Biotechnol Lett (2015) 37:837–847
DOI 10.1007/s10529-014-1736-6

ORIGINAL RESEARCH PAPER

Phylogenetic characterization and novelty of organic sulphur metabolizing genes of *Rhodococcus* spp. (Eu-32)

Nasrin Akhtar · Muhammad A. Ghauri · Munir A. Anwar · Shaun Heaphy

Received: 22 October 2014 / Accepted: 19 November 2014 / Published online: 10 December 2014
© Springer Science+Business Media Dordrecht 2014

Abstract *Rhodococcus* spp. (Eu-32) has the unique ability to metabolize organic sulphur containing compounds like dibenzothiophene through an extended sulphur specific pathway (Akhtar et al., in FEMS Microbiol Lett 301:95–102, 2009). Efforts were made to isolate and characterize the presumed desulphurizing genes (*dszABC)* involved in the sulphur specific pathway of isolate Eu-32 by employing standard and degenerate polymerase chain reaction primers. The partial *dszA* gene sequence of isolate Eu-32 showed 92 % sequence identity with a putative FMNH-2 dependent monooxygenase of *Rhodococcus erythropolis* PR4. The *dszC* gene sequence showed 99 % homology with the dibenzothiophene monooxygenase desulphurizing enzyme of another *Rhodococcus* species. The *dszB* gene was not unambiguously identified. A phylogenetic analysis by maximum likelihood method of the 16S rRNA gene and deduced DszA and C amino acid sequences suggest that horizontal gene transfer events might have taken place during the evolution of desulphurizing genes of *Rhodococcus* spp. (Eu-32).

Keywords Biodesulphurization · *dszABC* genes · Horizontal gene transfer (HGT) · PCR amplification · Phylogenetic analysis · *Rhodococcus* spp. (Eu-32)

Electronic supplementary material The online version of this article (doi:10.1007/s10529-014-1736-6) contains supplementary material, which is available to authorized users.

N. Akhtar · M. A. Ghauri (✉) · M. A. Anwar
Industrial Biotechnology Division, National Institute for Biotechnology and Genetic Engineering, Jhang Road, P. O. Box 577, Faisalabad, Pakistan
e-mail: ghauri1961@gmail.com; maghauri@nibge.org

S. Heaphy
Department of Infection, Immunity and Inflammation, University of Leicester, Leicester LE1 9HN, UK
e-mail: sh1@le.ac.uk

Introduction

Combustion of sulphur-containing compounds in fossil fuels emits sulphur oxides causing adverse effects on health, environment and economy (Diaz and Garcia 2010). Sulphur in fossil fuels can be removed through conventional methods however; refractory organic sulphur is very difficult to remove. The conventional methods are very costly and energy intensive. There are other methods, such as biodesulphurization (BDS), which have shown good potential for removing refractory sulphur under milder conditions (Soleimani et al. 2007).

Dibenzothiophene (DBT) and its derivatives are generally considered model target compounds for biodesulphurization studies (Mohebali and Ball 2008) and a significant number of microorganisms have been found that remove sulphur from DBT via the sulphur specific (4S) pathway. In this pathway, DBT

is desulphurized to 2-hydroxybiphenyl (2-HBP) without degradation of the carbon–carbon bond and thus the calorific value of the fuel remains conserved (Akhtar et al. 2009; Diaz and Garcia 2010; Soleimani et al. 2007; Mohebali and Ball 2008). Though, the genes involved in microbial DBT desulphurization have been isolated from various bacterial strains, only the genes of *R. erythropolis* strain IGTS8 (*dszA*, *dszB*, *dszC*) have been extensively studied and characterized (Piddington et al. 1995). The DBT desulphurizing genes constituting operons, have also been isolated and cloned from thermophilic bacteria *Paenibacillus* sp. A11-2 (*tdsABC*), *Bacillus subtilis* WU-S2B (*bdsABC*) and *Mycobacterium phlei* WU-F1 (*tdsABC*) (Konishi et al. 1997; Ishii et al. 2000; Kirimura et al. 2001, 2004; Furuya et al. 2001).

All known bacteria with the ability to desulphurize DBT into 2-HBP through the sulphur specific pathway possess three desulphurization enzymes. In *R. erythropolis* IGTS8, the DszC enzyme catalyzes two consecutive monooxygenation reactions converting DBT to DBT sulphone, subsequently DszA as a second monooxygenase enzyme converts the DBT sulphone to hydroxyphenyl benzenesulphinate (HPBS). Finally, DszB as a desulphinase enzyme transforms HPBS to 2-HBP and sulphite (Gray et al. 1996, 2003). All three genes are clustered on a 120 kb linear plasmid of strain IGTS8 (Denome et al. 1994).

Earlier, we have reported that *Rhodococcus* spp. (Eu-32) is an organic sulphur metabolizing culture with the ability to convert DBT into 2-HBP through an extended sulphur specific (4S) pathway with biphenyl as a second end product (Akhtar et al. 2009). In this paper, we report the work done for the isolation and identification of *dszABC* genes of *Rhodococcus* spp. (Eu-32). We used a range of different degenerate and standard (non degenerate) primers for the PCR amplification of these genes, individually or jointly. The *dszA* and *dszC* genes catalysing the first 3 steps of DBT desulphurization to the HPBS were identified and phylogenetically analysed. In the last step of desulphurization, *R. erythropolis* strain IGTS8 converts HPBS to 2-HBP using the *dszB* gene. In Eu-32 this last step is further followed by a dehydroxylation step to biphenyl suggesting that either a fourth enzyme or a significantly different DszB enzyme is involved in the pathway.

Materials and methods

Bacterial culture and growth conditions

Rhodococcus spp. (Eu-32) used in this study was previously isolated from a soil sample taken from the roots of a eucalyptus tree. Isolate Eu-32 was cultured either in MG or LB medium as described earlier (Akhtar et al. 2009).

Primers used to PCR amplify the *dszABC* genes

Two types of primers (standard and degenerate) were used in order to amplify the *dszABC* gene sequences of *Rhodococcus* spp. (Eu-32) by polymerase chain reaction (PCR). Standard primers were based on the 5′ and 3′ termini of *dsz* genes of *Rhodococcus erythropolis* IGTS8. The nucleotide sequences of the degenerate PCR primers were designed using conserved amino acid sequence regions of published DBT-desulphurizing enzymes from *R. erythropolis* IGTS8 (DszA AAA99482.1, DszB AAA99483.1 and DszC AAA99484.1), *Paenibacillus* sp. A11-2 (TdsA BAA94831.1, TdsB BAA 94832.1 and TdsC BAA94833.1), *Bacillus subtilis* WU-S2B (BdsA BAC20180.1, BdsB BAC20181 and BdsC BAC20182.1) and *Gordonia alkanivorans* strain 1B (DszA AAT78716.1, DszB AAT78717.1 and DszC AAT78718.1). Oligonucleotides for PCR amplification of *dszABC* genes were purchased from Integrated DNA Technologies, Inc., UK. The sequences of the standard and degenerate primers used in this study are shown in Table 1.

The DNA PCR and molecular biology techniques

Rhodococcus spp. (Eu-32) was grown in LB medium overnight. Total genomic DNA from *Rhodococcus* spp. was extracted using the method described by Ausubel et al. (1995). The extracted DNA was stored at −20 °C until required. PCR amplifications were performed with Bioline Biotaq polymerase (UK). The dNTPs mixture (10 mM) was purchased from Fermentas (ThermoFisher, Scientific, USA). Amplification was achieved with 1 cycle of 5 min of denaturation at 95 °C, 30 cycles of 1 min of denaturation at 95 °C, 1 min of annealing at 55 °C and 2 min of extension at 72 °C, plus a final additional extension

Table 1 List of primers used for the PCR amplification of *dszABC* genes of *Rhodococcus* spp. (Eu-32)

Primer name	Position in *dsz* genes of *R. erythropolis* IGTS8	Sequence (5′–3′)
Standard primers		
dszA Forward	1–22	ATGACTCAACAACGACAAATGC
dszA Reverse	1,342–1,362	TCATGAAGGTTGTCCTTGCAG
dszB Forward	1–20	ATGACAAGCCGCGTCGACCC
dszB Reverse	1,079–1,098	CTATCGGTGGCGATTGAGGC
dszC Forward	1–22	ATGACACTGTCACCTGAAAAGC
dszC Reverse	1,235–1,254	TCAGGAGGTGAAGCCGGGAA
AEu-32F	Based on *Rhodococcus* spp. (Eu-32) *dszA* sequence	CGGGGAGGGTTTCTCGACCACGA
Eu2AF	Based on *Rhodococcus* spp. (Eu-32) *dszA* sequence	CGACGCCACCCTCCCGCGGAGTG
Degenerate primers		
HLAGFF-AF	8–13	CATYTDGCCGGKTTYTT
AGNVTH-AF	15–20	GCBGGYAAYGTSACYCAT
ARTLERG-AF	44–50	GCSCGKACHCTSGAGCGSGGC
RYDRADEFL-AR	159–167	AAYTCRTCRGCVCGGTCRTABCG
PVILQAG-AR	223–229	CCSGCCTGMADGATSACMGG
GLSTL SSH-AR	309–316	TGRCTKGAYAGSGTMGASARWCC
EFVDQVVP-AF	413–419	GARTTYGTCGACMARGTGGTKCCK
GNASSENN-CF	128–135	GGRAAYGCMTCCAGCGARAACAA
GFDRFWR-CR	377–383	GTYGCGCCARAASCGGTCRAADCC
AYTRFGG-BF	67–73	GCYTACACCCGHTWTGGBGG
ASVWTVS-BR	252–258	GCTGACCGTCCAVACRCTSGC
VQRLVD-BF	269–274	GTBCAACGMYTSGTCGAC
WAAPEFL-BF	351–357	TGGGCGGCHCCSSAATTYCT
WAAPEFL-BR	351–357	AGRAATTSSGGDGCCGCCCA

Sequences are according to International Union of Biochemistry (IUB) group codes (N = any base, M = A or C, R = A or G, W = A or T, S = C or G, Y = C or T, K = G or T, B = not A, D = not C, H = not G, V = not T)

at 72 °C for 15 min with a Geneflow, thermal cycler, UK. The PCR amplified target bands were gel extracted by using NucleoSpin Extract II Gel Extraction Kit (Germany) as per manufacturer recommendations. *Escherichia coli* TOP10 one shot competent cells (Invitrogen, USA) were used as host strains for general cloning. Transformation of DNA was performed according to the supplier's recommendation. Published recombinant techniques (Sambrook et al. 1989) were used unless otherwise noted. DNA fragments amplified by PCR were ligated to the cloning vector pGEMT-easy (Promega). Plasmid DNA was isolated using PureLink^{TM} HQ Mini Plasmid Purification Kit (Invitrogen, USA). Cloned fragments were subsequently sequenced through Source Bioscience, UK using the M13 or T7 forward and reverse primers.

Blast and phylogenetic analysis of the genes

The determination of DNA sequence identity was performed using the basic local alignment search tool (BLAST) program "tblastx" of the National Center for Biotechnology Information (NCBI) (www.ncbi.nlm.nih.gov). Phylogenetic analysis was performed by maximum likelihood (ML) method using the programme MEGA5 (Tamura et al. 2011). The Tamura–Nei model (Tamura and Nei 1993) for 16S rRNA gene and Jones–Thornton–Taylor (JTT) matrix-based model (Jones et al. 1992) for DszA and DszC amino acid sequences were used. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using either Maximum

Table 2 Bacterial species names and NCBI accession numbers of 16S rRNA gene sequences used

Bacterial species	NCBI access on numbers
Rhodococcus erythropolis strain XP	DQ074453
Rhodococcus sp. IGTS8	AF001265
Paenibacillus humicus	AB681771
Paenibacillus lautus	AJ491842
Agrobacterium tumefaciens strain CCNWSX1287	GU645017
Agrobacterium tumefaciens strain X9	JX002661
Mycobacterium sp. MCRO3	AF152559
Mycobacterium goodie Strain X7B	AF513815
Gordonia amicalis	AF101418
Gordonia alkanivorans strain HKI 0136	NR_026488
Gordonia alkalivorans	AB065369
Bacillus subtilis strain 13B	HQ335318
Bacillus subtilis	AB018487

Table 3 Bacterial species names and NCBI accession numbers of DszA and DszC sequences used

Bacterial species	NCBI accession numbers	
	DszA	DszC
Rhodococcus erythropolis IGTS8	AAA99482.1	AAA99484.1
Rhodococcus sp. XP	AAP33509.1	AAP33510.1
Rhodococcus sp. SDUZAWQ	AAV49165.1	AAV49167.1
Paenibacillus sp. A11-2	BAA94831.1	BAA94833.1
Bacillus subtilis WU-S2B	BAC20180.1	BAC20182.1
Agrobacterium tumefaciens	AAX76900.1	AAX76902.1
Mycobacterium sp. G3	BAC41357.1	BAC41359.1
Gordonia alkanivorans strain 1B	AAT78716.1	AAT78718.1
Gordonia sp.CYKS2	AAQ96169.1	AAQ96171.1
Gordonia alkanivorans	AAU14817.1	AAU14819.1
Gordonia sp. RIPI	CAJ00733.1	CAJ00735.1
Gordonia amicalis	ABK20068.1	ABK20070.1

Composite Likelihood (MCL) approach or JTT model, and then selecting the topology with superior log likelihood value. Reliability for internal branch was assessed using the bootstrapping method (100 bootstrap replicates). All positions containing missing data and gaps were eliminated. The % identity between the deduced amino acid sequences and others was determined using the Clone Manager suite version 7.0.

Nucleotide sequence accession numbers

Nucleotide sequences from partial *dszA* and *dszC* genes were deposited with GenBank of NCBI (www.ncbi.nlm.nih.gov) under accession numbers KF031127 and KF031128 respectively. The nucleotide and protein sequences used in this study for phylogenetic analysis were obtained from the NCBI and their accession numbers are given in Tables 2 and 3. The 16S rRNA gene sequence for strain Eu-32 was previously deposited under accession number DQ386111.

Results

For PCR amplification and characterization of the presumed *dszABC* genes involved in C–S bond cleavage of *Rhodococcus* spp. (Eu-32), we designed specific primers based on the 5′ and 3′ regions of *dszABC* genes of *R. erythropolis* IGTS8. These primers generated multiple bands, despite stringent PCR conditions and did not amplify homologous *dsz* genes from *Rhodococcus* spp. (Eu-32), data not shown. Consequently degenerate PCR primers were designed to amplify the putative *dszABC* genes of Eu-32. For this purpose, an alignment of the protein sequences of dissimilar desulphurization enzymes of *Rhodococcus erythropolis* IGTS8 (AAA99482), *Gordonia alkanivorans* strain 1B (AAT78716), *Bacillus subtilis* WU-S2B (BAC20180) and *paenibacillus* sp. A11-2 (BAA94831) was made, which revealed multiple conserved regions (Supplementary Figs. 1, 2, 3). Based on these conserved amino acid regions and inferred nucleotide sequences degenerate primers were designed.

Isolation of *dszA* and *dszC* genes

For isolation of *dszA* gene of *Rhodococcus* spp. (Eu-32), PCR reactions were performed with primers based on BdsA of *B. subtilis* WU-S2B, TdsA of *Paenibacillus* spp. A11-2 and the DszA of *R. erythropolis* IGTS8 and *G. alkanivorans* 1B. The DszA enzyme of *R. erythropolis* IGTS8 and *G. alkanivorans* 1B consists of 453 and 474 amino acids respectively. The BdsA of *B. subtilis* WU-S2B and TdsA of *P. bacillus* A11-2 consist of 453 and 454 amino acids respectively. The DszA/BdsA/TdsA enzyme alignment contains 15 conserved regions

(Supplementary Fig. 1). Keeping in view the G + C contents and redundant positions in the corresponding nucleotides sequences of the conserved amino acid regions, different degenerate primers were designed to isolate *dszA* gene of Eu-32. Kilbane and Robbins in 2007 reported four conserved motifs that can be diagnostic for desulphurization enzymes and allow the construction of the most useful PCR primers. These amino acid sequences were HLAGFF, ARTLERG, FDLLFLPDGLA, and RYDRADEFL that correspond to positions 8–13, 44–50, 52–62, and 159–167, respectively in the DszA protein of *R. erythropolis* IGTS8. In our study, the amino acid regions ARTLERG (5′GCSCGKACHCTSGAGCGSGGC3′), RYDRADEFL (5′AAYTCRTCRGCVCGGTCRTABCG3′) and PVILQAG (5′CCSGCCTGMADGATSACMGG3′) were fully conserved and facilitated primer design for the PCR amplification of the target regions of the *dszA* gene of Eu-32. The sequences of amplicons obtained using these degenerate primers allowed us to construct other non-degenerate primers that were specific for the *dszA* gene of *Rhodococcus* spp. (Eu-32), see Table 1. Using these primers we determined a 496 bp sequence of *dszA* gene of isolate Eu-32. tblastx analysis of this 496 bp *dszA* DNA sequence from isolate Eu-32 showed 92 % homology with a putative-$FMNH_2$ dependent monooxygenase of *R. erythropolis* PR4 DNA, complete genome sequence. Alignment of the deduced amino acid sequences of DszA of isolate Eu-32 with DszA enzymes from other desulphurizing bacteria showed that DszA has low homology in the regions conserved within desulphurizing enzymes (Fig. 1a).

For isolation of *dszC* gene of *Rhodococcus* spp. (Eu-32), the two conserved regions of DszC protein sequences (Supplementary Fig. 2) were used to design PCR primers. A 23 bp forward PCR primer with four redundant positions was constructed targeting the region GNASSENN (5′GGRAAYGCMTCCAGCGARAACAA3′) and a 24 bp primer with five redundant positions was constructed targeting the GFDRFWR (5′GTYGCGCCARAASCGGTCRAADCC3′) region of DszC. The PCR amplification of *dszC* gene carried out using degenerate primer pair GNASSENN-CF and GFDRFWR-CR, resulted in a major amplicon of approximately the expected size (771 bp) which was cloned and sequenced. tblastx analysis of the sequence showed identity with desulphurizing enzymes of several genera including *Rhodococcus*, *Nocardia*, *Gordonia*, *Bacillus*, *Mycobacterium*, *Agrobacterium* and *Paenibacillus*. Alignment of the deduced amino acid sequences of *dszC* of isolate Eu-32 and equivalent proteins, from other desulphurizing bacteria showed that DszC has high homology in the regions conserved within desulphurizing enzymes (Fig. 1b).

Attempted isolation of *dszB* gene

The DszB/BdsB/TdsB of *R. erythropolis* IGTS8, *G. alkanivorans* 1B, *B. subtilis* WU-S2B and *P. bacillus* spp. A11-2 consists of about 353–365 amino acids and contains many conserved regions (Supplementary Fig. 3). Four conserved regions were judged to be appropriate for the construction of the degenerate primers. These conserved regions are AYTRFGG, ASVWTVS, VQRLVD, and WAAPEFL that correspond to amino acid positions 67–73, 252–258, 269–274, and 351–357, respectively, of DszB of *R. erythropolis* IGTS8. The PCR reactions set up using degenerate primers for amplification of *dszB* gene of Eu-32 yielded several amplicons, those of the expected size bands were cloned and sequenced. The DNA sequences showed identity with genome sequences of several *Rhodococcus* species, but not to *dszB* genes.

Phylogenetic analysis

Phylogenetic analysis is a powerful tool for sorting and interpreting molecular data. Phylogenetic comparisons using both the desulphurizing enzymes DszA and C and the 16S rRNA gene sequence of Eu-32 was carried out with other relevant organisms from the same genera. For the 16S gene comparison we choose 13 sequences, including at least two sequences of 1,200–1,400 bp from each genera and/or species included in the DszA and C trees. The phylogenetic analysis on the basis of 16S rRNA gene sequence showed that the species, as expected, are forming clades within their respective genera (Fig. 2). The phylogenetic analysis of the DszC sequence of isolate Eu-32 and other desulphurizing bacterial monooxygenases (Fig. 3) showed that Eu-32 also forms a close clade with *Agrobacterium tumefaciens*, *Gordonia alkanivorans, Rhodococcus* sp. SDUZAWQ, *Rhodococcus* sp. XP and *R. erythropolis* IGTS8. On the other hand, phylogenetic analysis based on DszA sequences

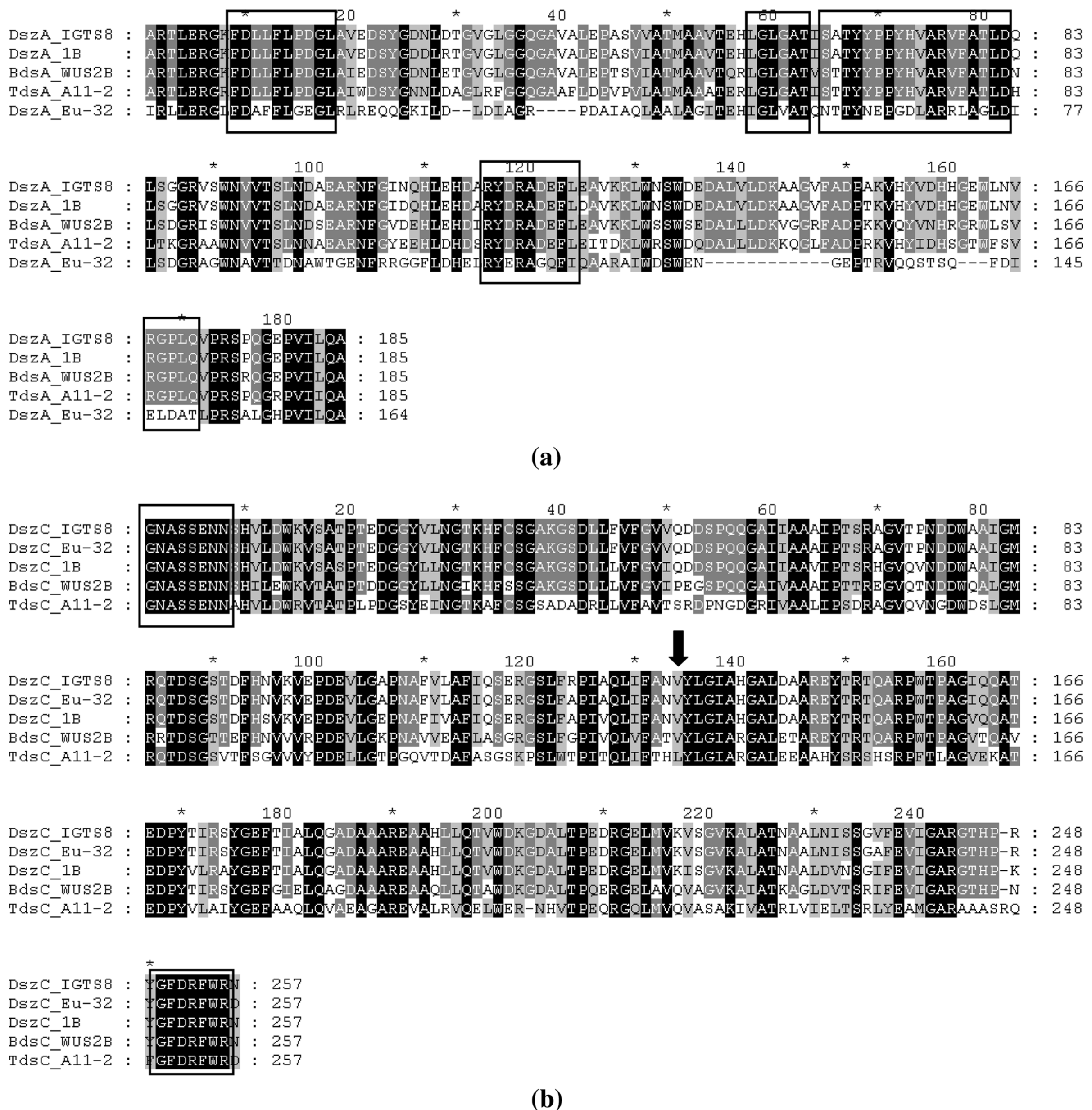

Fig. 1 **a** Multiple alignments of the partial deduced amino acid sequence of $DBTO_2$ monooxygenase from Eu-32 with its equivalents from other strains, showing that boxed regions are highly conserved among the desulphurizing bacteria but, not in isolate Eu-32. **b** Multiple alignments of the partial deduced amino acid sequence of DBTO monooxygenase from Eu-32 with its equivalents from other strains, showing that boxed regions are highly conserved among different desulphurizing bacterial species as well as in isolate Eu-32. The valine amino acid has been shown by a bold arrow sign. Alignment was carried out using ClustalX 1.81 software under the default settings (multiple alignment parameters: gap opening 10.00 and gap extension 0.20)

of the $FMNH_2$ dependent monooxygenases, Fig. 4, shows that Eu-32 is only distantly related to other *Rhodococcus* species, indeed it forms an independent clade along with *Paenibacillus* sp. A11-2.

Discussion

Generally, if two organisms are related, their DNA sequences for a particular gene will show some degree

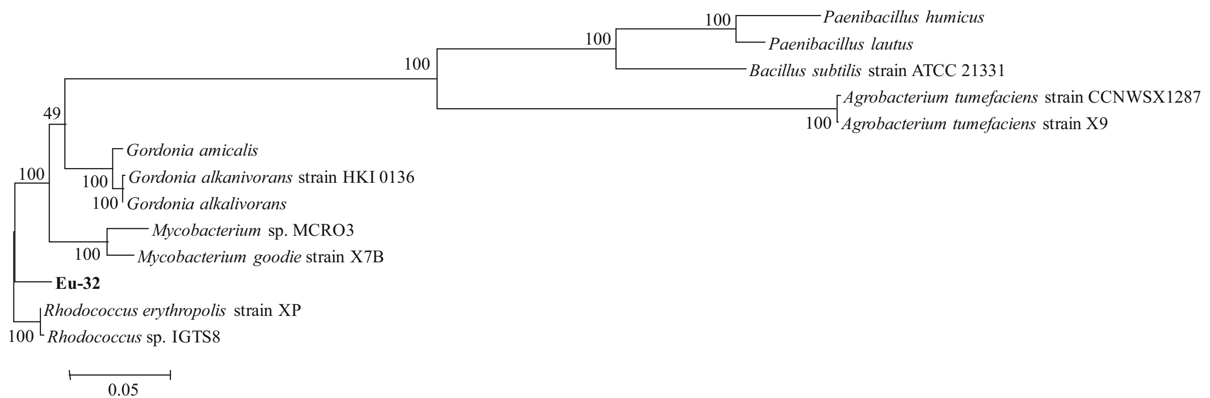

Fig. 2 Molecular Phylogenetic analysis based on 16S rRNA gene sequences, showing the relationship of isolate Eu-32 with other bacterial species by Maximum Likelihood method. The tree with the highest log likelihood (−4,420.5508) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.5580)). The rate variation model allowed for some sites to be evolutionarily invariable ([+ I], 31.1375 % sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site

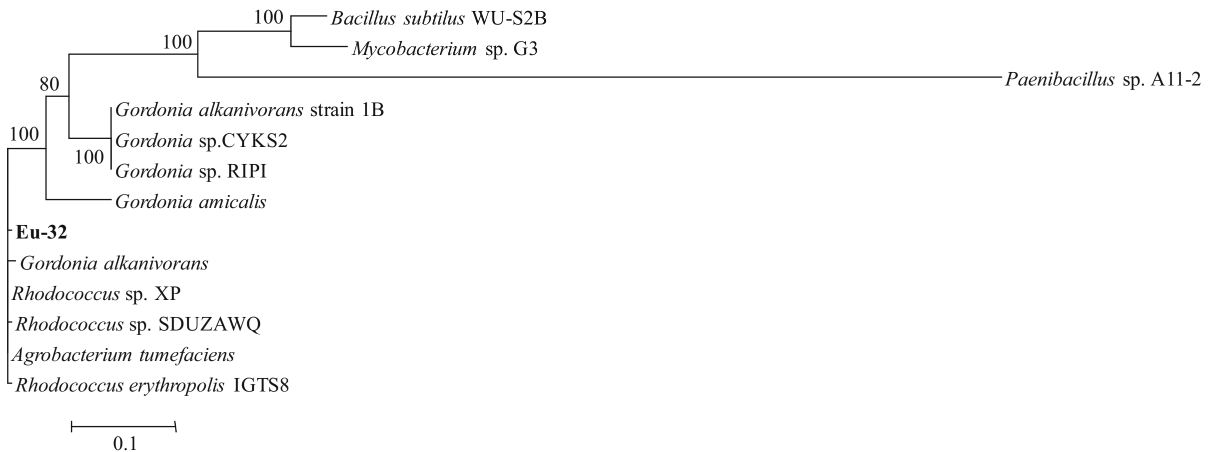

Fig. 3 Molecular Phylogenetic analysis based on the partial $FMNH_2$ dependent DBTO monooxygenase (DszC) sequence of Eu-32 (in this study) and other desulphurizing bacterial species by Maximum Likelihood method. The tree with the highest log likelihood (−1,773.7158) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 3.2778)). The rate variation model allowed for some sites to be evolutionarily invariable ([+ I], 20.8929 % sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site

of identity. The desulfurization genes have often been shown to be conserved among *Rhodococcus* species (Duarte et al. 2001). We reported earlier, that *Rhodococcus* spp. (Eu-32) desulphurized DBT through a novel extended 4S pathway (Akhtar et al. 2009) and being related to the genus *Rhodococcus,* we presumed that it may also harbour DBT desulphurizing genes (*dszABC* genes) similar to the well documented desulphurizing bacterium *Rhodococcus erythropolis* sp. IGTS8. But, the attempts to use PCR primers based on the 5′ and 3′ termini *R. erythropolis* IGTS8 desulfurization genes were completely unsuccessful in obtaining appropriate portions of the desulfurization genes from *Rhodococcus* spp. (Eu-32). The redundant nature of the genetic code makes it possible for DNA sequences to vary from one species of microorganism to another yet encode for proteins with identical amino acid sequences and vice versa. Therefore, examining protein sequences may allow the detection of conserved regions of genes that might

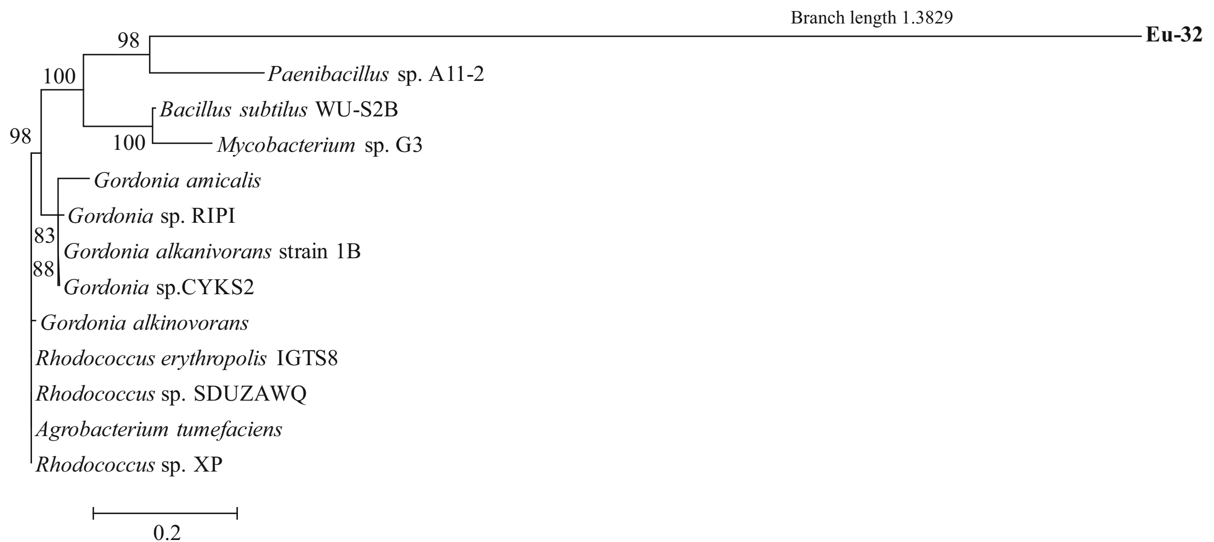

Fig. 4 Molecular phylogenetic analysis based on the partial $FMNH_2$ dependent $DBTO_2$ monooxygenase (DszA) sequence of Eu-32 (in this study) and other desulphurizing bacterial species by Maximum Likelihood method. The tree with the highest log likelihood (−1,261.6846) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 3.0844)). The rate variation model allowed for some sites to be evolutionarily invariable ([+ *I*], 12.3468 % sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site

not be as readily detected by comparing DNA sequence data (Kilbane and Robbins 2007). Conserved and unique regions of desulfurization proteins were identified, and degenerate PCR primers targeting these sites were designed.

The *dszA* gene (496 bp) of isolate Eu-32 showed that in BLAST search the most closely related sequence was from *Rhodococcus erythropolis* PR4 with 92 % sequence identity to a putative FMNH-2 dependent monooxygenase. An analysis of the deduced 164 amino acid sequence from the 496 bp *dszA* gene of isolate Eu-32 using the BLAST tools on the NCBI web site showed 77 % identity with putative $FMNH_2$ dependent monooxygenases of several bacterial species. Low identities (range 32–37 %) were found between the partial DszA of isolate Eu-32 and other corresponding DBT desulphurizing monooxygenases. The partial DszA of isolate Eu-32 showed only 36 % sequence homology with DszA of *Rhodococcus erythropolis* sp. IGTS8 (Table 4), despite the organisms being closely taxonomically related, see Fig. 2. This degree of homology was similar to that shown for other $FMNH_2$-dependent monooxygenases from taxonomically distant organisms such as the SnaA, pristinamycin II_A synthase subunit A of *Strepromyces pristinaespiralis,* 42 % homology

Table 4 The identity of deduced DszA and DszC partial sequences of *Rhodococcus* spp. (Eu-32) with related partial sequences of desulphurizing bacteria

Microorganism	% Identity	
	DszA	DszC
Rhodococcus erythropolis sp. IGTS8	36	98
Rhodococcus sp. XP	36	99
Rhodococcus sp. SDUZAWQ	36	98
Paenibacillus sp. A11-2	37	53
Bacillus subtilis WU-S2B	36	76
Gordonia alkinovorans strain 1B	36	90
Gordonia sp.CYKS2	36	90
Gordonia nitida	36	98
Gordonia sp. RIPI	36	90
Gordonia amicalis	36	90
Agrobacterium tumefaciens	36	99
Mycobacterium sp. G3	32	75

(Blanc et al. 1995) and NtaA, nitrilotriacetate monooxygenase of *Chelatobacter heintzii,* 39 % homology (Knobel et al. 1996). An analysis of changes in the putative amino acid sequence between the standard sequence of strain IGTS8 and Eu-32 sequence suggested that, over the 164-amino-acid region, there were 21 amino acids gaps at different positions, see Fig. 1a. Some highly conserved protein motifs present in all other known DszA enzymes are significantly mutated in Eu-32 yet a functional DszA protein is present. These changes did not result in a substantial modification of the function of the gene, as established from the analysis of the metabolites of extended 4S pathway of the Eu-32 (Akhtar et al. 2009).

The deduced 257 amino acid sequence from the 771 bp *dszC* gene of isolate Eu-32 showed high homologies (range 53–99 %) with dibenzothiophene monooxygenase from different bacterial species determined through alignments, as shown in Table 4. The valine at position 261 of DszC of *R. erythropolis* IGTS8 is important to the C–S bond cleavage specificity in refractory organic sulphur containing compounds (Kirimura et al. 2004; Kilbane and Robbins 2007; Arensdorf et al. 2002). This Val residue is also conserved in DszC of *Rhodococcus* spp. (Eu-32) (see Fig. 1b).

In other attempts to get complete PCR amplification of the *dszABC* operon of isolate Eu-32, the GFDRFWR region of DszC protein was used in combination with the primers specific for the *dszA* of isolate Eu-32. However, these primers did not amplify the target region of *dszABC* operon. Thus it is possible that the presumed three genes may not be present in an operon form in Eu-32, and each gene may have an independent promoter at 5′ end. Moreover, it may also be possible that the DNA sequences for presumed *dszB* gene of isolate Eu-32 are significantly different from those previously described sequences upon which we based our primer designs (Piddington et al. 1995; Ishii et al. 2000; Kirimura et al. 2004; Denome et al. 1994; Alves et al. 2007). Indeed Lee and co-workers (2006) observed that proteins having homology with DszB are often found in genomic contexts that differ from the *dsz* operon in many microbial species hinting at the presence of as yet to be discovered sulphur metabolic pathways. Likewise, we may conclude that as the *dszA* gene of Eu-32 was found in genome context of *R. erythropolis* PR4, that differ from the *dsz* operon with no discovered sulfur metabolic pathway so, it could be possible that other genes (like *dszB*) may also be homologous with genome sequence of various microbial species that has yet to be discovered for their desulfurizing abilities. Additionally Eu-32 shows a novel final step in the desulphurization pathway catalysed by DszB which liberates biphenyl from 2-HBP suggesting that the enzyme is significantly different to those previously described (Fig. 5).

Phylogenetic analysis and evidence for horizontal gene transfer (HGT)

Evidence for horizontal gene transfer events can be detected using common phylogenetic methods (Pylro et al. 2012). Four major methods namely distance, parsimony, maximum likelihood and bayesian are widely used for phylogenetic tree construction (Hall 2013). In our study we use the maximum likelihood (ML) method using the programme MEGA5 (Tamura et al. 2011). We choose ML method because it can perform better than distance or parsimony due to its statistical properties and inferential power on sequences of different lengths or divergence (Anisimova et al. 2013). In phylogenetic analysis, the long branch length shown by Eu-32 in Fig. 4 (1.3829) is attributed to the fact that DszA sequence of isolate Eu-32 is highly divergent as compared to other *Rhodococcus* species sequences. Consistent with the *dszA* gene of Eu-32 are having been acquired through horizontal gene transfer (HGT), as discussed by Denis-Larose et al. 1997 and Kirimura et al. 2004. The plasmid nature of the *dsz* genes (Piddington et al. 1995; Denome et al. 1994; DenisLarose et al. 1997; Oldfield et al. 1997) would facilitate successful transfer of these genes through HGT.

In conclusion the well characterized *dszABC* operon of *Rhodococcus erythropolis* IGTS8 strain is not found in our *Rhodococcus* Eu-32 strain with its novel desulphurization pathway. The divergence of the Eu-32 *dszA* gene from the well documented strain IGTS8 *dszA* gene and the close relatedness of the Eu-32 *dszC* gene to strain IGTS8 *dszC* gene probably reflect the events of HGT in the acquisition of desulphurization enzymes of Eu-32. A homologue to the *dszB* gene of *Rhodococcus* IGTS8 was not found in Eu-32 at all. This may be related to the novel 4S desulphurization pathway employed by Eu-32 (Fig. 5) in which HPBS is desulphurized to 2-HBP in a reaction catalysed by *dszB* and then uniquely to Eu-32

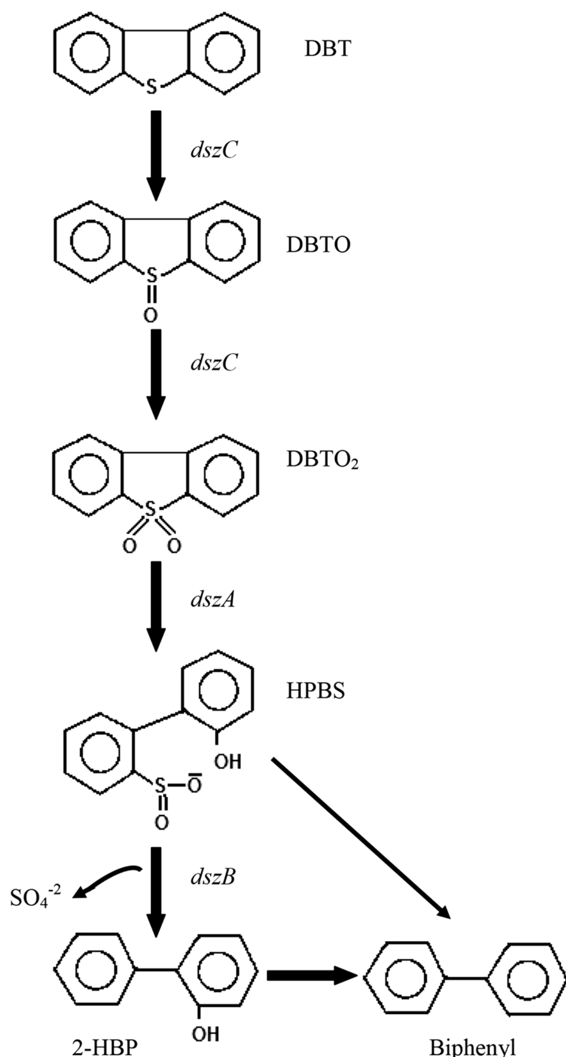

Fig. 5 Extended 4S pathway of DBT desulphurization (showing the genes involved in different steps) by isolate Eu-32 (modified from Akhtar et al. 2009). Abbreviations are as follows: *DBT* dibenzothiophene, *DBTO* DBT sulfoxide, *DBTO*$_2$ DBT sulfone, *HPBS* hydroxyphenyl benzenesulfinate, *2-HBP* 2-hydroxybiphenyl

further dehydroxylated to biphenyl. This suggests that the DszB enzyme is significantly different to other DszB enzymes previously described in biodesulphurization pathways. This hints at considerable diversity yet to be found in such pathways. Future studies will concentrate on fully characterising the activities of the *dszA* and *C* enzymes from Eu-32 and identification of the enzymic activities responsible for the formation of 2-HBP and biphenyl from HPBS i.e. the still missing *dszB* functionality.

Acknowledgments We are thankful to Ministry of Science and Technology, Pakistan, Higher Education Commission, Pakistan and University of Leicester, UK for financial support and cooperation. We thank Dr. John Kilbane of Illinois Institute of Technology, USA for providing a culture of *Rhodococcus erythropolis* IGTS8 as a gift and other technical help.

Supporting information Supplementary Fig. 1—Multiple alignment of the amino acid sequence of the desulphurizing enzyme, DBTO2 monooxygenase (DszA) from various desulphurizing bacteria. Amino acid residues conserved in all members are indicated by white letters in black boxes with an arrow sign. The regions that were diagnostic for the PCR amplification of dszA gene are shown in rectangle boxes. DszA_IGTS8, Rhodococcus erythropolis IGTS8 (AAA99482); DszA_1B, Gordonia alkinovorans strain 1B (AAT78716); BdsA_WUS2B, Bacillus subtilis WU-S2B (BAC20180) and TdsA_A11-2, Paenibacillus sp. A11-2 (BAA94831)

Supplementary Fig. 2—Multiple alignment of the amino acid sequence of the desulphurizing enzyme, DBTO monooxygenase (DszC), from various desulphurizing bacteria. Amino acid residues identical in all members are indicated by white letters in black boxes. The regions that were diagnostic for the PCR amplification of dszC gene are shown in rectangle boxes. DszC_IGTS8, Rhodococcus erythropolis IGTS8 (AAA99484); DszC_1B, Gordonia alkinovorans strain 1B (AAT78718); BdsC_WUS2B, Bacillus subtilis WU-S2B (BAC20182) and TdsC_A11-2, Paenibacillus sp. A11-2 (BAA94833).

Supplementary Fig. 3—Multiple alignment of the amino acid sequence of the desulphurizing enzyme, HPBS desulfinase (DszB), from various desulphurizing bacteria. Amino acid residues identical in all members are indicated by white letters in black boxes with an arrow sign. DszB_IGTS8, Rhodococcus erythropolis IGTS8 (AAA99483); DszB_1B, Gordonia alkinovorans strain 1B (AAT78717); BdsB_WUS2B, Bacillus subtilis WU-S2B (BAC20181) and TdsB_A11-2, Paenibacillus sp. A11-2 (BAA94832).

References

Akhtar N, Ghauri MA, Anwar MA, Akhtar K (2009) Analysis of the dibenzothiophene metabolic pathway in a newly isolated *Rhodococcus* spp. FEMS Microbiol Lett 301:95–102

Alves L, Melo M, Mendonca D, Simoes F, Matos J, Tenreiro R, Girio FM (2007) Sequencing, cloning and expression of the *dsz* genes required for dibenzothiophene sulfone desulfurization from *Gordonia alkanivorans* strain 1B. Enzyme Microb Technol 40:1598–1603

Anisimova M, Liberles DA, Philippe H, Provan J, Pupko T, Haeseler A (2013) State-of the art methodologies dictate new standards for phylogenetic analysis. BMC Evol Biol 13:161–169

Arensdorf JJ, Loomis AK, DiGrazia PM, Monticello DJ, Pienkos PT (2002) Chemostat approach for the directed evolution of biodesulfurization gain-of-function mutants. Appl Environ Microbiol 68:691–698

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (1995) Short protocols in molecular biology, 3rd edn. Wiley, New York

Blanc V, Lagneaux D, Didier P, Gil P, Lacroix P, Crouzet J (1995) Cloning and analysis of structural genes from *Streptomyces pristinaespiralis* encoding enzymes involved in the conversion of pristinamycin-II_B to pristinamycin-II_A (PII_A): PII_A synthase and NADH: riboflavin 5′-phosphate oxidoreductase. J Bacteriol 177:5206–5214

DenisLarose C, Labbe D, Bergeron H, Jones AM, Greer CW, AlHawari J, Grossman MJ, Sankey BM, Lau PCK (1997) Conservation of plasmid-encoded dibenzothiophene desulfurization genes in several rhodococci. Appl Environ Microbiol 63:2915–2919

Denome SA, Oldfield C, Nash LJ, Young KD (1994) Characterization of the desulfurization genes from *Rhodococcus* sp. strain IGTS8. J Bacteriol 176:6707–6716

Diaz E, Garcia JL (2010) Genetics engineering for removal of sulfur and nitrogen from fuel heterocycles. In: Timmis KN (ed) Handbook of hydrocarbon and lipid microbiology. Springer, Berlin

Duarte GF, van Elsas JD, Rosado AS, Seldin L, de Araujo W (2001) Analysis of bacterial community structure in sulfurous-oil-containing soils and detection of species carrying dibenzothiophene desulfurization (*dsz*) genes. Appl Environ Microbiol 67(3):1052–1062

Furuya T, Kirimura K, Kino K, Usami S (2001) Thermophilic biodesulfurization of dibenzothiophene and its derivatives by *Mycobacterium phlei* WU-F1. FEMS Microbiol Lett 204:129–133

Gray KA, Pogrebinsky OS, Mrachko GT, Xi L, Monticello DJ, Squires CH (1996) Molecular mechanisms of biocatalytic desulfurization of fossil fuels. Nat Biotechnol 14:1705–1709

Gray KA, Mrachko GT, Squires CH (2003) Biodesulfurization of fossil fuels. Curr Opin Microbiol 6:229–235

Hall BG (2013) Building phylogenetic trees from molecular data with MEGA. Mol Biol Evol 30(5):1229–1235

Ishii Y, Konishi J, Okada H, Hirasawa K, Onaka T, Suzuki M (2000) Operon structure and functional analysis of the genes encoding thermophilic desulfurizing enzymes of *Paenibacillus* sp. A11-2. Biochem Biophys Res Commun 270:81–88

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comp Appl Biosci 8:275–282

Kilbane JJ, Robbins J (2007) Characterization of the *dszABC* genes of *Gordonia amicalis* F.5.25.8 and identification of conserved protein and DNA sequences. Appl Microbiol Biotechnol 75:843–851

Kirimura K, Furuya T, Nishii Y, Ishii Y, Kino K, Usami S (2001) Biodesulfurization of dibenzothiophene and its derivatives through the selective cleavage of carbon-sulfur bonds by a moderately thermophilic bacterium *Bacillus subtilis* WU-S2B. J Biosci Bioeng 91:262–266

Kirimura K, Harada K, Iwasawa H, Tanaka T, Iwasaki Y, Furuya T, Ishii Y, Kino K (2004) Identification and functional analysis of the genes encoding dibenzothiophene-desulfurizing enzymes from thermophilic bacteria. Appl Microbiol Biotechnol 65:703–713

Knobel HR, Egli T, Vandermeer JR (1996) Cloning and characterization of the genes encoding nitrilotriacetate monooxygenase of *Chelatobacter heintzii* ATCC 29600. J Bacteriol 178:6123–6132

Konishi J, Ishii Y, Onaka T, Okumura K, Suzuki M (1997) Thermophilic carbon-sulfur-bond-targeted biodesulfurization. Appl Environ Microbiol 63:3164–3169

Lee WC, Ohshiro T, Matsubara T, Izumi Y, Tanokura M (2006) Crystal structure and desulfurization mechanism of 2′-hydroxybiphenyl-2-sulfinic acid desulfinase. J Biol Chem 281:32534–52539

Mohebali G, Ball AS (2008) Biocatalytic desulfurization (BDS) of petrodiesel fuels. Microbiol 154:2169–2183

Oldfield C, Pogrebinsky O, Simmonds J, Olson ES, Kulpa CF (1997) Elucidation of the metabolic pathway for dibenzothiophene desulphurization by *Rhodococcus* sp. strain IGTS8 (ATCC 53968). Microbiol 143:2961–2973

Piddington CS, Kovacevich BR, Rambosek J (1995) Sequence and molecular characterization of a DNA region encoding the dibenzothiophene desulfurization operon of *Rhodococcus* sp. strain IGTS8. Appl Environ Microbiol 61:468–475

Pylro VS, Vespoli LS, Duarte GF, Yotoko KSC (2012) Detection of horizontal gene transfers from phylogenetic comparisons. Int J of Evol Biol 2012:1–7

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbour, New York

Soleimani M, Bassi A, Margaritis A (2007) Biodesulfurization of refractory organic sulfur compounds in fossil fuels. Biotechnol Adv 25:570–596

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739