, Werkema DD, Rossbach S (2007) The microbial community structure in petroleum-contaminated sediments corresponds to geophysical signatures. App Environ Microbiol 73(9):2860–2870. doi:10.1128/aem.01752-06

Anderson I, Held B, Lapidus A, Nolan M, Lucas S, Tice H, Del Rio TG, Cheng JF, Han C, Tapia R, Goodwin LA, Pitluck S, Liolios K, Mavromatis K, Pagani I, Ivanova N, Mikhailova N, Pati A, Chen A, Palaniappan K, Land M, Brambilla EM, Rohde M, Spring S, Goker M, Detter JC, Woyke T, Bristow J, Eisen JA, Markowitz V, Hugenholtz P, Klenk HP, Kyrpides NC (2012) Genome sequence of the homoacetogenic bacterium *Holophaga foetida* type strain (TMBS4(T)). Stand Genomic Sci 6(2):174–184. doi:10.4056/sigs.2746047

Angov E (2011) Codon usage: nature's roadmap to expression and folding of proteins. Biotechnol J 6(6):650–659. doi:10.1002/biot.201000332

Anneser B, Pilloni G, Bayer A, Lueders T, Griebler C, Einsiedl F, Richters L (2010) High resolution analysis of contaminated aquifer sediments and groundwater—what can be learned in terms of natural attenuation? Geomicrobiol J 27(2):130–142. doi:10.1080/01490450903456723

Bosch R, García-Valdés E, Moore ERB (1999) Genetic characterization and evolutionary implications of a chromosomally encoded naphthalene-degradation upper pathway from *Pseudomonas stutzeri* AN10. Gene 236(1):149–157. doi:10.1016/s0378-1119(99)00241-3

Bouhajja E, Efthymiopoulos T, George IF, Moreels D, Van Houdt R, Mergeay M, Agathos SN (2016) Conjugative transfer of broad host range plasmids to an acidobacterial strain, *Edaphobacter aggregans*. J Biotechnol 221:107–113. doi:10.1016/j.jbiotec.2016.01.025

Brennerova MV, Josefiova J, Brenner V, Pieper DH, Junca H (2009) Metagenomics reveals diversity and abundance of meta-cleavage pathways in microbial communities from soil highly contaminated with jet fuel under air-sparging bioremediation. Environ Microbiol 11(9):2216–2227. doi:10.1111/j.1462-2920.2009.01943.x

Carmona M, Zamarro MT, Blazquez B, Durante-Rodriguez G, Juarez JF, Valderrama JA, Barragan MJ, Garcia JL, Diaz E (2009) Anaerobic catabolism of aromatic compounds: a genetic and genomic view. Microbiol Mol Biol Rev 73(1):71–133. doi:10.1128/mmbr.00021-08

Craig JW, Chang FY, Kim JH, Obiajulu SC, Brady SF (2010) Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. Appl Environ Microbiol 76(5):1633–1641. doi:10.1128/aem.02169-09

Dealtry S, Ding GC, Weichelt V, Dunon V, Schluter A, Martini MC, Del Papa MF, Lagares A, Amos GC, Wellington EM, Gaze WH, Sipkema D, Sjoling S, Springael D, Heuer H, van Elsas JD, Thomas C, Smalla K (2014) Cultivation-independent screening revealed hot spots of IncP-1, IncP-7 and IncP-9 plasmid occurrence in different environmental habitats. PLoS One 9(2):e89922. doi:10.1371/journal.pone.0089922

Dunon V, Sniegowski K, Bers K, Lavigne R, Smalla K, Springael D (2013) High prevalence of IncP-1 plasmids and IS1071 insertion sequences in on-farm biopurification systems and other pesticide-polluted environments. FEMS Microbiol Ecol 86(3):415–431. doi:10.1111/1574-6941.12173

Ekkers DM, Cretoiu MS, Kielak AM, Elsas JD (2012) The great screen anomaly-a new frontier in product discovery through functional metagenomics. Appl Microbiol Biotechnol 93(3):1005–1020. doi:10.1007/s00253-011-3804-3

Entsch B, van Berkel WJ (1995) Structure and mechanism of para-hydroxybenzoate hydroxylase. FASEB J 9(7):476–483

Ermolaeva MD (2001) Synonymous codon usage in bacteria. Curr Issues Mol Biol 3(4):91–97

Fuchs G, Boll M, Heider J (2011) Microbial degradation of aromatic compounds - from one strategy to four. Nat Rev Microbiol 9(11): 803–816. doi:10.1038/nrmicro2652

Gao J, Ellis LB, Wackett LP (2010) The University of Minnesota Biocatalysis/Biodegradation Database: improving public access. Nucleic Acids Res 38(Database issue):D488–D491. doi:10.1093/nar/gkp771

Garcia Costas AM, Liu Z, Tomsho LP, Schuster SC, Ward DM, Bryant DA (2012) Complete genome of *Candidatus Chloracidobacterium thermophilum*, a chlorophyll-based photoheterotroph belonging to the phylum *Acidobacteria*. Environ Microbiol 14(1):177–190. doi:10.1111/j.1462-2920.2011.02592.x

George IF, Liles MR, Hartmann M, Ludwig W, Goodman RM, Agathos SN (2009) Changes in soil *Acidobacteria* communities after 2,4,6-

trinitrotoluene contamination. FEMS Microbiol Lett 296(2):159–166. doi:10.1111/j.1574-6968.2009.01632.x

Gillespie DE, Brady SF, Bettermann AD, Cianciotto NP, Liles MR, Rondon MR, Clardy J, Goodman RM, Handelsman J (2002) Isolation of antibiotics turbomycin a and B from a metagenomic library of soil microbial DNA. Appl Environ Microbiol 68(9): 4301–4306

Gross SA, Avens HJ, Banducci AM, Sahmel J, Panko JM, Tvermoes BE (2013) Analysis of BTEX groundwater concentrations from surface spills associated with hydraulic fracturing operations. J Air Waste Manag Assoc 63(4):424–432

Heider J, Spormann AM, Beller HR, Widdel F (1998) Anaerobic bacterial metabolism of hydrocarbons. FEMS Microbiol Rev 22(5):459–473. doi:10.1111/j.1574-6976.1998.tb00381.x

Hendrickx B, Dejonghe W, Boënne W, Brennerova M, Cernik M, Lederer T, Bucheli-Witschel M, Bastiaens L, Verstraete W, Top EM, Diels L, Springael D (2005) Dynamics of an oligotrophic bacterial aquifer community during contact with a groundwater plume contaminated with benzene, toluene, ethylbenzene, and xylenes: an in situ mesocosm study. App Environ Microbiol 71(7):3815–3825. doi:10.1128/aem.71.7.3815-3825.2005

Hendrickx B, Dejonghe W, Faber F, Boënne W, Bastiaens L, Verstraete W, Top EM, Springael D (2006) PCR-DGGE method to assess the diversity of BTEX mono-oxygenase genes at contaminated sites. FEMS Microbiol Ecol 55(2):262–273. doi:10.1111/j.1574-6941.2005.00018.x

Jackson C, Harper J, Willoughby D, Roden E, Churchill P (1997) A simple, efficient method for the separation of humic substances and DNA from environmental samples. Appl Environ Microbiol 63(12):4993–4995

Jacquiod S, Demaneche S, Franqueville L, Ausec L, Xu Z, Delmont TO, Dunon V, Cagnon C, Mandic-Mulec I, Vogel TM, Simonet P (2014) Characterization of new bacterial catabolic genes and mobile genetic elements by high throughput genetic screening of a soil metagenomic library. J Biotechnol 20(190):18–29. doi:10.1016/j.jbiotec.2014.03.036

Janssen PJ, Van Houdt R, Moors H, Monsieurs P, Morin N, Michaux A, Benotmane MA, Leys N, Vallaeys T, Lapidus A, Monchy S, Medigue C, Taghavi S, McCorkle S, Dunn J, van der Lelie D, Mergeay M (2010) The complete genome sequence of *Cupriavidus metallidurans* strain CH34, a master survivalist in harsh and anthropogenic environments. PLoS One 5(5):e10433. doi:10.1371/journal.pone.0010433

Jobelius C, Ruth B, Griebler C, Meckenstock RU, Hollender J, Reineke A, Frimmel FH, Zwiener C (2011) Metabolites indicate hot spots of biodegradation and biogeochemical gradients in a high-resolution monitoring well. Environ Sci Technol 45(2): 474–481. doi:10.1021/es1030867

Johnson GR, Olsen RH (1997) Multiple pathways for toluene degradation in *Burkholderia* sp. strain JS150. Appl Environ Microbiol 63(10):4047–4052

Kahng HY, Cho K, Song SY, Kim SJ, Leem SH, Kim SI (2002) Enhanced detection and characterization of protocatechuate 3,4-dioxygenase in *Acinetobacter lwoffii* K24 by proteomics using a column separation. Biochem Biophys Res Commun 295(4):903–909

Kakirde KS, Wild J, Godiska R, Mead DA, Wiggins AG, Goodman RM, Szybalski W, Liles MR (2011) Gram negative shuttle BAC vector for heterologous expression of metagenomic libraries. Gene 475(2): 57–62. doi:10.1016/j.gene.2010.11.004

Kielak AM, van Veen JA, Kowalchuk GA (2010) Comparative analysis of acidobacterial genomic fragments from terrestrial and aquatic metagenomic libraries, with emphasis on *Acidobacteria* subdivision 6. Appl Environ Microbiol 76(20): 6769–6777. doi:10.1128/aem.00343-10

Kimura N, Sakai K, Nakamura K (2010) Isolation and characterization of a 4-nitrotoluene-oxidizing enzyme from activated sludge by a metagenomic approach. Microbes Environ 25(2):133–139. doi:10.1264/jsme2.ME10110

Larentis M, Hoermann K, Lueders T (2013) Fine-scale degrader community profiling over an aerobic/anaerobic redox gradient in a toluene-contaminated aquifer. Environ Microbiol Rep 5(2):225–234. doi:10.1111/1758-2229.12004

Lau PC, Bergeron H, Labbe D, Wang Y, Brousseau R, Gibson DT (1994) Sequence and expression of the *todGIH* genes involved in the last three steps of toluene degradation by *Pseudomonas putida F1*. Gene 146(1):7–13

Lee EH, Ryu HW, Cho KS (2011) Effect of switching gas inlet position on the performance of a polyurethane biofilter under transient loading for the removal of benzene, toluene and xylene mixtures. J Environ Sci Health, Part A: Tox Hazard Subst Environ Eng 46(13):1570–1578. doi:10.1080/10934529.2011.609444

Lee SH, Lee WS, Lee CH, Kim JG (2008) Degradation of phenanthrene and pyrene in rhizosphere of grasses and legumes. J Hazard Mater 153(1–2):892–898. doi:10.1016/j.jhazmat.2007.09.041

Leis B, Angelov A, Liebl W (2013) Screening and expression of genes from metagenomes. Adv Appl Microbiol 83:1–68. doi:10.1016/b978-0-12-407678-5.00001-5

Liles MR, Williamson LL, Rodbumrer J, Torsvik V, Goodman RM, Handelsman J (2008) Recovery, purification and cloning of high molecular weight DNA from soil microorganisms. Appl Environ Microbiol 74(10):3302–3305. doi:10.1128/aem.02630-07

Meckenstock RU, Elsner M, Griebler C, Lueders T, Stumpp C, Aamand J, Agathos SN, Albrechtsen HJ, Bastiaens L, Bjerg PL, Boon N, Dejonghe W, Huang WE, Schmidt SI, Smolders E, Sorensen SR, Springael D, van Breukelen BM (2015) Biodegradation: updating the concepts of control for microbial cleanup in contaminated aquifers. Environ Sci Technol 49(12): 7073–7081. doi:10.1021/acs.est.5b00715

Mergeay M, Houba C, Gerits J (1978) Extrachromosomal inheritance controlling resistance to cadmium, cobalt, copper and zinc ions: evidence from curing in a *Pseudomonas*. Arch Int Physiol Biochim 86(2):440–442

Mettel C, Kim Y, Shrestha PM, Liesack W (2010) Extraction of mRNA from soil. Appl Environ Microbiol 76(17):5995–6000

Notomista E, Lahm A, Di Donato A, Tramontano A (2003) Evolution of bacterial and archaeal multicomponent monooxygenases. J Mol Evol 56(4):435–445. doi:10.1007/s00239-002-2414-1

Ochman H, Lawrence JG, Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. Nature 405(6784):299–304. doi:10.1038/35012500

Okunishi S, Morita Y, Higuchi T, Maeda H, Nishi K (2012) Transformation of microflora during degradation of gaseous toluene in a biofilter detected using PCR-DGGE. J Air Waste Manag Assoc 62(7):748–757

Olsen RH, Kukor JJ, Kaphammer B (1994) A novel toluene-3-monooxygenase pathway cloned from *Pseudomonas pickettii* PKO1. J Bacteriol 176(12):3749–3756

Pansegrau W, Lanka E, Barth PT, Figurski DH, Guiney DG, Haas D, Helinski DR, Schwab H, Stanisich VA, Thomas CM (1994) Complete nucleotide sequence of Birmingham *IncP* alpha plasmids. Compilation and comparative analysis. J Mol Biol 239(5):623–663. doi:10.1006/jmbi.1994.1404

Parsley LC, Linneman J, Goode AM, Becklund K, George I, Goodman RM, Lopanik NB, Liles MR (2011) Polyketide synthase pathways identified from a metagenomic library are derived from soil *Acidobacteria*. FEMS Microbiol Ecol 78(1): 176–187. doi:10.1111/j.1574-6941.2011.01122.x

Patil KR, Roune L, McHardy AC (2012) The PhyloPythiaS web server for taxonomic assignment of metagenome sequences. PLoS One 7(6):e38581. doi:10.1371/journal.pone.0038581

Rakoczy J, Remy B, Vogt C, Richnow HH (2011) A bench-scale constructed wetland as a model to characterize benzene biodegradation

processes in freshwater wetlands. Environ ScTechnol 45(23): 10036–10044. doi:10.1021/es2026196

Rawat SR, Mannisto MK, Starovoytov V, Goodwin L, Nolan M, Hauser L, Land M, Davenport KW, Woyke T, Haggblom MM (2012) Complete genome sequence of *Terriglobus saanensis* type strain SP1PR4(T), an *Acidobacteria* from tundra soil. Stand Genomic Sci 7(1):59–69. doi: 10.4056/sigs.3036810

Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA, Minor C, Tiong CL, Gilman M, Osburne MS, Clardy J, Handelsman J, Goodman RM (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl Environ Microbiol 66(6):2541–2547

Saul DJ, Aislabie JM, Brown CE, Harris L, Foght JM (2005) Hydrocarbon contamination changes the bacterial diversity of soil from around Scott Base, Antarctica. FEMS Microbiol Ecol 53(1): 141–155. doi:10.1016/j.femsec.2004.11.007

Scow KM, Hicks KA (2005) Natural attenuation and enhanced bioremediation of organic contaminants in groundwater. Curr Opin Biotechnol 16(3):246–253. doi:10.1016/j.copbio.2005.03.009

Sevastsyanovich YR, Krasowiak R, Bingle LE, Haines AS, Sokolov SL, Kosheleva IA, Leuchuk AA, Titok MA, Smalla K, Thomas CM (2008) Diversity of IncP-9 plasmids of *Pseudomonas*. Microbiology 154(Pt 10):2929–2941. doi: 10.1099/mic.0.2008/017939-0

Sharma P, Capalash N, Kaur J (2007) An improved method for single step purification of metagenomic DNA. Mol Biotechnol 36(1):61–63. doi:10.1007/s12033-007-0015-3

Sierra-Garcia IN, Correa Alvarez J, de Vasconcellos SP, Pereira de Souza A, dos Santos Neto EV, de Oliveira VM (2014) New hydrocarbon degradation pathways in the microbial metagenome from Brazilian petroleum reservoirs. PLoS One 9(2):e90087. doi:10.1371/journal.pone.0090087

Singleton DR, Hu J, Aitken MD (2012) Heterologous expression of polycyclic aromatic hydrocarbon ring-hydroxylating dioxygenase genes from a novel pyrene-degrading betaproteobacterium. Appl Environ Microbiol 78(10):3552–3559. doi:10.1128/aem.00173-12

Suenaga H (2012) Targeted metagenomics: a high-resolution metagenomics approach for specific gene clusters in complex microbial communities. Environ Microbiol 14(1):13–22. doi:10.1111/j.1462-2920.2011.02438.x

Suenaga H, Ohnuki T, Miyazaki K (2007) Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. Environ Microbiol 9(9):2289–2297

Szczepanowski R, Eikmeyer F, Harfmann J, Blom J, Rogers LM, Top EM, Schluter A (2011) Sequencing and comparative analysis of IncP-1alpha antibiotic resistance plasmids reveal a highly conserved backbone and differences within accessory regions. J Biotechnol 155(1):95–103. doi:10.1016/j.jbiotec.2010.11.018

Taghavi S, van der Lelie D, Mergeay M (1994) Electroporation of *Alcaligenes eutrophus* with (mega) plasmids and genomic DNA fragments. Appl Environ Microbiol 60(10):3585–3591

Trevors JT, Starodub ME (1990) Electroporation of pKK1 silver-resistance plasmid from *Pseudomonas stutzeri* AG259 into *Pseudomonas putida* CYM318, vol 21. Springer, New York

van Elsas JD, Costa R, Jansson J, Sjöling S, Bailey M, Nalin R, Vogel TM, van Overbeek L (2008) The metagenomics of disease-suppressive soils—experiences from the METACONTROL project. Trends Biotechnol 26(11):591–601. doi:10.1016/j.tibtech.2008.07.004

Vishnivetskaya TA, Mosher JJ, Palumbo AV, Yang ZK, Podar M, Brown SD, Brooks SC, Gu B, Southworth GR, Drake MM, Brandt CC, Elias DA (2011) Mercury and other heavy metals influence bacterial community structure in contaminated Tennessee streams. Appl Environ Microbiol 77(1):302–311. doi:10.1128/aem.01715-10

Wang Y, Rawlings M, Gibson DT, Labbe D, Bergeron H, Brousseau R, Lau PC (1995) Identification of a membrane protein and a truncated LysR-type regulator associated with the toluene degradation pathway in *Pseudomonas putida* F1. Mol Gen Genet 246(5):570–579

Ward NL, Challacombe JF, Janssen PH, Henrissat B, Coutinho PM, Wu M, Xie G, Haft DH, Sait M, Badger J, Barabote RD, Bradley B, Brettin TS, Brinkac LM, Bruce D, Creasy T, Daugherty SC, Davidsen TM, DeBoy RT, Detter JC, Dodson RJ, Durkin AS, Ganapathy A, Gwinn-Giglio M, Han CS, Khouri H, Kiss H, Kothari SP, Madupu R, Nelson KE, Nelson WC, Paulsen I, Penn K, Ren Q, Rosovitz MJ, Selengut JD, Shrivastava S, Sullivan SA, Tapia R, Thompson LS, Watkins KL, Yang Q, Yu C, Zafar N, Zhou L, Kuske CR (2009) Three genomes from the phylum *Acidobacteria* provide insight into the lifestyles of these microorganisms in soils. Appl Environ Microbiol 75(7):2046–2056. doi:10.1128/aem.02294-08

Wexler M, Johnston AW (2010) Wide host-range cloning for functional metagenomics. Methods Mol Biol 668:77–96. doi:10.1007/978-1-60761-823-2_5

Whited GM, Gibson DT (1991) Separation and partial characterization of the enzymes of the toluene-4-monooxygenase catabolic pathway in *Pseudomonas mendocina* KR1. J Bacteriol 173(9):3017–3020

Winderl C, Anneser B, Griebler C, Meckenstock RU, Lueders T (2008) Depth-resolved quantification of anaerobic toluene degraders and aquifer microbial community patterns in distinct redox zones of a tar oil contaminant plume. Appl Environ Microbiol 74(3):792–801. doi:10.1128/AEM.01951-07

Yoon YH, Yun SH, Park SH, Seol SY, Leem SH, Kim SI (2007) Characterization of a new catechol branch of the beta-ketoadipate pathway induced for benzoate degradation in *Acinetobacter lwoffii* K24. Biochem Biophys Res Commun 360(3):513–519. doi:10.1016/j.bbrc.2007.05.132

Zhou JZ, Bruns MA, Tiedje JM (1996) DNA recovery from soils of diverse composition. Appl Environ Microbiol 62(2):316–322