



# Genomic analysis provides insights into the *Gordonia* sp. W3S5 taxonomy and sulfur metabolism-related genes

Sana Parveen<sup>1</sup> · Nasrin Akhtar<sup>1</sup> · Javeria Akram<sup>1</sup>

Received: 13 October 2020 / Accepted: 17 May 2021 / Published online: 29 May 2021  
© King Abdulaziz City for Science and Technology 2021

## Abstract

The *Gordonia* sp. W3S5, isolated from oil-polluted soil samples can remove sulfur from a variety of symmetric and asymmetric thiophenic compounds and diesel oil. Its draft genome sequence was comprised of 49 contigs, total genome size 4.86 Mb, and a G + C content of 67.50%. According to the current bacterial taxonomy procedures (16S rRNA gene sequence and overall genome-related index), the W3S5 was affiliated to *Gordonia rubripertincta*. Rapid Annotation using Subsystem Technology (ClassicRAST) server revealed that the W3S5 contains 4435 coding sequences, 404 subsystems and 60 sulfur metabolism genes. The RAST comparative genomic analysis showed that the genes connected with organic sulfur metabolism are majorly related to *ssu* and *dszABC* operons. Moreover, the comparison of orthologous gene clusters using OrthoVenn2 web server revealed a total of 4869 clusters, 2685 core orthologs, 632 shared orthologs and 112 unique ortholog clusters among the W3S5 and other type strains of *Gordonia*. This is the first report describing genome-based characterization of a *Gordonia rubripertincta* strain desulfurizing thiophenic compounds and diesel oil. The desulfurization potential of *Gordonia rubripertincta* W3S5 and genomic analyses revealed it as a valuable biocatalyst for process development to desulfurize a broad range of thiophenic sulfur-containing compounds, which are a major component of organic sulfur in petroleum oil.

**Keywords** Thiophenic sulfur · Desulfurizing bacteria · Whole-genome sequencing · ANI · dDDH · Comparative genomics · Orthologous gene clusters

## Introduction

Combustion of petroleum-driven high sulfur fuels releases vast amounts of sulfur dioxide into the atmosphere and is a major cause of environmental pollution and serious health problems (Feng et al. 2016; Moheballi and Ball 2008). According to legislation, the sulfur content in diesel oil should be < 10 ppm (Sadare et al. 2017). The hydrodesulfurization (HDS) process is used in oil refineries to remove sulfur from non-volatile organic sulfur compounds in petroleum (Soleimani et al. 2007). However, it demands specific operating procedures and also ineffective to remove sulfur from recalcitrant thiophenic compounds like dibenzothiophene (DBT) and its homologous alkylated derivatives (Moheballi

and Ball 2008). The potential complementary and/or alternative to HDS is biodesulfurization where microorganisms are used which remove sulfur residues from the hydrocarbon chains in petroleum by cleaving C-S bond in the organic sulfur compounds through a specific metabolic pathway, called the “4S” pathway (Bhanjadeo et al. 2018; Boniek et al. 2015). During this sulfur-specific 4S pathway, the DBT is converted to sulfur-free compound, i.e. 2-hydroxybiphenyl (2-HBP) by an operon comprised of three genes (*dszA*, *dszB*, *dszC*) and a chromosome-born *dszD* gene (Aggarwal et al. 2013). In the majority of desulfurizing bacterial species (like *Rhodococcus*) the *dszABC* operon is located on plasmid, however; the *dsz* operon in some *Gordonia* species is also located on bacterial chromosome (Shavandi et al. 2010; Aggarwal et al. 2013).

In the current scenario, the biodesulfurization is a recommended approach to remove sulfur from complex compounds like dibenzothiophene (DBT), benzothiophene (BT), or their alkylated forms (Cx-DBTS, Cx-BTS) under mild conditions (Moheballi and Ball 2008; Akhtar et al. 2018). However, due to limited knowledge on desulfurizing traits

✉ Nasrin Akhtar  
nasrin\_379@yahoo.com; nasrin@nibge.org

<sup>1</sup> Industrial Biotechnology Division, National Institute for Biotechnology and Genetic Engineering College, Pakistan Institute of Engineering and Applied Sciences (NIBGE-C, PIEAS), Jhang Road, Faisalabad, Pakistan

of the bacterial species, the large-scale process has not been developed (Akhtar et al. 2019; Parveen et al. 2020). Genomics is one of the fastest evolving discipline of science and the next generation sequencing (NGS) has made it possible to have whole-genome sequences of various organisms not only in limited time but with minimal cost as well (Land et al. 2015). To date (NCBI January 2021), more than 0.3 million bacterial genomes have been sequenced (<https://www.ncbi.nlm.nih.gov/genome/browse/>) and this genomic information can provide an insight into the bacterial gene pool. Moreover, genomic analysis can expand the scale of microbial screening processes without labor-intensive growth experiments.

Previously, no study comprehensively describes the genome-based characterization, ortholog gene clusters identification and comparison of the sulfur metabolism-related genes of a *Gordonia* species desulfurizing thiophenic compounds and diesel oil. The current study provides an insight into the metabolic potential of the isolate W3S5 to use various alkylated derivatives of DBT, BT and thiophene as a sulfur source as well as to biodesulfurize the hydrodesulfurized diesel oil. Moreover, comparative genomic analysis of the obtained draft genome sequence has enabled us to clearly establish the taxonomic position of the bacterium, to explore its orthologous gene clusters as well as the genomic background associated with sulfur metabolism.

## Materials and methods

### Bacterial culture and the medium used

The *Gordonia* sp. W3S5 used in this study was isolated from oil-contaminated soil samples collected from a local oil drilling company (OGDCL-Rajian Oil Field, Chakwal, Pakistan). A chemically defined medium (MG medium) used for growth and other experimental schemes contained 2.0 g  $\text{KH}_2\text{PO}_4$ , 4.0 g  $\text{K}_2\text{HPO}_4$ , 1.0 g  $\text{NH}_4\text{Cl}$ , 0.2 g  $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$ , 5.0 g glucose, 10.0 ml metal solution, and 1.0 ml vitamin mixture in 1000 ml distilled water (pH 7.0) (Akhtar et al. 2018). The enriched culture was evaluated for its desulfurization activity by Gibbs test as described by Akhtar et al. 2009.

### Biodesulfurization of thiophenic compounds and diesel oil

The potential of the isolated bacterium to use different thiophenic sulfur-containing compounds as a sole source of sulfur was studied. Inoculum was prepared by growing the bacterial cells in LB media for 48 h. Then 250 mL flasks containing 50 mL MG media were provided with 0.2 mM concentration of different thiophenic compounds

and were inoculated (1% v/v, OD660 ~ 0.45 nm) in duplicates. These flasks were incubated at 30 °C in a rotary shaker with 180 rpm. The bacterial growth (OD 660 nm) was recorded every 24 h for 10 days and the production of phenolic end products was estimated by Gibbs test (Akhtar et al. 2019). Diesel oil containing 245 ppm sulfur was obtained from a local oil refinery (Attock Refinery Limited, Rawalpindi, Pakistan) and the desulfurization studies were carried out in 1L capacity flasks (5% v/v oil pulp density) using the DBT-adapted bacterial cells prepared in MG-media at the optimized growth conditions (pH 7.0, 30 °C, 180 rpm). The sulfur was estimated on the EDX-RF sulfur analyzer (Tanaka Scientific Limited, Japan) based on the American Society for Testing and Materials (ASTM) method D5453.

### Genome sequencing and assembly

The genomic DNA extraction and genome sequencing were carried out commercially on Illumina (NovaSeq) platform (2 × 100 bp paired-end) by Macrogen Inc. Korea. De novo assembly of the raw reads was performed with the SPAdes assembler (version 3.5.0), with the following assembly parameters: k automatic selection based on read length, mismatch careful mode turned ON, repeat resolution enabled, mismatch corrector not skipped, coverage cutoff turned OFF. The whole-genome SRA sequence data and bio-project/bio-sample information of the *Gordonia* sp. reported in this study are available at DDBJ/ENA/GenBank.

### Genome and taxonomic analysis

The gene prediction and annotation were done by Rapid Annotation using Subsystem Technology (ClassicRAST) server version 2.0 (default settings) (Aziz et al. 2008) and NCBI Prokaryotic Genome Annotation Pipeline (PGAP). Moreover, BlastKOALA and KEGG were used to annotate the genome sequence. The default settings parameters in BlastKOALA were set as-Taxonomy group: Prokaryotes, Bacteria; KEGG database to be searched: genus\_prokaryotes.pep. For taxonomic studies, the closely related *Gordonia* species' 16S rRNA gene sequences were taken from EzBioCloud and the National Center for Biotechnology Information databases (Yoon et al. 2017). The phylogenetic tree was constructed (tree topologies supported by bootstrap analysis) using the software package MEGA 6 (Tamura et al. 2013). The Genome-to-Genome Distance Calculator (GGDC) version 2.1 (<http://ggdc.dsmz.de>) (Meier-Kolthoff et al. 2013) was used to calculate the genomic digital DNA-DNA Hybridization (dDDH) values while the Kostas Lab webserver (<http://enve-omics.ce.gatech.edu/ani/>) (Goris

et al. 2007), was used to compute the Average Nucleotide Identity (ANI) values between the closely related species.

## Comparative genomic analysis

For comparative genomic analysis, the genome sequences of the closely related strains of *Gordonia/Rhodococcus* were obtained from the NCBI GenBank database and were uploaded in the RAST server. The accession numbers (in parentheses) of the selected genome sequences were as follows: *Gordonia rubripertincta* NBRC 101908<sup>T</sup> (NZ\_BAHB000000000.1), *Gordonia rubripertincta* CWB2 (NZ\_CP022580.1), *Gordonia rubripertincta* SD5 (NZ\_CP059694.1), *Gordonia namibiensis* NBRC 108229<sup>T</sup> (NZ\_BAHE000000000.1), *Gordonia alkanivorans* NBRC 16433<sup>T</sup> (NZ\_BACI000000000.1), *Gordonia terrae* NBRC 100016<sup>T</sup> (NZ\_BAFD000000000.1), *Gordonia terrae* C-6 (NZ\_AQPW000000000.1), *Gordonia amicalis* NBRC 100051<sup>T</sup> (NZ\_BANS01000001.1), *Gordonia amicalis* BDS-1 (NZ\_JACFXQ010000094.1), *Gordonia desulfuricans* NBRC 100010<sup>T</sup> (NZ\_BCNF000000000.1), *Gordonia desulfuricans* 213E (NZ\_JAADZU010000001.1), *Gordonia hydrophobica* NBRC 16057<sup>T</sup> (NZ\_BCWU000000000.1), *Rhodococcus* sp. Eu-32 (QGNK000000000.1), *Rhodococcus qingshengii* IGTS8 (CP029297.1) and *Rhodococcus erythropolis* XP (NZ\_AGCF000000000.1).

The orthologous gene cluster analysis of the W3S5 isolate with other type strains of Genus *Gordonia* was performed by uploading the protein sequence FASTA files in the OrthoVenn2 web platform (<https://orthovenn2.bioinfotoolkits.net/home>), keeping the threshold e-value 1e-5 and inflation as 1.5 (Xu et al. 2019; Datta et al 2020).

## Results and discussion

### Thiophenic compounds and diesel oil desulfurization potential of the isolate W3S5

The thiophenic compounds desulfurization potential of the isolate W3S5, as determined in MG media supplemented with different thiophenic sulfur-containing compounds is presented in Table 1. The growth (OD660nm) showed that the isolate W3S5 not only used DBT, BT and Thiophene but also their alkylated derivatives as a sulfur source (Supplementary Fig. 1). The growth was higher in the presence of DBT, 4-methyl DBT, 2,8-dimethyl DBT and 3-methyl BT as compared to the other compounds (Table 1). The high cell growth indicated that these organic sulfur-containing compounds have been metabolized/transformed by the W3S5 to fulfill its cellular sulfur need, suggesting its ability to decrease the sulfur contents of the fossil fuels. The culture broths of isolate W3S5, supplemented

**Table 1** Substrate spectra of strain W3S5 for various thiophenic compounds

Substrate	*Growth OD (660 nm)	*Gibbs test	*Phenolic product (mM)
Dibenzothiophene (DBT)	9.300	+	0.19
DBT sulfone	7.890	+	0.12
4-Methyl DBT	9.594	+	0.10
2,8-Dimethyl DBT	9.468	–	–
Benzothiophene (BT)	5.544	+	0.10
3-Methyl BT	9.189	+	0.10
Thiophene	3.843	–	–
2-Methyl thiophene	4.356	–	–
3-Methyl thiophene	4.761	–	–
Control (no sulfur source)	2.70	–	–

\*Growth OD 660 nm and Gibbs test recorded in stationary phase

with DBT, DBT sulfone, 4-methyl DBT, BT and 3-methyl BT were also positive for Gibbs test (Table 1) indicating that the W3S5 has desulfurized and converted these compounds into sulfur-free alkylated/non-alkylated phenolic end products. The negative Gibbs test for 2,8-dimethyl DBT showed that either the isolate W3S5 is converting it into some non-phenolic end product or the produced phenolic end product is not free to react with the Gibbs reagent. The quantification of the produced phenolic end product by Gibbs test showed that the isolate W3S5 had completely desulfurized the DBT into 2-HBP (1.99 mM) (Table 1). However, the other DBT and BT derivatives except the 2,8-dimethyl DBT were 50% desulfurized in the stationary phase.

A considerable number of bacteria that are able to desulfurize the symmetric heterocyclic sulfur-containing compounds like DBT appear to be impotent of desulfurizing the asymmetric heterocyclic sulfur-containing compounds like BT (Mohamed et al. 2015; Akhtar et al. 2018). The bacterial isolate W3S5 described in this study can be categorized as symmetric and asymmetric sulfur compound (BT and DBT) desulfurizing bacterium. Such bacteria are thought to be of incalculable value biocatalysts for process development to desulfurize a broad range of thiophenic sulfur-containing compounds present as organic sulfur in petroleum oil (Mohamed et al. 2015).

After biodesulfurization by the isolate W3S5, the total sulfur content of the diesel oil was 147 ppm which is related to a total reduction of ~40% in 30 days as compared to the control, demonstrating it as a potent desulfurizing candidate for the up-scaled studies for removal of sulfur from petroleum oils. These findings validate the ability of the isolate W3S5 to desulfurize the thiophenic

sulfur-containing compounds and diesel oil through enzymes/metabolites produced by putative genes that need to be studied in its genome.

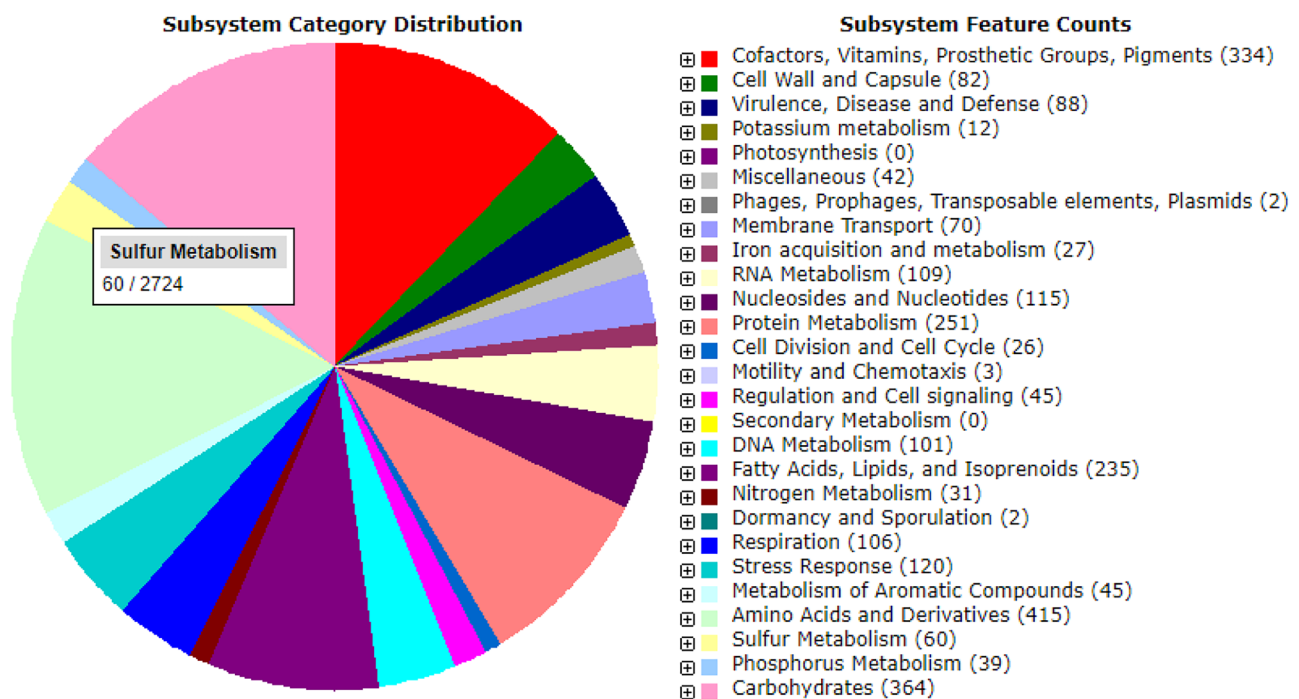
### Genome assembly and annotation

The total number of raw reads produced was 26,583,530 with Q20 and Q30 values as 97.2% and 92.14%, respectively. The high Q scores value indicated a smaller probability of error in the genomic data of the W3S5. The de novo genome assembly was comprised of 49 contigs with a genome coverage of 241X. To support the authenticity of the genome assembly (assembled using SPAdes version

3.5.0), the 16S gene sequence was used, as described by (Chun et al. 2018). The genome assembly was classified as undecided for the potential contaminations as determined by ContEst16S (Contamination Estimator by 16S) algorithm. In ClassicRAST analysis, the draft genome sequences of the isolate W3S5 comprised of total length (4,857,317 bp); contig count (49); N50 (430,858 bp), L50 (4) and G + C content (67.50%). Moreover, a total of 4,435 protein-coding sequences and 51 RNAs (48 tRNA genes and 3 copies of rRNA) were found in the genome (Table 2). Further ClassicRAST based functional gene subsystem clustering analysis revealed that 404 subsystems are present in the genomic island of the W3S5 (Table 2). The subsystems representing the amino acids and derivatives (415 ORFs); cofactors, vitamins, prosthetic groups, pigments (334 ORFs); carbohydrate metabolism (364 ORFs); protein metabolism (251 ORFs); and fatty acids, lipids, and isoprenoids (235 ORFs) were in large number. Furthermore, the subsystems connected with membrane transport (70 ORFs); stress response (120 ORFs); sulfur metabolism (60 ORFs); as well as aromatic compounds metabolism (45 ORFs) were also identified (Fig. 1). Of particular interest to organic sulfur metabolism, subsystems associated with alkanesulfonate assimilation/ utilization were indicated in the genome of the isolate W3S5.

**Table 2** Genome features of strain W3S5

Attributes	Values
Genome size	4.86 Mb
Genome coverage	241 X
G + C content	67.5%
N50	430,858
L50	4
Number of contigs	49
Number of protein coding sequences	4435
Number of subsystems	404
Number of RNAs	51
Sulfur metabolism genes	60



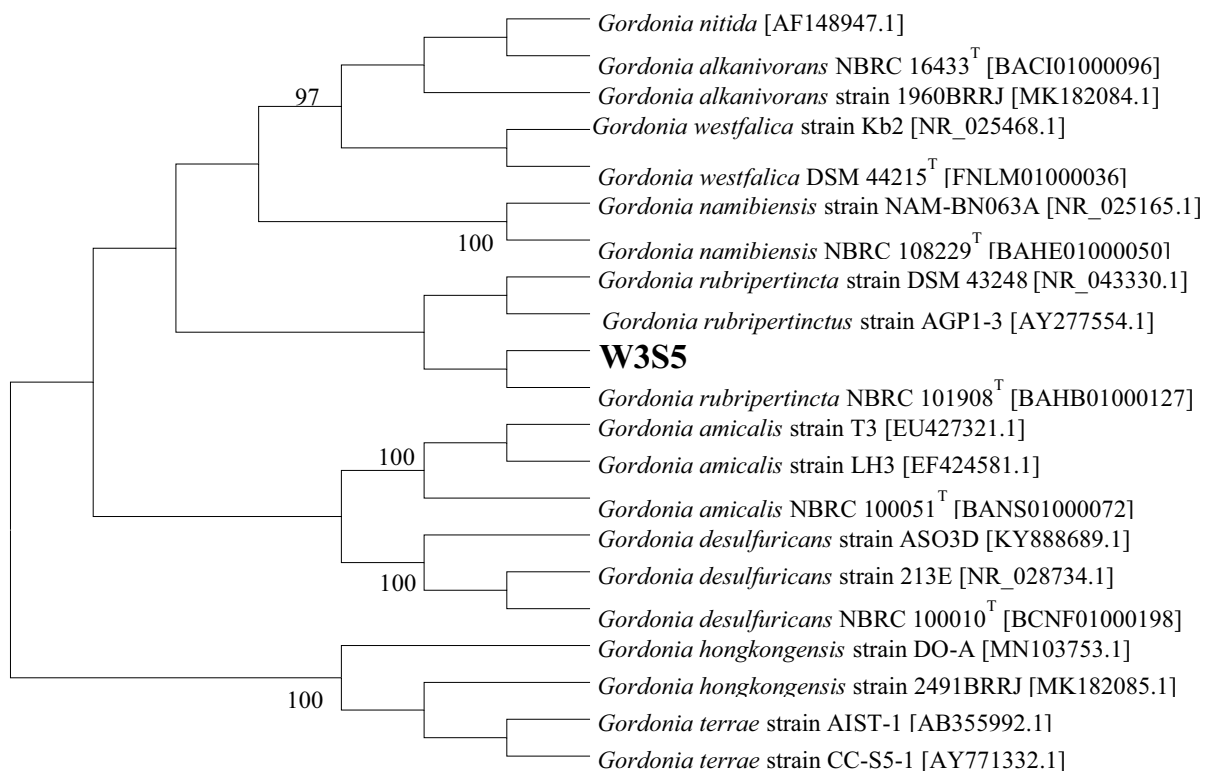
**Fig. 1** Subsystem category distribution from strain W3S5, generated through ClassicRAST pipeline (default settings)

**Table 3** The 16S rRNA gene similarity, Average Nucleotide Identity (ANI) and digital DNA-DNA Hybridization (dDDH) values by the genome comparison of W3S5 to its more closely related type strains (16S rRNA gene similarity  $\geq 98.7\%$ )

Strain Name and accession No	16S % identity	dDDH	ANI
<i>Gordonia rubripertincta</i> NBRC 101908 <sup>T</sup> [BAHB01000127]	100	83.00	98.15
<i>Gordonia namibiensis</i> NBRC 108229 <sup>T</sup> [BAHE01000050]	99.44	48.10	92.20
<i>Gordonia westfalica</i> DSM 44215 <sup>T</sup> [FNLM01000036]	99.17	35.80	88.13
<i>Gordonia alkanivorans</i> NBRC 16433 <sup>T</sup> [BACI01000096]	98.96	47.70	92.05
<i>Gordonia amicalis</i> NBRC 100051 <sup>T</sup> [BANS01000072]	98.96	34.30	87.53
<i>Gordonia desulfuricans</i> NBRC 100010 <sup>T</sup> [BCNF01000198]	98.70	21.20	80.39

## Phylogenetic and genome-based classification at the species level

For taxonomic evaluation of the bacterium, a combination of 16S similarity and OGRI (overall genome-related index, i.e. ANI and dDDH) were used systematically. As a first step, to find the closely related strains, the 16S rRNA gene sequence of the isolate W3S5 was submitted to NCBI and EzBioCloud databases (Chun et al. 2018; Yoon et al. 2017). The BLASTn and EzBioCloud search showed that the 16S nucleotide sequence of the isolate W3S5 was highly similar to different species of the genus *Gordonia* (*rubripertincta*, *namibiensis*, *alkanivorans*, *westfalica*, *amicalis*, *desulfuricans*, *nitida*, *hongkongensis* and *terrae*). On the basis of 16S rRNA gene sequence, maximum similarity (100%) of the W3S5 was found with *Gordonia rubripertincta* NBRC 101908<sup>T</sup> [BAHB01000127], followed by *Gordonia namibiensis* NBRC 108229<sup>T</sup> [BAHE01000050], 99.44%, and *Gordonia westfalica* DSM 44215<sup>T</sup> [FNLM01000036], 99.17% (Table 3). Likewise, phylogenetic analysis of the closely related 16S rRNA gene sequences showed that the W3S5



**Fig. 2** Neighbor-joining phylogenetic tree based on 16S rRNA gene sequences. The bootstrap consensus tree was inferred from 100 replicates. The evolutionary distances were computed using the Tamura-

Nei method and are in the units of the number of base substitutions per site. Evolutionary analyses were conducted in MEGA 6

formed a separate branch with the *Gordonia rubripertincta* NBRC 101908<sup>T</sup> (Fig. 2). Further, to establish more specific taxonomic position at the species level, a comparison of the genome of the W3S5 was carried out to its closely related type strains using ANI and dDDH values (Table 3). According to the current bacterial taxonomy, the projected and generally accepted dDDH and ANI values are 70% and 95–96%, respectively, between genomes of the same species (Chun et al. 2018). A comparison of the W3S5 and *Gordonia rubripertincta* NBRC 101908<sup>T</sup> revealed dDDH value above 80% and the ANI value above 98% supporting the W3S5 and *Gordonia rubripertincta* NBRC 101908<sup>T</sup> being the same species (Table 3). Taking together the 16S rRNA gene, dDDH and ANI the isolate W3S5 described here is closely related to *Gordonia rubripertincta*.

### Potential genes and regulatory sequences providing desulfurization trait in isolate W3S5

Sulfur is required by microbial cells for growth and various biochemical processes. Bacterial cell dry mass is composed of 0.5–1% sulfur (as part of amino acids, proteins and enzyme cofactors) (Feng et al. 2016). Microorganisms acquire sulfur from different sources. Some are proficient in extracting sulfur from organosulfur compounds such as DBT

and its derivatives while others can acquire from inorganic sources (Soleimani et al. 2007).

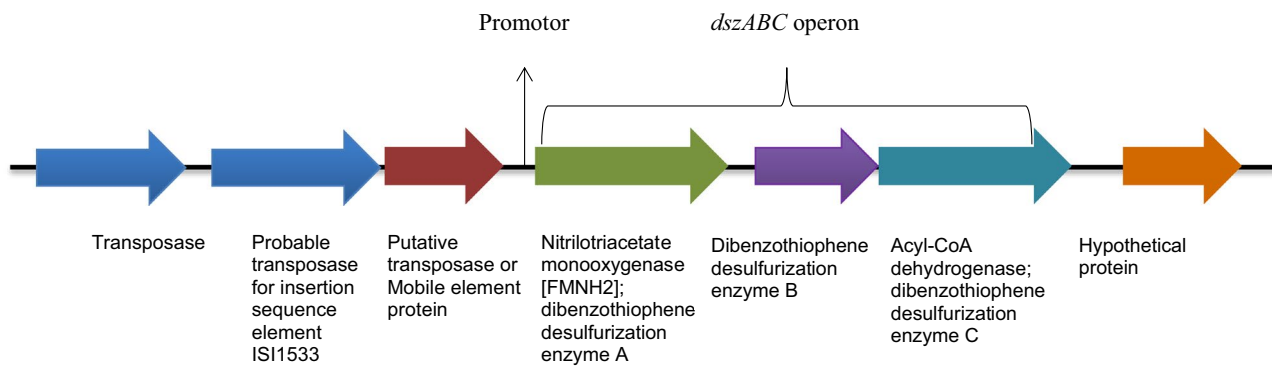
A large number of *Gordonia* species are being used in various industrial and environmental biotechnology settings. One of the important features of this genus is the capabilities of its members to transform/biodesulfurize different organic sulfur-containing compounds (Drzyzga 2012). To identify the potential genes involved in sulfur metabolism, the genome of the isolate W3S5 was annotated. The ClassicRAST based genome annotation revealed that a total of 60 genes are involved in the metabolism of sulfur (Fig. 1), of which 23 genes are associated with organic sulfur assimilation/utilization. Furthermore, the organic sulfur metabolism in the W3S5 genome consists of *ssu* and *dszABC* operons (Table 4), responsible for the desulfonation of organic sulfur-containing compounds. In *ssu* operon, mainly the *ssuD* (encoding FMNH<sub>2</sub> dependent monooxygenase) and *ssuE* (encoding NADPH dependent FMN reductase) genes are involved in the desulfonation of alkanesulfonates via carbon–sulfur bond cleavage (Aggarwal et al. 2013). Both SsuD and SsuE enzymes were found in the genome of isolate W3S5 (Table 4).

To find the genes associated with *dszABC* operon, the genomic analysis indicated that the isolate W3S5 harbored a 1251 bp gene fragment, located on contig#22 with the designated function of Acyl-CoA dehydrogenase probably

**Table 4** Comparison of the key sulfur metabolism encoding genes present in W3S5 and other related species genomes

Microorganism	Genome size (Mb)	Enzymes encoded by different genes												
		SsuA	SsuD	SsuB	SsuE	SsuC	SO	DszA	DszB	DszC	Pc	Ac	Prc	TD
W3S5 (this study)	4.86	–	SsuD	SsuB	SsuE	SsuC	SO	DszA	DszB	DszC	Pc	Ac	Prc	TD
<i>Gordonia rubripertincta</i> NBRC 101908 <sup>T</sup>	5.20	–	SsuD	SsuB	–	SsuC	SO	–	–	DszC	–	Ac	Prc	TD
<i>Gordonia rubripertincta</i> CWB2	5.33	SsuA	SsuD	SsuB	SsuE	SsuC	SO	–	–	DszC	Pc	Ac	Prc	TD
<i>Gordonia rubripertincta</i> SD5	5.26	–	SsuD	SsuB	SsuE	SsuC	SO	–	–	DszC	Pc	Ac	Prc	TD
<i>Gordonia namibiensis</i> NBRC 108229	4.94	SsuA	SsuD	SsuB	SsuE	SsuC	SO	–	–	DszC	Pc	Ac	Prc	TD
<i>Gordonia alkanivorans</i> NBRC 16433 <sup>T</sup>	5.07	–	SsuD	SsuB	SsuE	–	SO	–	–	DszC	Pc	–	–	TD
<i>Gordonia terrae</i> NBRC 100016 <sup>T</sup>	5.67	SsuA	SsuD	SsuB	SsuE	SsuC	SO	DszA	DszB	DszC	Pc	Ac	Prc	TD
<i>Gordonia terrae</i> C-6	5.17	SsuA	SsuD	SsuB	SsuE	SsuC	SO	–	DszB	–	Pc	Ac	Prc	TD
<i>Gordonia amicalis</i> NBRC 100051 <sup>T</sup>	4.92	–	SsuD	SsuB	SsuE	–	SO	–	–	DszC	Pc	Ac	–	TD
<i>Gordonia amicalis</i> BDS-1	4.98	–	SsuD	SsuB	SsuE	–	SO	DszA	DszB	DszC	Pc	Ac	–	TD
<i>Gordonia desulfuricans</i> NBRC 100010 <sup>T</sup>	5.43	SsuA	SsuD	SsuB	SsuE	SsuC	SO	–	–	DszC	Pc	Ac	Prc	TD
<i>Gordonia desulfuricans</i> 213E	5.55	SsuA	SsuD	SsuB	SsuE	SsuC	SO	–	–	DszC	Pc	Ac	Prc	TD
<i>Gordonia hydrophobica</i> NBRC 16057 <sup>T</sup>	4.58	–	SsuD	SsuB	SsuE	–	–	–	–	–	–	–	–	–
<i>Rhodococcus</i> sp. Eu-32	5.61	–	SsuD	SsuB	SsuE	SsuC	SO	DszA	–	DszC	Pc	Ac	Prc	TD
<i>Rhodococcus qingshengii</i> IGTS8 ( <i>Rhodococcus</i> sp. IGTS8)	6.93	SsuA	SsuD	SsuB	SsuE	SsuC	SO	DszA	DszB	DszC	–	Ac	Prc	TD
<i>Rhodococcus erythropolis</i> XP	7.23	SsuA	SsuD	SsuB	SsuE	SsuC	SO	DszA	DszB	DszC	–	Ac	Prc	TD

*SsuA* Alkanesulfonates binding protein; *SsuD* Alkanesulfonate monooxygenase; *SsuB* Alkanesulfonates ABC transporter ATP-binding protein; *SsuE* FMN reductase; *SsuC* Alkanesulfonates transport system permease protein; *SO* Sulfonate monooxygenase, *DszABC* Probable dibenzothio-phene desulfurization enzymes; *Pc* ABC-type nitrate/sulfonate/bicarbonate transport systems, periplasmic components; *Ac* ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase component; *Prc* ABC-type nitrate/sulfonate/bicarbonate transport system, permease component; *TD* Alpha-ketoglutarate-dependent taurine dioxygenase



**Fig. 3** Genetic structure of W3S5 *dszABC* operon (on contig#22) showing associated genes upstream and downstream of the operon. The function assignment of the ORFs was done by ClassicRAST and NCBI blastn analysis

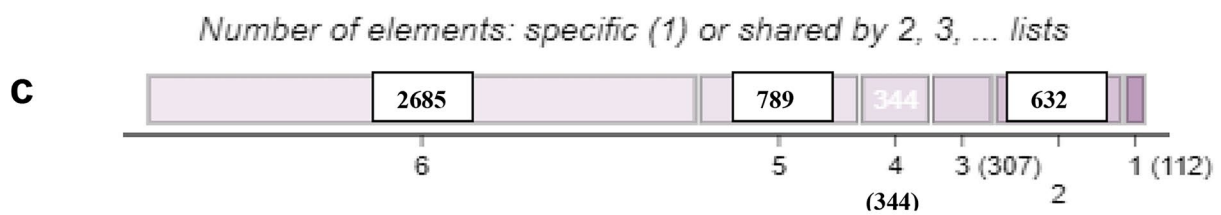
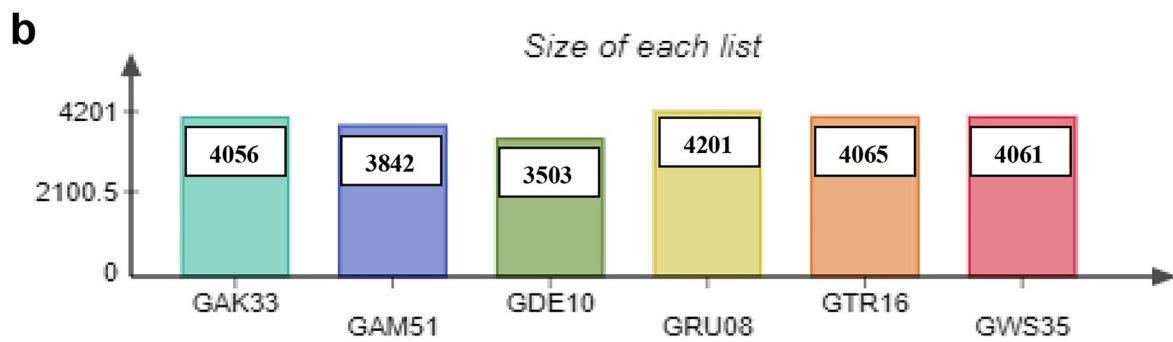
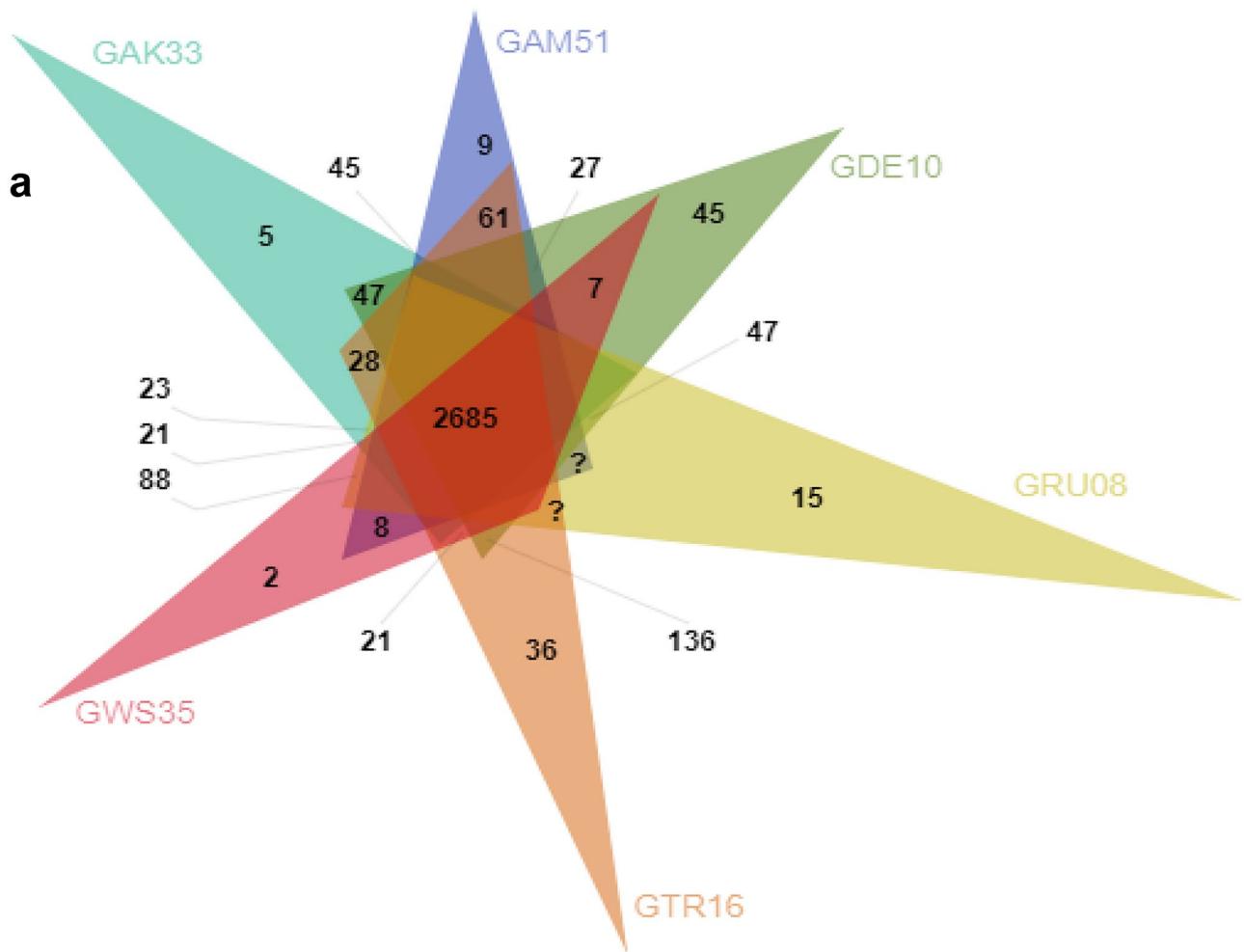
dibenzothiophene desulfurization enzyme (Fig. 3). This gene was taken from the genome sequence of the W3S5 and was further analyzed by NCBI blastn tool which showed 100% similarity with the *dszC* gene of different *Gordonia* species. Adjacent to the *dszC* gene, an 852 bp gene fragment, also identified on contig#22, with the assigned function of dibenzothiophene desulfurization enzyme B was revealed in the genome of the strain W3S5 (Fig. 3). Furthermore, a 1425 bp gene fragment, also found on contig#22, showed high similarity to nitrilotriacetate monooxygenase (FMNH<sub>2</sub>) dibenzothiophene desulfurization enzyme A. The *dszABC* genes were also verified in the genome of isolate W3S5 by performing PCR amplification (Supplementary Figs. 2, 3 and 4), cloning and sequencing. The blastn analysis of the PCR amplified *dsz* gene sequences showed 100% sequence identity to *dsz* gene sequences of different species of the *Gordonia*. Moreover, in RAST sequence alignment, the sequences of the PCR amplified *dszABC* genes were aligning with the *dszABC* gene sequences present on the contig#22 in the genome of W3S5, authenticating the genome-based detection of the sulfur metabolism genes. The *dsz* genes belong to the alkanesulfonate assimilation family, which regulates *dszABC* operon by gene activation in the presence of DBT (Shavandi et al. 2010). The sequence similarity analysis of a 200 bp nucleotide sequence upstream to the starting codon ATG of the *dszA* gene (on contig#22) using the NCBI blastn tools showed 100% similarity to the *Gordonia* strains R1P190A and CYKS1 promoter element sequences. Moreover, ClassicRAST based further analysis showed transposases like sequences in the upstream region of the promoter element (Fig. 3). A similar type of transposase sequences has been reported near the *dsz* operon of several other desulfurizing bacterial strains including *Gordonia alkanivorans* strain 1B (GenBank accession number AY678116), *Gordonia* sp. CYKS2 (GenBank accession number AY396519), and *Rhodococcus* sp. (GenBank accession number U08850). It has been reported that the

transposases in the proximity of *dsz* operon might be responsible for horizontal transfer of the desulfurization operon/genes between taxonomically dissimilar bacterial strains and its transfer from plasmid to the chromosome or vice versa (Shavandi et al. 2010).

In addition to DBT desulfurization enzymes, alkanesulfonates transport system permease proteins, alkanesulfonates ABC transporters, ATP-binding proteins and sulfonate monooxygenases encoding sequences, which are probably involved in the uptake of alkanesulfonates inside the cell, were also identified in the genome of W3S5 (Table 4). The annotation performed using KEGG in BlastKOALA (Kanehisa et al. 2016) revealed that other than the sulfur metabolism, the genome of W3S5 has a wide range of enzymes, including trehalose (17), hydratases (33) oxygenases (69), cytochromes (17), reductases (104), proteases (15) and transaminases (6), supporting it as a potential candidate for applications in industrially important bioprocesses like biodesulfurization.

### Comparative genomics studies for organic sulfur metabolism

To develop our understanding of similarities and divergences of the organic sulfur metabolism genes present in the W3S5, its genome sequence was compared with closely related type strains as well as desulfurizing strains of *Gordonia* species by uploading the genome sequences in RAST. Members of the genus *Rhodococcus* have the ability to desulfurize the thiophenic sulfur-containing compounds present in fossil fuels. Therefore, some reported biodesulfurizing *Rhodococcus* species, whose genome sequences were available in the NCBI database, were also included in the comparative analysis (Table 4). The alkanesulfonate utilization/assimilation genes associated with organic sulfur metabolism were detected in all *Gordonia* and *Rhodococcus* species. The *ssuEADCB* gene cluster is required for the utilization of





**Fig. 4** Venn diagram and the bar plots generated by OrthoVenn2 represents the distribution of shared and unique gene clusters among different biodesulfurizing *Gordonia* species. The specie names used in the analysis are; GAK33: *Gordonia alkanivorans* NBRC 16433<sup>T</sup>, GAM51: *Gordonia amicalis* NBRC 100051<sup>T</sup>, GDE10: *Gordonia desulfuricans* NBRC 100010<sup>T</sup>, GRU08: *Gordonia rubripertincta* NBRC 101908<sup>T</sup>, GTR16: *Gordonia terrae* NBRC 100016<sup>T</sup>, GWS35: W3S5. **a** The Venn diagram represents the distribution of core ortholog clusters, shared clusters and unique clusters in all the six species **b** The bar plot represents the cumulative ortholog clusters found in each species. **c** The bar plot illustrates the cumulative core, shared and unique clusters in all the six species, where label 1 on the horizontal scale shows the cumulative number of unique clusters (112) for all the six species, while label 2 shows the total number of clusters shared by two species (632) and so on

alkanesulfonates as sulfur sources. The RAST comparative genomic analysis showed that all the genes involved in this gene cluster are present in the W3S5 except the *ssuA* gene, encoding for alkanesulfonates binding proteins. The operon *dszABC* involved in the biodesulfurization of DBT, a major constituent of the thiophenic compounds present in fossil fuels was not detected in some of the species. A complete *dszABC* operon was only present in *Gordonia rubripertincta* W3S5, *Gordonia terrae* NBRC 100016<sup>T</sup>, *Rhodococcus* sp. IGTS8 (now *Rhodococcus qingshengii* IGTS8) and *Rhodococcus erythropolis* XP.

The ABC-type nitrate/sulfonate/bicarbonate transport systems, periplasmic components (PC), necessary for the initial step of sulfur oxidation pathway (Urich et al. 2006) were present in all the annotated genomes except *Gordonia rubripertincta* NBRC 101908<sup>T</sup>, *Gordonia hydrophobica* NBRC 16057<sup>T</sup>, *Rhodococcus qingshengii* IGTS8 and *Rhodococcus erythropolis* XP (Table 4). The sulfonate monooxygenase (SO), Alpha-ketoglutarate-dependent taurine dioxygenase (TD), ABC-type nitrate/sulfonate/bicarbonate transport system, permease and ATPase components (Prc & Ac), associated with the alkanesulfonate utilization/assimilation subsystem, were present in the genome of W3S5 (Table 4). However, genes for the enzymes SO, Prc, Ac and TD were either completely absent (*Gordonia hydrophobica* NBRC 16057<sup>T</sup>) or partially absent (*Gordonia alkanivorans* NBRC 16433<sup>T</sup>, *Gordonia amicalis* NBRC 100051<sup>T</sup>, *Gordonia amicalis* BDS-1) in some of the bacterial species genome sequences (Table 4).

The genome-wide analysis of orthologous gene clusters can enable us to explain the evolution and correlation of proteins across multiple species (Singh et al. 2020). The genome comparative studies of six species of *Gordonia*, including the W3S5 using OrthoVenn2 tool, generating a Venn diagram from user-defined cluster files (Xu et al. 2019; Datta et al 2020) revealed that the six species make a total of 4869 clusters, of which 2257 are orthologous clusters (contains at least two species) and 2612 are single-copy gene clusters. The Venn diagram and the bar plot (Fig. 4a, c) showed that

the numbers of core ortholog clusters shared by all the six species were 2685 that suggests their conservation in the lineage after speciation events. The cumulative number of ortholog clusters shared between any two genomes, including the W3S5, was 632. A total of 112 gene clusters were unique to only a single genome. These clusters are probably gene clusters within multiple genes or in-paralog clusters which suggest that a lineage-specific gene expansion has occurred in these gene families. The Venn diagram showed that the W3S5 shared the highest (88) clusters with *Gordonia rubripertincta* NBRC 101908<sup>T</sup> which also supported the ANI and dDDH based taxonomic evaluation of W3S5 as shown in Table 3. Additionally, the bar plot below the Venn diagram showed that the number of ortholog clusters found in the W3S5, *Gordonia terrae* NBRC 100016<sup>T</sup>, *Gordonia rubripertincta* NBRC 101908<sup>T</sup>, *Gordonia desulfuricans* NBRC 100010<sup>T</sup>, *Gordonia amicalis* NBRC 100051<sup>T</sup> and *Gordonia alkanivorans* NBRC 16433<sup>T</sup> were 4061, 4065, 4201, 3503, 3842 and 4056, respectively (Fig. 4b). The complete ortholog gene cluster analysis of protein sequences of all the six *Gordonia* strains revealed the correlation of orthologous genes among these strains.

## Conclusions

The genome-based taxonomic studies indicate that the isolate W3S5 belongs to *Gordonia rubripertincta*. It is anticipated that the identified sulfur metabolizing genes and regulatory sequences in the *Gordonia rubripertincta* W3S5 provide ecologically critical traits for cellular survival in sulfur-rich environments. The comparative genomic analysis showed that *dsz* operon encoding thiophenic sulfur desulfurization enzymes was present in *Gordonia* and *Rhodococcus* species, suggesting that it might be transferable between species of the different genera. Moreover, the orthologous gene clusters comparison among the W3S5 and *Gordonia* species revealed a conservation of developmental-related core genes and the presence of many unique genes, suggesting a lineage-specific gene expansion in these gene families. In addition, the results indicated that the *Gordonia rubripertincta* W3S5 could desulfurize different types of alkylated/non-alkylated organosulfur compounds, thus it might be a useful biocatalyst for its application in the various biodesulfurization processes.

## Accession Numbers

The whole-genome SRA sequence data of W3S5 is available in the NCBI database under the SRA accession PRJNA555169. Moreover, the whole-genome shotgun project of *Gordonia* sp. W3S5 reported in this study is available at DDBJ/ENA/GenBank under the accession NZ\_VLNS00000000 (BioProject: PRJNA555169; BioSample: SAMN12302752), version VLNS00000000.1. The W3S5 partial 16S rRNA gene sequence is available at GenBank under accession number MH569672.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s13205-021-02850-4>.

**Acknowledgements** This study was supported by the International Foundation for Science (IFS), Sweden (Grant No. F-5379-2) and Higher Education Commission (HEC), Pakistan (Project No. 3884).

## Declarations

**Conflict of interest** The authors declare that there is no conflict of interests.

## References

- Aggarwal S, Karimi IA, Ivan GR (2013) In silico modeling and evaluation of *Gordonia alkanivorans* for biodesulfurization. *Mol Biosyst* 9:2530–2540
- 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(1):95–102
- Akhtar N, Akhtar K, Ghauri MA (2018) Biodesulfurization of thiophenic compounds by 2-hydroxybiphenyl resistant *Gordonia* sp. HS126-4N carrying *dszABC* genes. *Curr Microbiol* 75:597–603
- Akhtar N, Ghauri MA, Akhtar K, Parveen S, Farooq M, Ali A, Schirack P (2019) Comparative analysis of draft genome sequence of *Rhodococcus* sp. Eu-32 with other *Rhodococcus* species for its taxonomic status and sulfur metabolism potential. *Curr Microbiol* 76:1207–1214
- Aziz RK, Bartels D, Best AA et al (2008) The RAST server: rapid annotation using subsystem technology. *BMC Genomics* 9:75
- Bhanjadeo MM, Rath K, Gupta D, Pradhan N, Biswal SK, Mishra BK, Subudhi U (2018) Differential desulfurization of dibenzothiophene by newly identified MTCC strains: influence of operon array. *PLoS ONE* 13:e0192536
- Boniek D, Figueiredo D, dos Santos AFB, de Resende Stoianoff MA (2015) Biodesulfurization: a mini review about the immediate search for the future technology. *Clean Technol Environ Policy* 17:29–37
- Chun J, Oren A, Ventosa A et al (2018) Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int J Syst Evol Microbiol* 68:461–466
- Datta S, Saha D, Chattopadhyay L et al (2020) Genome comparison identifies different *Bacillus* species in a bast fibre-retting bacterial consortium and provides insights into pectin degrading genes. *Sci Rep* 10:8169
- Drzyzga O (2012) The strengths and weaknesses of *Gordonia*: a review of an emerging genus with increasing biotechnological potential. *Crit Rev Microbiol* 38(4):300–316
- Feng S, Yang H, Zhan X, Wang W (2016) Enhancement of dibenzothiophene biodesulfurization by weakening the feedback inhibitions effects based on a systematic understanding of the biodesulfurization mechanism by *Gordonia* sp. through the potential “4S” pathway. *RSC Adv* 6:82872–82881
- Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM (2007) DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 57:81–91
- Kanehisa M, Sato Y, Morishima K (2016) BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 428:726–731
- Land M et al (2015) Insights from 20 years of bacterial genome sequencing. *Funct Integr Genomics* 15:141–161
- Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinform* 14:60
- Mohamed ME-S, Al-Yacoub ZH, Vedakumar JV (2015) Biocatalytic desulfurization of thiophenic compounds and crude oil by newly isolated bacteria. *Front Microbiol* 6:1–12
- Mohebbali G, Ball AS (2008) Biocatalytic desulfurization (BDS) of petrodiesel fuels. *Microbiol* 154:2169–2183
- Parveen S, Akhtar N, Ghauri MA, Akhtar K (2020) Conventional genetic manipulation of desulfurizing bacteria and prospects of using CRISPR-Cas systems for enhanced desulfurization activity. *Crit Rev Microbiol* 46:300–320
- Sadare OO, Obazu F, Daramola MO (2017) Biodesulfurization of petroleum distillates—current status. *Opportun Future Challeng Environ* 4:85
- Shavandi M, Sadeghizadeh M, Khajeh K, Mohebbali G, Zomorodipour A (2010) Genomic structure and promoter analysis of the *dsz* operon for dibenzothiophene biodesulfurization from *Gordonia alkanivorans* RIPI90A. *Appl Microbiol Biotechnol* 87:1455–1461
- Singh AK, Joshi I, Kumar K, Sirohi, (2020) Comparative analysis of orthologous genes between *Anguina tritici*, *Ditylenchus destructor* and *Meloidogyne incognita*. *Int J Curr Microbiol App Sci* 9:391–398
- Soleimani M, Bassi A, Margaritis A (2007) Biodesulfurization of refractory organic sulfur compounds in fossil fuels. *Biotechnol Adv* 25:570–596
- Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725–2729
- Urich T, Gomes CM, Kletzin A, Frazao C (2006) X-ray structure of a self-compartmentalizing sulfur cycle metalloenzyme. *Science* 311:996–1000
- Xu L, Dong Z, Fang L et al (2019) OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res* 47:52–58
- Yoon SH, Ha SM, Kwon S, Lim J, Kim Y et al (2017) Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int J Syst Evol Microbiol* 67:1613–1617