J Korean Soc Appl Biol Chem (2014) 57(4), 419–427
DOI 10.1007/s13765-014-4158-1

Online ISSN 2234-344X
Print ISSN 1738-2203

ARTICLE

Construction of Aligned Database of *dsr*A, a Gene Encoding Dissimilatory Sulfite Reductase Alpha Subunit, for Metagenomic Studies of Sulfate-Reducing Bacteria

Ji-Hoon Lee · Bo-A Kim · Min-Gyu Ki

Received: 19 May 2014 / Accepted: 4 June 2014 / Published Online: 31 August 2014

Abstract A sulfite reductase gene sequence database, which could facilitate analysis of metagenomic data of a functional gene from sulfate-reducing bacteria, is described. The database contains 127 aligned nucleotide sequences of a gene (*dsr*A), encoding a dissimilatory sulfite reductase alpha subunit, retrieved from Reference Sequence of GenBank that provides only well-annotated genome sequences. The *dsr*A gene sequences were screened for length and aligned using four multiple sequencing alignment programs, Mafft, Muscle, Mothur, and Clustal Omega. In addition to the *dsr*A gene sequences, 16S ribosomal RNA (rRNA) gene sequences for the bacterial species appearing in the database were also retrieved from the GenBank and aligned by a public database of SILVA in Mothur to examine any potential relationship between taxonomy of the bacteria and distribution of a functional gene. The aligned *dsr*A and 16S rRNA gene sequences were used to construct neighbor-joining phylogenetic trees for comparison of the alignments for relative agreement among the aligned sequence databases. This method of aligned reference database construction using public database can help provide better insight into microbial communities by suggesting an alternative approach of studying metagenomic analysis of functional genes involved in important ecological processes.

Keywords dissimilatory sulfite reductase gene · functional gene · metagenomics · reference database

J.-H. Lee (✉)
Groundwater Department, Korea Institute of Geoscience and Mineral Resources (KIGAM), Daejeon 305-350, Republic of Korea
Department of Mineral & Groundwater Resources, Korea University of Science and Technology (UST), Daejeon 305-350, Republic of Korea
E-mail: jihoonlee@kigam.re.kr

B.-A Kim · M.-G. Ki
Groundwater Department, Korea Institute of Geoscience and Mineral Resources (KIGAM), Daejeon 305-350, Republic of Korea

Introduction

Biogeochemical contributions of microorganisms on subsurface soil and groundwater have been investigated intensively to discover diverse reactions, since the discovery of metal-reducing microorganisms (Lovley and Phillips, 1988; Myers and Nealson, 1988). Together with and before the blossom of the topic of microbial metal reduction, microbial sulfate reduction has also been one of the main research topics in microbial biogeochemistry for several decades (Postgate, 1984). Sulfate-reducing bacteria (SRB) are ubiquitously found in nature, especially in anoxic environments. Through electron transport chain system to generate chemical energy in the form of ATP, SRB typically use sulfate as the terminal electron acceptor for anaerobic respiration of hydrogen and various organic acids as the electron donors, while generating hydrogen sulfide (H_2S) as a reduction product, a highly reactive and toxic chemical. SRB play important roles in linking the sulfur and carbon cycles in many ecological niches, such as hydrothermal vents and acid mine drainages. In addition, for environmental applications, SRB have been studied for bioremediation of toxic metal ions utilizing their metabolism of direct electron deliver to oxidized metals, including uranium(VI), technetium(VII), and chromium(VI), reducing them to less soluble and less mobile forms (Lovley and Phillips, 1992; 1994; Lloyd et al., 1999; Michel et al., 2001; Payne et al., 2002), as well as the metabolism of H_2S generation inducing precipitation of metal sulfides in acidic aquatic environments.

In the center of dissimilatory sulfate reduction pathway, dissimilatory sulfite reductase (Dsr) catalyzes the six-electron reduction of SO_3^{2-} to H_2S, the main energy-conserving step in sulfate respiration, with prior steps of SO_4^{2-} activation to adenosine 5'-phosphosulfate (APS) and APS reduction to SO_3^{2-}, respectively by sulfate adenylyltransferase (Sat) and adenylylsulfate reductase. Microorganisms capable of reducing sulfite to sulfide during anaerobic respiration have been reported to harbor Dsr

systems; therefore, genes encoding these enzymes (*dsr*AB) have been widely used as phylogenetic markers (Wagner et al., 1998; Stahl et al., 2002; Geets et al., 2006). By virtue of recent leap of metagenomic technology (*e.g.*, next generation sequencing, NGS), microbial DNAs extracted from environmental samples can be directly sequenced, generating enormous sizes of data. Although the majority of bacterial community analyses are performed with 16S rRNA gene sequences for their phylogenetic investigations, any gene can be analyzed by the metagenomic technology, using nucleotide sequences. However, the reference databases or libraries, to which the NGS-generated sequences are compared or aligned, are crucial to the downstream analyses. Quality of reference sequences and their alignments are also very important factors controlling the quality of the analytical results to understand the relationship between gene diversity and function for important environmental processes. Although public databases for 16S rRNA genes have been well established as such in SILVA (Quast et al., 2013) and greengenes (DeSantis et al., 2006), databases for functional genes to be used as NGS references are still in development, with database of *nif*H gene encoding a dinitrogenase reductase subunit (Gaby and Buckley, 2014) and a database repository, FunGene (Fish et al., 2013). Here we provide a database of *dsr*A gene, encoding a dissimilatory sulfite reductase alpha subunit, to be used as a reference for metagenomic analyses.

Materials and Methods

***dsr*A gene collection and screening.** The nucleotide sequences of *dsr*A gene, encoding dissimilatory sulfite reductase alpha subunit were retrieved from GenBank (Benson et al., 2013) database as of 4 March 2014. A total of 150 sequences of *dsr*A were manually collected only from the Reference Sequence (RefSeq) collection, which suggests well-annotated genomic DNAs to ensure quality of the sequences. Due to the wide range of sequence lengths of 150 sequences (111 to 1500 nucleotides with median and mean of 1221 and 993, respectively); sequences shorter than 500 nucleotides were removed for further analysis, resulting in a total of 127 sequences (546 to 1500 nucleotides with median and mean of 1254 and 1134) (Fig. 1), considering coverage of *dsr*A gene sequences, which would be generated by NGS technologies such as pyrosequencing and Illumina.

16S rRNA gene collection, alignment, and phylogenetic tree construction. In addition to *dsr*A gene, 16S rRNA gene sequences of 83 bacterial species composed 127 *dsr*A sequences were also retrieved manually from the GenBank to compare the relationships of the bacteria in taxonomic space with the relationships of the functional gene, *dsr*A without taxonomic consideration. Some of the bacterial species contained more than one *dsr*A gene. Eighty-three 16S rRNA gene sequences were aligned against SILVA database (Quast et al., 2013) in Mothur (Schloss et al., 2009), and the aligned sequences were added to The All-Species Living Tree (LTP) (Yarza et al., 2008) in ARB (Ludwig et al., 2004).

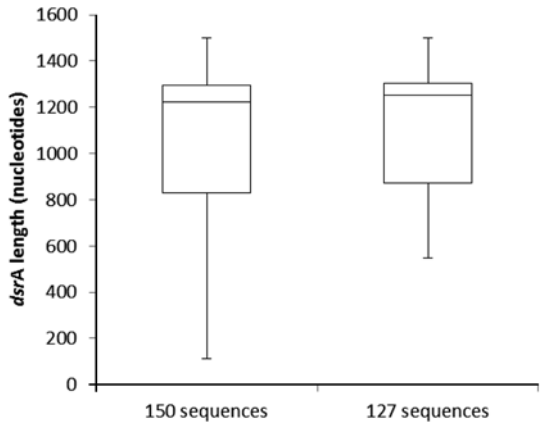

Fig. 1 Box plots displaying summaries of the *dsr*A gene sequences before and after removal of sequences less than 500 nucleotides, showing maximum, 75th percentile, median, 25th percentile, and minimum values in length.

***dsr*A gene alignment.** A total of 127 *dsr*A sequences were aligned using multiple sequencing alignment (MSA) programs, including Mafft (Katoh et al., 2002), Muscle (Edgar, 2004), Mothur, and Clustal Omega (Sievers et al., 2011). Unlike the other standalone MSAs, a seed reference sequence database was utilized for alignment by Mothur. A preliminary alignment of the 127 sequences by Muscle was clustered with dissimilarity of 3%, and one representative sequence was chosen from each group (operational taxonomic unit, OTU) in Mothur, which was used as the seed for the alignment by Mothur.

Phylogenetic trees for *dsr*A gene sequences. For comparison of the alignments generated by the above programs, qualitative approach was employed rather than quantitative comparison, by comparing phylogenetic trees of *dsr*A genes on some of representative bacterial strains with each other and also with the tree of 16S rRNA genes. For each alignment, neighbor-joining phylogenetic tree was constructed from Juke-Cantor-corrected distance matrix with a bootstrap of 1,000 replicates using ARB.

Results and Discussion

Analysis of the retrieved *dsr*A gene sequences. Among the collected *dsr*A gene sequences retrieved from the public library of GenBank RefSeq as mentioned above, as much as 68% of the screened 127 *dsr*A sequences originated from the members of classes Clostridia and Deltaproteobacteria (Fig. 2), and used for the following downstream analysis. Most abundant genera included *Desulfitobacterium*, *Desulfosporosinus*, *Desulfotomaculum*, *Desulfomonile*, and *Desulfovibrio* (Fig. 2). The sequences were aligned by four MSA programs as mentioned above, including Mafft, Muscle, Mothur, and Clustal Omega. The alignments generated by the MSA programs were not manually modified to

Phylum	Class	127
Actinobacteria	Actinobacteria	4
Chlorobi	Chlorobea	11
Firmicutes	Bacilli	1
	Clostridia	52
	Negativicutes	2
Nitrospirae	Nitrospira	1
Proteobacteria	Alphaproteobacteria	4
	Betaproteobacteria	6
	Deltaproteobacteria	34
	Epsilonproteobacteria	1
	Gammaproteobacteria	8
Spirochaetae	Spirochaetes	1
Thermodesulfobacteria	Thermodesulfobacteria	2

Fig. 2 Taxonomic summary of 127 *dsr*A gene sequences retrieved from well-annotated public database of genome sequences (RefSeq). Two bar charts display compositions of the most abundant Classes (Clostridia and Deltaproteobacteria) at genus level.

compare them. However, the alignments by any of current MSA programs might not be perfect or might not be easy to assess accuracy of the alignments. Although many advances have been achieved in MSA, there are still many challenges in alignment methods in today's high throughout biotechnologies (Thompson et al., 2011).

Comparison analysis of phylogenetic trees for 16S rRNA and *dsr*A genes. The phylogenetic tree for 16S rRNA genes was visualized with only 33 bacterial species, which were selected as representative species for genera that contained more than two different species, and an outgroup of *Thermaerobacter marianensis* type strain (Fig. 3). In addition, for the *dsr*A gene phylogenetic

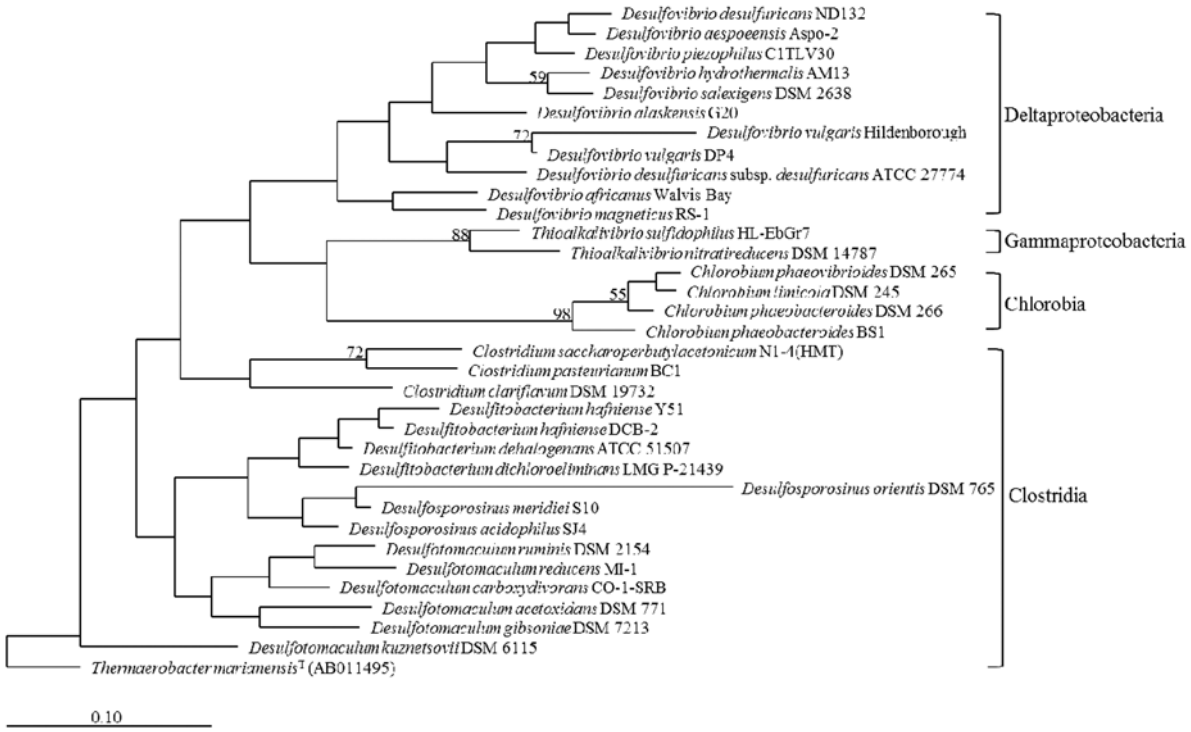

Fig. 3 Neighbor-joining phylogenetic tree of 16S rRNA gene sequences for 33 representative species appeared in the *dsr*A gene collection, and *Thermaerobacter marianesis* type strain as the outgroup constructed in ARB. Bootstrap percentages larger than 50 were only shown from 1000 replicates. Scale bar indicates 0.1 nucleotide substitution per site.

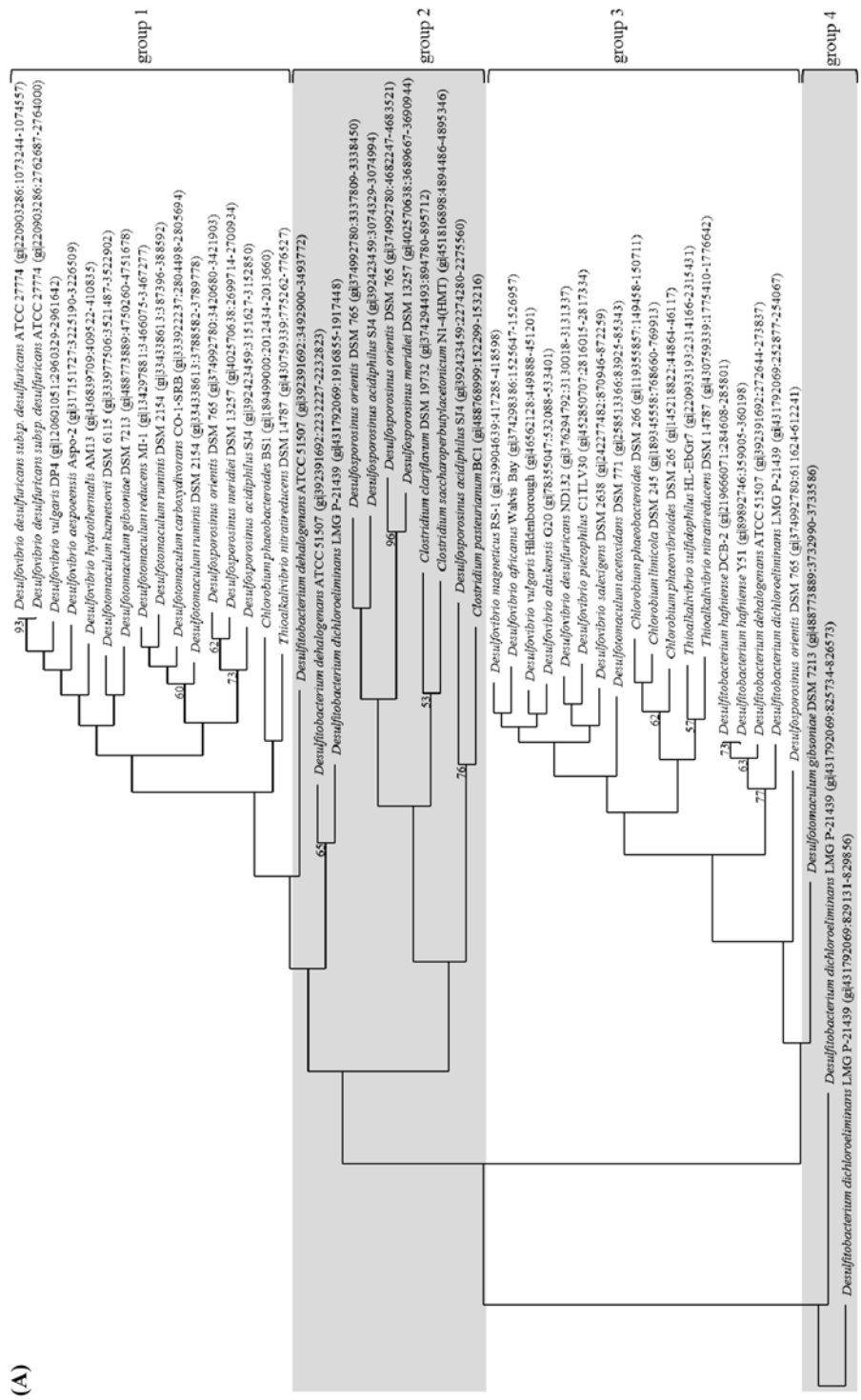

Fig. 4 Neighbor-joining phylogenetic trees of 48 *dsr*A gene sequences from the representative bacterial species analyzed for 16S rRNA gene, that were aligned by Mafft (A), Muscle (B), Mothur (C), and Clustal Omega (D), constructed in ARB. Bootstrap percentages larger than 50 were only shown from 1000 replicates. Scale bar indicates 0.1 nucleotide substitution per site. GenBank sequence identification number, gi is provided with region of the gene for each sequence in parentheses.

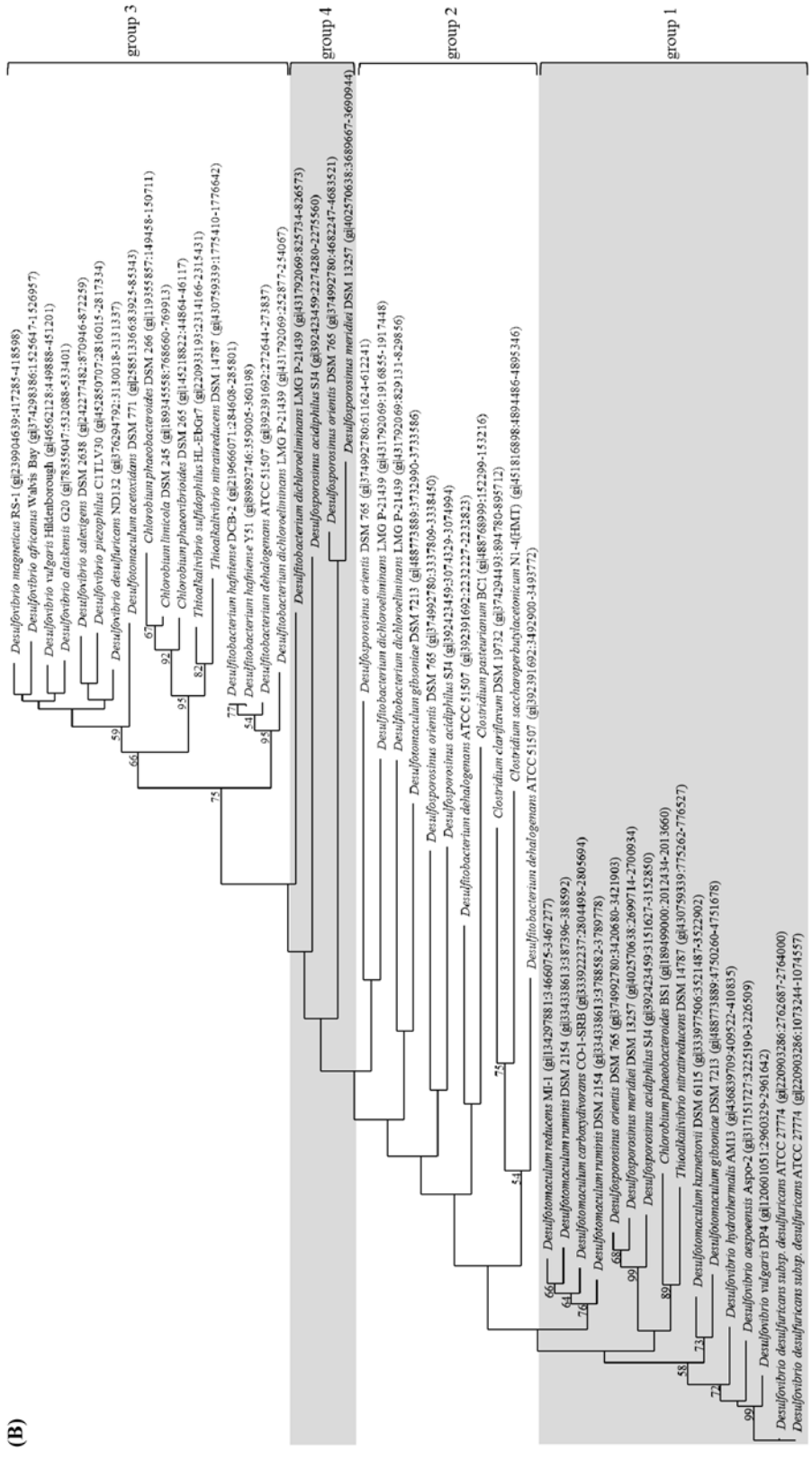

Fig. 4 Continued.

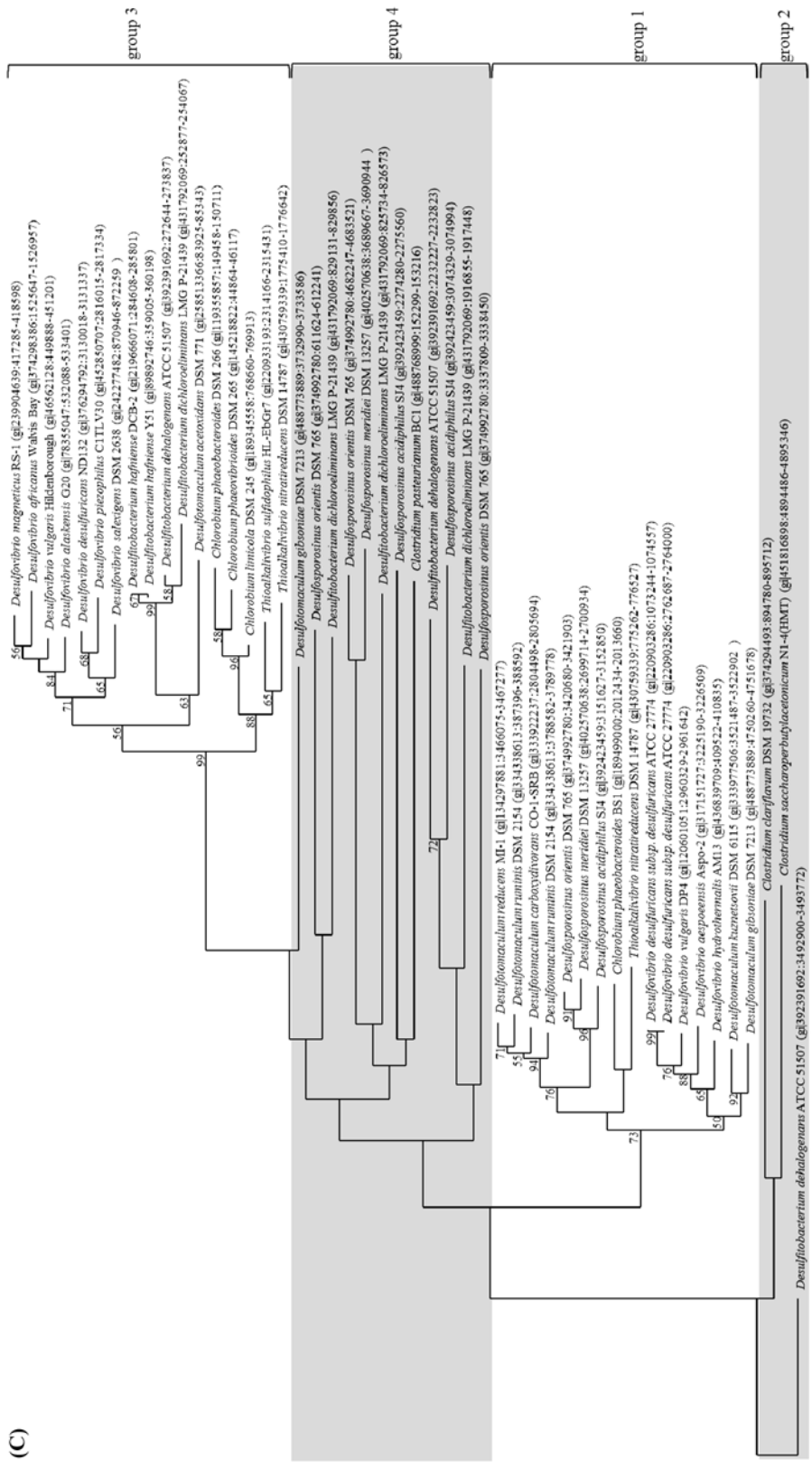

Fig. 4 Continued.

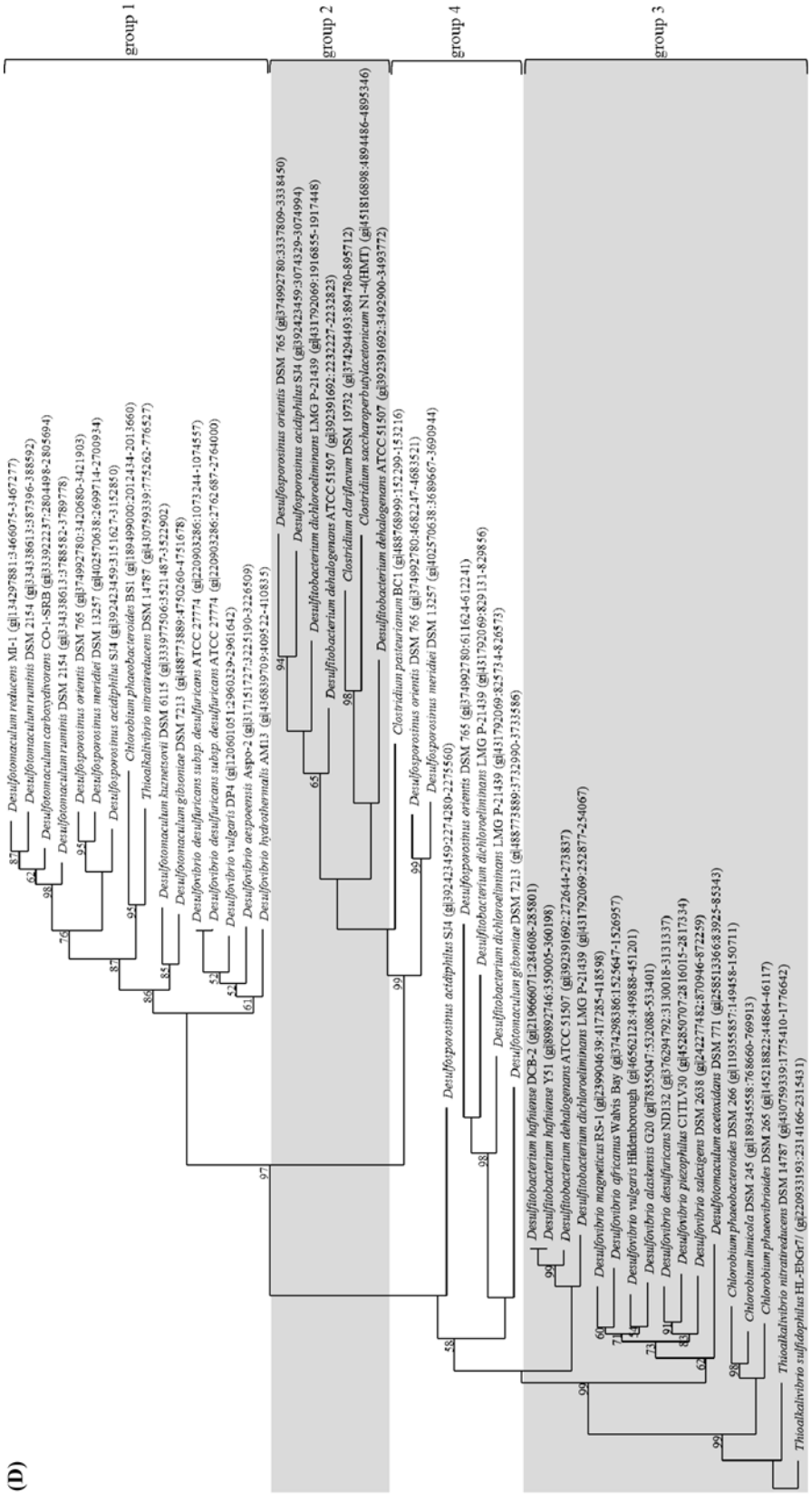

Fig. 4 Continued.

trees, due to their large sizes, the trees were trimmed to manageable sizes with only 48 sequences from the 33 bacterial species that were used as representatives for 16S rRNA genes, and the others were excluded from the trees. The trees built for the alignments by Mafft, Muscle, Mothur, and Clustal Omega are shown in Fig. 4. The aligned *dsr*A sequences were grouped into four by comparing the four trees manually, as indicated in the figures. Whereas groups 1 and 3 were relatively stable in all the trees, groups 2 and 4 varied between the trees (Fig. 4A, B, C, and D), an indication that two distinctive types of *dsr*A genes were clustered separately and the other two groups (2 and 4) of the gene might share similar sequences with both the two distinctive groups (1 and 3) partially or have experienced different evolutionary pathways. The two distinctive groups have also higher percentages of bootstrap replicates for the internal nodes and leaves of the trees, more than the other two variable groups, supporting the obvious divergent genes by groups 1 and 3. Looking into the trees, sequences from the same bacterial strains were not frequently found both in groups 1 and 3 with the exception of 2 strains (*Desulfosporosinus orientis* DSM 765, *Thioalkalivibrio nitratireducens* DSM 14787) (Fig. 4A and B), whereas two sequences from a single strain, *Desulfovibrio desulfuricans* subsp. *desulfuricans* ATCC 27774 (gi|220903286) were the closest neighbors to each other. In this case, two very similar genes are duplicated in the same organism. Thus, understanding and revealing the reason for the presence of duplicated genes in the same organisms might be a questionable issue in the future researches. Certain sequences from two strains of a species were close with each other in certain cases (*e.g.*, *Desulfitobacterium hafniense* strains DCB-2 and Y51 in group 3), and also were divergent in some other cases (*Desulfovibrio vulgaris* strains Hildenborough and DP4, and *Chlorobium phaeobacteroides* strains DSM 266 and BS1 in groups 3 and 1, respectively) (Fig. 4C). Certain strains showed multiple inclusion of *dsr*A genes widely distributed in 3 groups, such as *Desulfosporosinus orientis* DSM 765 in groups 1, 2, and 4, and *Desulfitobacterium dichloroeliminans* LMG P-21439 in groups 2, 3, and 4 (Fig. 4D). Upon comparison of the 4 trees, the trees for Muscle and Clustal Omega alignments match well rather than the others (Fig. 4B and D), though all 4 trees resemble each other regardless of the orders of the groups. Tree for Mothur alignment appears to be close to the one for Mafft, rather than the Muscle tree, although the alignment by Mothur was based on a Muscle-seed alignment. Considering higher bootstrap percentages on groups 1 and 3, and sequence relationships between neighbors on the phylogenetic trees described above, alignments by Mothur and Clustal Omega seem to suggest slightly better results than those by Mafft and Muscle, although quantitative comparison was not suggested. The processes for database construction and its validation described in this study is not a decisive method including certain quantitative parameters. On the other hand, the alignments by those MSA programs and examining the results with phylogenetic trees could suggest an alternative method to construct reference database for use in metagenomic data processing of functional genes, by investing humane efforts to examine the positions and relationships of the sequences in the trees. The methodology of construction of a functional gene database and alignment can be extended to diverse genes, and evolutional investigations also can be considered by comparing phylogenetic relationships of both bacterial taxonomy and the functional genes.

Acknowledgment This work was supported by the basic research projects (14-3211 and 14-3211-2) of the Korea Institute of Geoscience and Mineral Resources (KIGAM) funded by Korean Ministry of Science, ICT and Future Planning.

References

Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J et al. (2013) GenBank. *Nucleic Acids Res* **41**, D36–42.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**, 5069–72.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–7.

Fish JA, Chai B, Wang Q, Sun Y, Brown CT, and Tiedje JM (2013) FunGene: the functional gene pipeline and repository. *Front Microbiol* **4**, 291.

Gaby JC and Buckley DH (2014) A comprehensive aligned *nif*H gene database: a multipurpose tool for studies of nitrogen-fixing bacteria. *Database* **2014**, bau001.

Geets J, Borremans B, Diels L, Springael D, Vangronsveld J, van der Lelie D et al. (2006) DsrB gene-based DGGE for community and diversity surveys of sulfate-reducing bacteria. *J Microbiol Methods* **66**, 194–205.

Katoh K, Misawa K, Kuma Ki, and Miyata T (2002) MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059–66.

Lloyd JR, Ridley J, Khizniak T, Lyalikova NN, and Macaskie LE (1999) Reduction of technetium by *Desulfovibrio desulfuricans*: Biocatalyst characterization and use in a flowthrough bioreactor. *Appl Environ Microbiol* **65**, 2691–6.

Lovley DR and Phillips EJ (1988) Novel mode of microbial energy metabolism: Organic carbon oxidation coupled to dissimilatory reduction of iron or manganese. *Appl Environ Microbiol* **54**, 1472–80.

Lovley DR and Phillips EJP (1992) Reduction of uranium by *Desulfovibrio desulfuricans*. *Appl Environ Microbiol* **58**, 850–6.

Lovley DR and Phillips EJP (1994) Reduction of chromate by *Desulfovibrio vulgaris* and its c_3 cytochrome. *Appl Environ Microbiol* **60**, 726–8.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar et al. (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32**, 1363–71.

Michel C, Brugna M, Aubert C, Bernadac A, and Bruschi M (2001) Enzymatic reduction of chromate: comparative studies using sulfate-reducing bacteria. Key role of polyheme cytochromes c and hydrogenases. *Appl Microbiol Biotechnol* **55**, 95–100.

Myers CR and Nealson KH (1988) Bacterial manganese reduction and growth with manganese oxide as the sole electron acceptor. *Science* **240**, 1319–21.

Payne RB, Gentry DM, Rapp-Giles BJ, Casalot L, and Wall JD (2002) Uranium reduction by *Desulfovibrio desulfuricans* strain G20 and a cytochrome c_3 mutant. *Appl Environ Microbiol* **68**, 3129–32.

Postgate JR (1984) In T*he Sulphate-Reducing Bacteria*, (2nd ed). Cambridge University Press, UK.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P et al. (2013) The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res* **41**, D590–6.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB et al. (2009) Introducing mothur: Open-Source, Platform-Independent, Community-

Supported Software for Describing and Comparing Microbial Communities. *Appl Environ Microbiol* **75**, 7537–41.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W et al. (2011) Fast, scalable generation of highquality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539.

Stahl DA, Fishbain S, Klein M, Baker BJ, and Wagner M (2002) Origins and diversification of sulfate-respiring microorganisms. *Antonie Van Leeuwenhoek* **81**, 189–95.

Wagner M, Roger AJ, Flax JL, Brusseau GA, and Stahl DA (1998) Phylogeny of dissimilatory sulfite reductases supports an early origin of sulfate respiration. *J Bacteriol* **180**, 2975–82.

Yarza P, Richter M, Peplies J, Euzeby J, Amann R, Schleifer KH et al. (2008) The All-Species Living Tree project: a 16S rRNA-based phylogenetic tree of all sequenced type strains. *Syst Appl Microbiol* **31**, 241–50.