Supplementary File 1. Description of Supplementary Files S2-S6

Supplementary File S2

Reference *dsrAB* sequence file for classification of query sequences with mothur's implementation of RDP's Naïve Bayesian classifier (classify.seqs). This multi-fasta file contains all 1292 core nucleotide sequences from the reference database (Müller *et al.*, 2015).

Supplementary File S3

Reference *dsrAB* taxonomy file for classification of query sequences with mothur's implementation of RDP's Naïve Bayesian classifier (classify.seqs). This file contains the taxonomic information of all 1292 core sequences. First column is the identifier that matches the sequence header in Supplementary File S2. The five hierarchical taxonomic levels are separated by semicolons and are '*dsrAB* type' (for reductive bacterial-type or archaeal-type or oxidative bacterial-type *dsrAB* sequences), 'supercluster' (within the reductive bacterial-type *dsrAB* lineage), 'phylum/class' (based on the taxonomy of known isolates), 'family' (taxonomic families based on known isolates or uncultured family-level lineages), and 'strain/clone'.

Supplementary File S4

Reference DsrAB sequence alignment file for classification of query sequences with RAxML's Evolutionary Placement Algorithm (EPA). This multi-fasta file contains all aligned 1292 core amino acid sequences from the reference database (Müller *et al.*, 2015). Any query dataset has to be aligned to this reference sequence alignment, i.e. by MAFFT (Katoh *et al.*, 2002). The format of the combined output file has to be changed from ".fasta" into ".phy" and can then be directly used as input for EPA.

Supplementary File S5

Reference DsrAB tree file for placement with EPA. This file is an export of the consensus tree contained in the reference *dsrAB*/DsrAB ARB database (Müller *et al.*, 2015) in Newick format.

Supplementary File S6

ID-mapping file for classification of sequences placed into the reference tree with EPA. The first column of the file shows the ID's that are generated for the 1292 core sequences by EPA with the reference DsrAB tree (Supplementary File S5) as an input. The five hierarchical taxonomic levels are '*dsrAB* type' (for reductive bacterial-type or archaeal-type or oxidative bacterial-type *dsrAB* sequences), 'supercluster' (within the reductive bacterial-type *dsrAB* lineage), 'phylum/class' (based on the taxonomy of known isolates), 'family' (taxonomic families based on known isolates or uncultured family-level lineages), and 'strain/clone'. This file has to be compared to the EPA output file "RAxML_classification" by using e.g. R or Excel.

References

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**: 3059–3066.

Müller, A., Kjeldsen, K.U., Rattei, T., Pester, M., and Loy, A. (2015) Phylogenetic and environmental diversity of DsrAB-type dissimilatory (bi)sulfite reductases. *ISME J.* **9**: 1152–1165.