



ELSEVIER

Contents lists available at ScienceDirect

Data in brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Sequence alignments and validation of PCR primers used to detect phylogenetically diverse *nrfA* genes associated with dissimilatory nitrate reduction to ammonium (DNRA)



Jordan Cannon ^a, Robert A. Sanford ^b, Lynn Connor ^c,
Wendy H. Yang ^{a, b}, Joanne Chee-Sanford ^{c, *}

^a Dept. of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA

^b Dept. of Geology, University of Illinois at Urbana-Champaign, Urbana, IL, USA

^c USDA-ARS, Urbana, IL, USA

ARTICLE INFO

Article history:

Received 25 March 2019

Received in revised form 9 May 2019

Accepted 10 May 2019

Available online 7 June 2019

ABSTRACT

PCR primer sets were designed to target *nrfA*, the gene encoding the pentaheme nitrite reductase NrfA that catalyzes the nitrite ammonification step in the process of dissimilatory nitrate reduction to ammonium (DNRA). Details of the nucleotide alignments of the primer target regions of 271 *nrfA* sequences from reference genomes representing 18 distinct clades of NrfA are shown here along with validation of application to PCR-based methodology including the use of amplified fragment length polymorphism (AFLP) profiling and Illumina platform amplicon-based sequencing of environmental samples and selected reference strains. Summary data tables illustrate the specificity of forward primers *nrfAF2awMOD* and *nrfAF2awMODgeo* when paired with the new reverse primer *nrfAR1MOD* in relation to consensus target reference sequences associated with members of 18 NrfA clades. Specificity of the new primers to *nrfA* sequences in environmental samples is shown in AFLP analysis and amino acid-translated amplicon sequences obtained with the new primer sets. We also provide sequence alignment files of the full length *nrfA* genes, PCR reference amplicon alignment, NrfA amino-acid alignment and NrfA translated PCR amplicon-amino acid alignment. The full nucleotide and protein alignments contain 271

DOI of original article: <https://doi.org/10.1016/j.mimet.2019.03.020>.

* Corresponding author.

E-mail address: Joanne.Cheesanford@ars.usda.gov (J. Chee-Sanford).

<https://doi.org/10.1016/j.dib.2019.104016>

2352-3409/Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

reference genomes that represent the 18 identified *NrfA* clades as a tool to further aid practitioners in examining new sequences corresponding to the primer target regions and allow further primer design modifications if deemed pertinent to specific studies. A more comprehensive analysis of this data may be obtained from (“Optimization of PCR primers to detect phylogenetically diverse *nrfA* genes associated with nitrite ammonification” Cannon et al., 2019).

Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Specifications table

Subject area	Microbiology
More specific subject area	Molecular method for bacterial gene detection
Type of data	Tables, figures, FASTA files with nucleotide and amino acid sequences
How data was acquired	<i>nrfA</i> sequences downloaded from the Functional Gene Pipeline and Repository (FUNGENE) (http://fungene.cme.msu.edu/) database version 9.5, sequence alignment tools from MacVector software (v. 16.0.8), nucleotide stack displays using WebLogo (v. 2.8.2 (2005-09-08), < https://weblogo.berkeley.edu/ >), PCR, services provided by University of Illinois Carver Biotechnology Center include: fragment size analysis (AFLP) of amplicons, Fluidigm Access Array™ for multiplex amplification, Illumina sequencing platform
Data format	Raw AFLP profiles, analyzed sequence alignments, phylogenetic trees, FASTA formatted sequence files
Experimental factors	New primers targeting the <i>nrfA</i> gene were compared using PCR to old primers for performance and product quality.
Experimental features	Aligned sequences of the <i>nrfA</i> gene from 271 references were used to design new primer sets that were optimized and validated for use by PCR, AFLP, and Illumina sequencing.
Data source location	Urbana, IL
Data accessibility	Data is provided with this article
Related research article	Title: Optimization of PCR primers to detect phylogenetically diverse <i>nrfA</i> genes associated with nitrite ammonification. Author list: Cannon J, Sanford RA, Connor L, Yang WH, Chee-Sanford, J. Status: Published

Value of the data

- Provides extensive nucleotide alignments of reference *nrfA* sequences in PCR-targeted regions corresponding to highly conserved motifs that are diagnostic for the pentaheme nitrite reductase protein *NrfA*, the key enzyme in the N-cycling process dissimilatory nitrate reduction to ammonium (DNRA).
- Provides a detailed view of the sequence coverage by our primers for a majority of the known *nrfA* diversity, and the methodological approach we took to validating the primer design to demonstrate efficient amplification of *nrfA* from different environments. This type of methodology is still an indispensable tool for study of microbial communities and relies on updated primer sets for genes like *nrfA* that are harbored by highly diverse taxa.
- The availability of an extensive sequence alignment of *nrfA* will provide users with a highly useful bioinformatics tool and a starting scaffold for assessing newly obtained sequences and evaluating the effectiveness of the primers presented in this study.
- DNRA includes the key step of nitrite ammonification and is a microbial process that is more prevalent than previously thought in wide ranging terrestrial and aquatic environments. Having updated molecular tools is paramount for researchers to assess the potential for this process in mixed community systems.

1. Data

Summary percent coverage to each *NrfA* clade of the forward primers *nrfAF2awMOD* and *nrfA-F2awMODgeo* along with reverse primer *nrfAR1MOD* are shown in Table 1. WebLogos depicting the consensus of the primers aligned to reference genomes in the target regions for forward primers

nrfAF2awMOD and nrfAF2awMODgeo, and reverse primer nrfAR1MOD are shown in Fig. 1. Alignments of the primer target regions from reference genomes to primers nrfAF2awMOD and nrfAR1MOD grouped accordingly with members of each NrfA clade are shown in Fig. 2. Alignments of forward primer nrfAF2awMODgeo targeting its corresponding region in specific reference genomes that are not covered by nrfAF2awMOD are shown in Fig. 3. The primer sequence alignments to target regions of the *nrfA* genes are derived from full length and partial *nrfA* sequence alignments made available in FASTA format (Supplemental *nrfA* and *nrfA*-amplicon FASTA files: “NrfA-gene-complete-Nucleotide-alignment.fasta”; “NrfA-Gene-Amplicon-alignment.fasta”; and “NrfA-protein-alignment.fasta”). Graphic representations of AFLP data demonstrating the utility of using the individual primer pairs or multiplexed together are shown in Figs. 4 and 5 for reference genomic DNA and soil DNA, respectively. Translated amplicon sequences obtained from different soil and groundwater samples using the Fluidigm amplicon array followed by high throughput sequencing yielded sequences of the expected size range (230–300 bp) from multiple clades (Fig. 6). Selected sequences were translated and aligned to reveal common amino acids conserved among both reference amplicons and environmental DNA-derived amplicons (Fig. 7 and Supplemental file “NrfA-Environ-amplicon-Translation-AA-alignment.fasta”). The data demonstrate the utility of the new primer sets for use in the detection of *nrfA* genes and the sequence alignment data available here will provide a reference tool and starting point for data analysis by researchers conducting DNRA studies.

2. Experimental design, materials and methods

2.1. NrfA sequence selection

A previous phylogenetic analysis of 272 full-length NrfA protein sequences, based on Bayesian inference, distinguished 18 clades possessing conserved features diagnostic of pentaheme NrfA proteins [2]. The resulting final sets of new primers were ultimately tested *in silico* against a library of 271 aligned *nrfA* sequences assembled here (Figs. 1 and 2). NrfA sequences from three metagenome-assembled genomes (European Nucleotide Archive # PRJEB20068) belonging to Clades K and N and derived from the Illinois agricultural soils used in this study (described below) were included for this analysis [3].

2.2. Sequence alignment and primer design

All sequence alignments, mismatch identification, and analyses of temperature characteristics were made *in silico* using tools in MacVector software (v. 16.0.8, MacVector, Inc.). The resulting primer sequences were further analyzed for consensus alignment *in silico* against reference sequences grouped by clade membership.

2.3. Validation of primers

DNA extracts from reference strains from different NrfA clades and originating from a variety of environments were used to test new primer pair candidates. The subset of accessible reference DNA included *Serratia fonticola* strain HAC5 (Clade A) (Genbank #JX293824.1), *Shewanella oneidensis* MR-1 (Clade C), *Geobacter bemidjensis* Bem (Clade I), *Anaeromyxobacter dehalogenans* st. 2CP-1 (Clades J and K). Full *nrfA* sequences were obtained from the Functional Gene Pipeline and Repository (FUNGENE) (<http://fungene.cme.msu.edu/>) database, version 9.5 (February 2018). *S. fonticola* strain HAC-5 was previously isolated from agricultural soils and a draft genome was previously obtained (Chee-Sanford, unpublished). DNA was extracted from reference cultures and soil using a phenol: chloroform extraction method [4]. Soil extracts were modified by the addition of glycogen (20 mg/mL) to enhance the recovery of DNA during precipitation. Soil DNA samples consisted of equal volumes of DNA pooled accordingly from extracts of soil taken in April 2012 and November 2012 from depths of 0–5 cm, 5–20 cm, and 20–30 cm at agricultural sites near Havana, Illinois (HW) and Urbana, Illinois (UM). DNA from additional soil and groundwater samples used specifically for amplicon sequencing were extracted

Table 1
Percent (%) of *nrfA* sequences covered by primers to the corresponding target regions of 271 *nrfA* sequences (see Figs. 2 and 3) based on given numbers of allowable mismatches and position of mismatch.

Primer/Clade	Percent coverage of primers to target region					
nrfAF2awMOD	0 mismatches	1 mismatch	2 mismatches	1 mismatch not in the 3' end	2 mismatches not in 3' end	2 mismatches with one allowed in 3' end
A	84.0	100.0	100.0	100.0	100.0	100.0
B	100.0	100.0	100.0	100.0	100.0	100.0
C	100.0	100.0	100.0	100.0	100.0	100.0
D	45.0	50.0	50.0	50.0	50.0	50.0
E	0.0	0.0	89.5	0.0	89.5	89.5
F	62.5	75.0	87.5	75.0	87.5	87.5
G	100.0	100.0	100.0	100.0	100.0	100.0
H	55.3	60.5	100.0	60.5	100.0	100.0
I	0.0	0.0	13.0	0.0	0.0	0.0
J	100.0	100.0	100.0	100.0	100.0	100.0
K	91.3	91.3	95.7	91.3	95.7	95.7
L	21.4	71.4	100.0	71.4	100.0	100.0
M	50.0	50.0	83.3	50.0	83.3	83.3
N	45.5	45.5	54.5	45.5	54.5	54.5
O	100.0	100.0	100.0	100.0	100.0	100.0
P	100.0	100.0	100.0	100.0	100.0	100.0
Q	27.3	36.4	90.9	36.4	90.9	90.9
R	41.7	41.7	75.0	41.7	58.3	58.3
All Clades	57.9	63.8	83.8	63.8	81.9	81.9
A	0.0	0.0	64.0	0.0	0.0	0.0
B	0.0	0.0	23.5	0.0	0.0	0.0
C	0.0	0.0	50.0	0.0	0.0	0.0
D	50.0	50.0	80.0	50.0	50.0	50.0
E	0.0	0.0	0.0	0.0	0.0	0.0
F	0.0	0.0	25.0	0.0	0.0	0.0
G	0.0	0.0	50.0	0.0	0.0	0.0
H	0.0	0.0	13.2	0.0	0.0	0.0
I	82.6	87.0	100.0	82.6	95.7	100.0
J	0.0	0.0	55.6	0.0	0.0	0.0
K	4.3	4.3	82.6	4.3	4.3	4.3
L	0.0	0.0	14.3	0.0	0.0	0.0
M	0.0	16.7	33.3	16.7	16.7	16.7
N	36.4	45.5	54.5	45.5	45.5	45.5
O	0.0	0.0	28.6	0.0	0.0	0.0
P	0.0	0.0	100.0	0.0	0.0	0.0
Q	0.0	0.0	0.0	0.0	0.0	0.0
R	0.0	16.7	33.3	0.0	0.0	16.7
All Clades	12.5	14.4	44.3	13.3	14.4	15.5
A	100.0	100.0	100.0	100.0	100.0	100.0
B	94.1	100.0	100.0	100.0	100.0	100.0
C	100.0	100.0	100.0	100.0	100.0	100.0
D	25.0	25.0	85.0	25.0	85.0	85.0
E	84.2	100.0	100.0	100.0	100.0	100.0
F	100.0	100.0	100.0	100.0	100.0	100.0
G	75.0	100.0	100.0	100.0	75.0	100.0
H	94.7	94.7	100.0	94.7	94.7	100.0
I	87.0	100.0	100.0	87.0	87.0	100.0
J	77.8	100.0	100.0	77.8	77.8	100.0
K	78.3	100.0	100.0	91.3	91.3	100.0
L	85.7	100.0	100.0	92.9	92.9	100.0
M	100.0	100.0	100.0	100.0	100.0	100.0
N	100.0	100.0	100.0	100.0	100.0	100.0
O	57.1	71.4	100.0	71.4	100.0	100.0
P	100.0	100.0	100.0	100.0	100.0	100.0
Q	90.9	100.0	100.0	90.9	90.9	100.0
R	58.3	58.3	58.3	58.3	58.3	58.3
All Clades	83.0	90.4	97.0	87.1	92.6	97.0

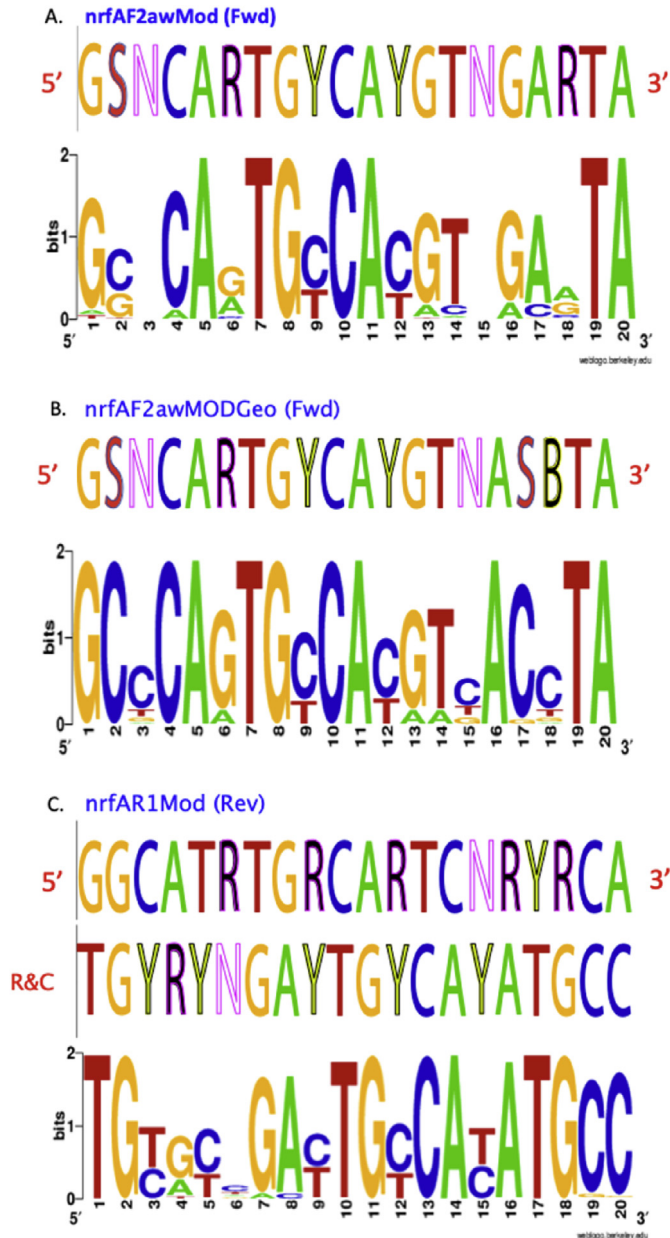


Fig. 1. WebLogos of the consensus regions depicted by stacks of nucleotides (one stack for each position in the sequence) corresponding to (A) forward primer nrfAF2awMOD sequence for the 271 reference *nrfA* gene sequences, (B) forward primer nrfAF2awMODGeo corresponding to the consensus region of Clade I; and (C) reverse primer nrfAR1Mod corresponding to the consensus of 271 *nrfA* gene sequences aligned in the primer target region. The overall height of the stack indicates the sequence conservation at that position, while the height of symbols within the stack indicates the relative frequency of each nucleic acid at that position.

Clado	Organism	nrFA2wMOD		nrFAR1MOD	
		5' G S I C A R T G C A Y G T T G A A T A		T G Y R Y I G A Y T T G C A Y A T G C C	
A	<i>Escherichia coli</i> JIS	G G T C A G T G C C A T G T T G A A T A		T G T A T C G A C T G C C A T A T G C C	
A	<i>Escherichia coli</i> TAI43 1st	G G T C A G T G C C A T G T T G A A T A		T G T A T C G A C T G C C A T A T G C C	
A	<i>Escherichia coli</i> TAI43 2nd	G G T C A G T G C C A T G T T G A A T A		T G T A T C G A C T G C C A T A T G C C	
A	<i>Escherichia alberti</i> TW07627 2	G G T C A G T G C C A C G T G G A A T A		T G C A T C G A C T G C C A C A T G C C	
A	<i>Shigella flexneri</i> 2747 71	G G T C A G T G C C A T G T T G A A T A		T G T A T C G A C T G C C A T A T G C C	
A	<i>Citrobacter koseri</i> ATCC BAA-895	G G G C A G T G C C A T G T T G A A T A		T G T A T T G C A T T G T C A T A T G C C	
A	<i>Salmonella typhimurium</i> LT2	G G T C A G T G C C A C G T G G A A T A		T G T A T C G A C T G C C A C A T G C C	
A	<i>Citrobacter youngae</i> ATCC 29220	G G T C A G T G C C A C G T G G A A T A		T G T A T C G A C T G T C A C A T G C C	
A	<i>Citrobacter rodentium</i> ICG168	G G C C A G T G C C A C G T G G A A T A		T G T A T C G A C T G C C A T A T G C C	
A	<i>Providencia alcalifaciens</i> DSM 30120	G G T C A A T G C C A C G T A G A A T A		T G T A T T G A C T G T C A T A T G C C	
A	<i>Providencia rustigiani</i> DSM 4541	G G T C A A T G C C A C G T T G A A T A		T G T A T T G A C T G C C A C A T G C C	
A	<i>Providencia rettgeri</i> DSM 1131	G G C C A A T G T C A C G T G G A A T A		T G T A T T G A C T G C C A T A T G C C	
A	<i>Providencia stuartii</i> ATCC 25827	G G G C A A T G C C A C G T T G A A T A		T G T A T C G A C T G C C A C A T G C C	
A	<i>Pectobacterium carotovorum</i> subsp. brasiliensis PBRI692	G G G C A A T G C C A C G T T G A A T A		T G T A T C G A C T G C C A T A T G C C	
A	<i>Erwinia carotovora</i> subsp. atroposeica SCR1043	G G C A A A T G C C A C G T T G A A T A		T G T A T C G A C T G C C A T A T G C C	
A	<i>Pectobacterium carotovorum</i> subsp. carotovorum PCI	G G A C A A T G C C A C G T T G A A T A		T G T A T C G A C T G C C A T A T G C C	
A	<i>Edwardsiella ictaluri</i> 93 146	G G T C A G T G C C A C G T G G A A T A		T G C G T C G A T T G C C A T A T G C C	
A	<i>Edwardsiella tarda</i> ATCC 23685	G G C C A A T G C C A C G T G G A A T A		T G C G T C G A T T G C C A C A T G C C	
A	<i>Yersinia bercovieri</i> ATCC 43970	A G T C A A T G C C A C G T T G A A T A		T G C A T C G A T T G C C A C A T G C C	
A	<i>Yersinia enterocolitica</i> subsp. palearctica Y11	A G C C A G T G C C A C G T T G A A T A		T G T A T C G A C T G C C A T A T G C C	
A	<i>Yersinia ruckeri</i> ATCC 43380	A G C C A A T G C C A T G T T G A A T A		T G C A T C G A T T G T C A T A T G C C	
A	<i>Yersinia frederiksenii</i> ATCC 33641	A G C C A A T G C C A T G T T G A A T A		T G C A T C G A T T G C C A T A T G C C	
A	<i>Aeromonas caviae</i> Ae398	G G C C A G T G C C A C G T G G A A T A		T G C G T C G A C T G C C A C A T G C C	
A	<i>Aeromonas salmonicida</i> subsp. salmonicida A449	G G C C A A T G C C A C G T T G A A T A		T G T G T C G A C T G T C A C A T G C C	
A	<i>Aeromonas veronii</i> B565	G G T C A A T G C C A C G T T G A A T A		T G T G T C G A C T G C C A C A T G C C	
B	<i>Vibrio orientalis</i> CIP 102891	G G T C A G T G T C A C G T T G A A T A		T G T G T G G A T T G T C A T A T G C C	
B	<i>Vibrio brasiliensis</i> LM0 20546	G G T C A G T G T C A C G T T G A A T A		T G T G T T G A C T G T C A T A T G C C	
B	<i>Vibrio coralliilyticus</i> ATCC BAA 450	G C C C A A T G T C A C G T T G A A T A		T G T T G G A C T G C C A T A T G C C	
B	<i>Vibrio parahaemolyticus</i> 16	G G T C A G T G C C A C G T T G A A T A		T G T G T G G A T T G T C A T A T G C C	
B	<i>Vibrio strolatus</i> DSM 21326 2	G G T C A A T G T C A T G T T G A A T A		T G T G T G A T T G T C C A C A T G C C	
B	<i>Vibrio shiloni</i> AK1 2	G G T C A G T G T C A C G T A G A A T A		T G T G T T G A C T G T C A T A T G C C	
B	<i>Vibrio mimicus</i> MB451	G G T C A G T G C C A C G T G G A A T A		T G T G C G G A T T G C C A C A T G C C	
B	<i>Aliivibrio salmonicida</i> 1F11238	G G T C A A T G T C A T G T T G A A T A		T G T T G A T T G T C A T A T G C C	
B	<i>Vibrio fischeri</i> ES114 2	G G T C A G T G C C A C G T T G A A T A		T G T G T T G A C T G T C A T A T G C C	
B	<i>Vibrio splendidus</i> 12D01 2	G G T C A G T G C C A C G T T G A A T A		T G T G C A G A C T G T C A T A T G C C	
B	<i>Vibrionales bacterium</i> SWAT3 2	G C G C A G T G C C A C G T T G A A T A		T G T G T T G A C T G T C A T A T G C C	
B	<i>Vibrio alginolyticus</i> 12G01 2	G G T C A A T G T C A C G T T G A A T A		T G T T G G A C T G C C A C A T G C C	
B	<i>Vibrio harveyi</i> ATCC BAA 1116	G G T C A A T G T C A C G T T G A A T A		T G T G T G G A C T G C C A T A T G C C	
B	<i>Vibrio parahaemolyticus</i> 10329	G G T C A G T G C C A T G T C G A A T A		T G C G T C G A T T G T C A T A T G C C	
B	<i>Vibrio vulnificus</i> CMC96	G G T C A G T G C C A C G T T G A A T A		T G T G T G G A C T G T C A C A T G C C	
B	<i>Moraxella</i> sp. PE36	G G C G A A T G C C A C G T T G A A T A		T G T G T T G A C T G T C C A T A T G C C	
B	<i>Photobacterium profundum</i> 3TCC 2	G G T C A A T G T C A C G T T G A A T A		T G T G T T G A C T G T C A C A T G C C	
C	<i>Shewanella piezotolerans</i> WP3 3	G G G C A G T G T C A C G T T G A A T A		T G T A T C G A T T G C C A T A T G C C	
C	<i>Shewanella</i> sp. ANA 3	G G T C A G T G C C A C G T T G A A T A		T G T G T T G A C T G T C A C A T G C C	
C	<i>Shewanella baltica</i> OS155	G G T C A G T G C C A C G T T G A A T A		T G T G T T G A T T G T C A C A T G C C	
C	<i>Shewanella putrefaciens</i> 200 4	G C C C A G T G C C A C G T T G A A T A		T G T G T T G A C T G C C A C A T G C C	
C	<i>Shewanella oneidensis</i> MR1	G C C C A G T G C C A C G T T G A A T A		T G C G T T G A C T G C C A T A T G C C	
C	<i>Shewanella amnensis</i> SBE2	G C C C A G T G C C A C G T T G A A T A		T G T G T G A C T G T C A C A T G C C	
C	<i>Shewanella frigidimarina</i> NCIMB 400	G G T C A A T G T C A C G T T G A A T A		T G T G T G A T T G T C A T A T G C C	
C	<i>Shewanella piezotolerans</i> WP3 6	G G T C A G T G T C A C G T T G A A T A		T G T A C T G A C T G T C A C A T G C C	
C	<i>Shewanella lothica</i> PV4 2	G G T C A G T G T C A C G T T G A A T A		T G T A C C G A C T G T C A T A T G C C	
C	<i>Shewanella sediminis</i> HAW EB3 2	G G T C A G T G C C A C G T T G A A T A		T G T A C T G A C T G T C A T A T G C C	
C	<i>Shewanella baltica</i> KT99 2	G G C C A A T G T C T G T C G A A T A		T G T A C C G A T T G T C A C A T G C C	
C	<i>Shewanella woodyi</i> ATCC 51908	G G C C A G T G T C A C G T T G A A T A		T G T A C T G A C T G T C A C A T G C C	
C	<i>Shewanella halifaxensis</i> HAW EB4	G G T C A G T G C C A C G T T G A A T A		T G T A C T G A C T G T C A C A T G C C	
C	<i>Shewanella piezotolerans</i> ATCC 700345	G G T C A G T G T C A C G T T G A A T A		T G T A C T G A C T G T C A C A T G C C	
D	<i>Shewanella piezotolerans</i> ATCC 700345 3	G G A C A G T G T C A T G T A G A A T A		T G T A C C G A T T G T C A T A T G C C	
D	<i>Shewanella halifaxensis</i> HAW EB4 4	G G G C A G T G C C A C G T T G A A T A		T G T A C C G A T T G T C A T A T G C C	
D	<i>Shewanella halifaxensis</i> HAW EB4 3	G G A C A G T G T C A C G T T G A A T A		T G C A C C G A T T G C C A T A T G C C	
D	<i>Shewanella frigidimarina</i> NCIMB 400 2	G G T C A A T G T C A T G T A G A A T A		T G T A T A C C T G T C A T A T G C C	
D	<i>Shewanella lothica</i> KT99	G G T C A A T G T C A T G T A G A A T A		T G T A T G A C T G T C A T A T G C C	
D	<i>Shewanella putrefaciens</i> 200	G G T C A G T G T C A T G T A G A A T A		T G C A T A C T T G C C A T A T G C C	
D	<i>Shewanella</i> sp. ANA 3 2	G G A C A A T G C C A C G T T A C C T A		T G C G T G A C T G C C A T A T G C C	
D	<i>Shewanella sediminis</i> HAW EB3	G G T C A A T G T C A T G T A G A A T A		T G C A T T A C C T G T C A T A T G C C	
D	<i>Shewanella halifaxensis</i> HAW EB4 2	G G T C A G T G T C A T G T A C T T A		T G T A T T A C C T G T C A T A T G C C	
D	<i>Shewanella piezotolerans</i> ATCC 700345 2	G G T C A G T G C C A T G T G A C T T A		T G T A T T A C T G T C A T A T G C C	
D	<i>Shewanella putrefaciens</i> WP3 5	G G C C A A T G T C A C G T T G A A T A		T G T G T G A C C T G C C A C A T G C C	
D	<i>Shewanella lothica</i> PV4	G G C C A G T G C C A T G T A C T T A		T G C A T C G A T T G C C A T A T G C C	
D	<i>Feromonas balnearia</i> DSM 9799 4	G G T C A G T G C C A T T A G A A T A		T G C A C C A C A T G T C A T A T G C C	
D	<i>Shewanella piezotolerans</i> WP3	G G T C A A T G C C A T T G A A T A T G A A T A		T G T A T T G A C T G T C A T A T G C C	
D	<i>Shewanella sediminis</i> HAW EB4 4	G C C C A A T G C C A C G T T G A A T A		T G T A A A A C T T G T C A C A T G C C	
D	<i>Shewanella sediminis</i> HAW EB3 5	G G T C A G T G T C A C G T T G A A T A		T G T A A A A C T T G T C A C A T G C C	
D	<i>Feromonas balnearia</i> DSM 9799	G C A C A G T G C C A C G T T G A A T A		T G T G C C A C T T G C C A C A T G C C	
D	<i>Feromonas balnearia</i> DSM 9799 2	G G C A G T G T C A C G T T G A A T A		T G T A C A C T G T C A C A T G C C	
D	<i>Feromonas balnearia</i> DSM 9799 5	G G C C A G T G C C A C G T T G A A T A		T G C A C C A T T G C C A T A T G C C	
E	<i>Aggregatibacter actinomycetemcomitans</i> D781	G C G A A C T G T C A C G T T G A A T A		T G T A T C G A T T G C C A C A T G C C	
E	<i>Aggregatibacter actinomycetemcomitans</i> D1181	G C G A A C T G T C A C G T T G A A T A		T G T A T C G A T T G C C A C A T G C C	
E	<i>Aggregatibacter aphrophilus</i> N31800	G C A A A C T G T C A C G T T G A A T A		T G T A T T G A T T G T C A C A T G C C	
E	<i>Haemophilus parainfluenzae</i> ATCC 33392	G C A A A C T G T C A C G T T G A A T A		T G T A T C G A C T G T C A C A T G C C	
E	<i>Aggregatibacter segnis</i> ATCC 33393	G C A A A C T G T C A C G T T G A A T A		T G T G T A G A T T G C C A T A T G C C	
E	<i>Haemophilus aegyptius</i> ATCC 11116	G C A A A C T G C C A T G T T G A A T A		T G T G T A G A T T G C C A T A T G C C	
E	<i>Haemophilus influenzae</i> 3655	G C A A A C T G C C A C G T T G A A T A		T G T G T A G A T T G C C A T A T G C C	
E	<i>Haemophilus influenzae</i> 22.1 21	G C A A A C T G C C A C G T T G A A T A		T G T G T A G A T T G C C A T A T G C C	
E	<i>Pasteurella multocida</i> subsp. multocida str. Pm70	G C A A A C T G C C A C G T T G A A T A		T G T A T T G C A T T G C C A T A T G C C	
E	<i>Pasteurella dagmatis</i> ATCC 43325	G C A A A C T G C C A C G T T G A A T A		T G T A T T G A C T G C C A T A T G C C	
E	<i>Mannheimia succiniciproducens</i> MB1655E	G C T A A C T G C C A C G T T G A A T A		T G T G T T G A T T G C C A C A T G C C	
E	<i>Actinobacillus succinogenes</i> 1302	G C T A A C T G C C A C G T T G A A T A		T G T G T G A C T G T C A C A T G C C	
E	<i>Actinobacillus pleuropneumoniae</i> 120	G C C A A A C T G C C A C G T T G A A T A		T G T A T T G A C T G T C A C A T G C C	
E	<i>Actinobacillus ureae</i> ATCC 25976	G C G A A C T G T C A C G T T G A A T A		T G T A C G A C T G T C A T A T G C C	
E	<i>Actinobacillus minor</i> 202	G C A A A C T G C C A C G T T G A A T A		T G T A T C G A C T G C C A T A T G C C	
E	<i>Haemophilus ducreyi</i> 35000HP	A G T A A T T G T C A C G T T G A A T A		T G T A T T G A T T G C C A T A T G C C	
E	<i>Mannheimia haemolytica</i> PH1213	G C T A A C T G C C A C G T T G A A T A		T G T A T T G A T T G C C A T A T G C C	
E	<i>Haemophilus parvus</i> 2075	T G C A A C T G C C A C G T T G A A T A		T G T A T C G A C T G T C A T A T G C C	
E	<i>Gallibacterium anatis</i> UNM179	G C T A A C T G C C A C G T T G A A T A		T G T G T T G A T T G C C A T A T G C C	
F	<i>Wolnella succinogenes</i> DSM 1740	G C C C A G T G C C A T G T G G A A T A		T G T G C G G A T T G C C A T A T G C C	
F	<i>Salinispirillum deleyanum</i> DSM 6946	G C C A A C T G T C A C G T T G A A T A		T G T G C C A T T G T C A T A T G C C	
F	<i>Danilofishia salicaria</i> DSM 2638	G C C C A G T G C C A C G T T G A A T A		T G T G C A G A C T G C C A C A T G C C	
F	<i>Danilofishia psychrophila</i> LS-54	G C C C A A T G T C A C G T T G A A T A		T G T G C C G A C T G C C A T A T G C C	
F	<i>Danilofishia propionica</i> DSM 2032	G C C C A A T G C C A C G T T G A A T A		T G C G C C G A C T G T C A C A T G C C	
F	<i>Danilofishia acetosidans</i> DSM 684 2	G G T G A G T G C C A T T C C G A A T A		T G T G C G G A T T G C C A T A T G C C	
F	<i>Arco bacter butleri</i> JV22	G C C A A A T G T C A C G T T G A A T A		T G T G C A G A T T G T C A C A T G C C	
F	<i>Arco bacter butleri</i> ED1	G C C A A A T G T C A C G T T G A A T A		T G T A C A G A T T G T C A C A T G C C	

Fig. 2. Alignment of primers nrFA2wMOD and nrFAR1MOD with the target regions of 271 *nrfa* reference sequences grouped according to clade. Shaded nucleotides match corresponding bases in the primer. Unshaded nucleotides are mismatches between primer and target site. nrFAR1MOD is depicted as the reverse and complement of the primer sequence. Note that sequences beginning with "Havana" or "Urbana" represent *nrfa* contigs recovered from metagenomic sequencing.

Clade	Organism	nrFAP2awMOD		nrFARI1MOD	
		5'	3'	5'	3'
G	<i>Marivirga tractuosa</i> DSM 4126	GCTCAATGTCATGTCAGAGTA	GCTCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
G	<i>Bizonia argentinensis</i> AFZ021000069	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
G	<i>Chlorobium phaeobacteroides</i> BSI NZ AAI01000277	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
G	<i>Leptospira araneosa</i> HTCC2155	GCTCAGTGTCACTGTCAGTA	GCTCAGTGTCACTGTCAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Prevotella oris</i> C735	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCTGATTGTCACATGCCC	TGTGCTGATTGTCACATGCCC
H	<i>Prevotella salivae</i> DSM 15606	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Prevotella buccae</i> ATCC 33574	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Prevotella</i> sp. oral taxon 299 str. F0039	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Prevotella oridis</i> ATCC 33269	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Prevotella tammerei</i> ATCC-51259	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Prevotella veroralis</i> F0319	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Prevotella melaninogenica</i> ATCC 25845	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Prevotella denticola</i> CRIS 18CA	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Prevotella multiformis</i> DSM 16608	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Prevotella disiens</i> F8035 09AN	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Prevotella buccalis</i> ATCC 35310	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Prevotella</i> sp. oral taxon 317 str. F0108	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Prevotella marshalli</i> DSM 16973	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Porphyromonas gingivalis</i> W83	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Bacteroides</i> oral taxon 274 str. F0058	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Bacteroides</i> sp. 1114	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Bacteroides</i> sp. 116	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Bacteroides plebeius</i> DSM 17135	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Bacteroides caccae</i> ATCC 43185	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Bacteroides</i> sp. 2122	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Bacteroides cellulosilyticus</i> DSM 14838	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Bacteroides</i> sp. D20	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Bacteroides fragilis</i> 3112	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Bacteroides xylanisolvens</i> X1A1	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Bacteroides</i> sp. 2133B2	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Bacteroides</i> sp. 203	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Bacteroides</i> sp. 2133B	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Dysgonomonas massii</i> DSM 22836	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Bacteroides capensis</i> DSM 18011	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Capnocytophaga</i> sp. oral taxon 329 str. F0087	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Capnocytophaga gingivalis</i> JCVIHM016	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Capnocytophaga ochracea</i> DSM 7271	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Capnocytophaga</i> sp. oral taxon 338 str. F0234	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Riemerella anatipestifer</i> DSM 15868	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
H	<i>Ignitibacterium album</i> JCM 16511	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
Close to H	<i>Opitutaceae bacterium</i> TAVS	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
Close to H	<i>Solitalea candensis</i>	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
I	<i>Geobacter bemiensis</i> Bem	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
I	<i>Geobacter</i> sp. M13	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
I	<i>Geobacter</i> sp. M18.3	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
I	<i>Geobacter lovleyi</i> SZ	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
I	<i>Geobacter</i> sp. M18.2	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
I	<i>Geobacter bemiensis</i> Bem 2	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
I	<i>Geobacter</i> sp. M21.2	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
I	<i>Pelobacter propionicus</i> DSM 2379	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
I	<i>Geobacter uranireducens</i> R4	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
I	<i>Geobacter metallireducens</i> GS15.2	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
I	<i>Geobacter sulfurreducens</i> PCA 3	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
I	<i>Desulfurimonas acetidans</i> DSM 684	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
I	<i>Geobacter lovleyi</i> SZ.2	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
I	<i>Desulfotribrio africanus</i> str. Walvis Bay	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
I	<i>Geobacter metallireducens</i> GS15	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
I	<i>Pelobacter propionicus</i> DSM 2379.2	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
I	<i>Syntrophobacter fumaroxidans</i> MPOB.2	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
I	<i>Desulfotribrio magnetus</i> RS1	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
I	<i>Desulfotribrio</i> sp. FW1012B	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
I	<i>Thermodesulfobacterium yellowstonii</i> DSM 11347	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
I	<i>Desulfotomaculum gilvum</i> DSM 7213	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
I	<i>Ammonifex degensii</i> KCA	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
I	<i>Desulfotomaculum kasnetsovii</i> DSM 6115	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
J	<i>Planctomycetes brasiliensis</i> DSM 5305	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
J	<i>Rhodospirillum rubrum</i> SH 12	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
J	<i>Thiorhodospirillum</i> sp. 970	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
J	<i>Verrucomicrobium bacterium</i> DG1235	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
J	<i>Coralimargarita akajimensis</i> DSM 45221.2	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
J	<i>Pelobacter carbinolicus</i> DSM 2380	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
J	<i>Anaeromyxobacter</i> sp. K.2	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
J	<i>Anaeromyxobacter dehalogenans</i> 2CP1.2	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
J	<i>Anaeromyxobacter dehalogenans</i> 2CPC.2	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
K	<i>Holophaga fortida</i> DSM 6591	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
K	<i>Oxylolochloris trichodes</i> DG6	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
K	<i>Opitutaceae</i> PB901.3	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
K	<i>Metothermus silvestris</i> DSM 9946	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
K	<i>Anaeromyxobacter</i> sp. K	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
K	<i>Anaeromyxobacter dehalogenans</i> 2CP1	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
K	<i>Anaeromyxobacter dehalogenans</i> 2CPC	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
K	<i>Anaeromyxobacter</i> sp. Fw109.5	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
K	<i>Anaeromyxobacter</i> sp. Fw109.5.2	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
K	<i>Planctomycetes maris</i> DSM 8797.2	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
K	<i>Idellotribrio bacteriivorus</i> HD100	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
K	<i>Mycosphaerella</i> sp. HW.1	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
K	<i>Mycosphaerella</i> sp. DK.1622	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
K	Urban deep 015 Clade K	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
K	Havana deep 038 Clade K	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
K	<i>Coralloporococcus coralloides</i> DSM 2259	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
K	<i>Sorangium cellulosum</i> 'So ce 56'	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
K	<i>Chlorobacterium flavum</i> F11428	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
K	<i>Actinomyces</i> sp. oral taxon 448 str. F0400	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
K	<i>Arcanobacterium haemolyticum</i> DSM 20595	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
K	<i>Intrasporangium calvum</i> DSM 43043	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
K	<i>Corynebacterium pseudotuberculosis</i> CIP 52.97	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC
K	<i>Corynebacterium ulcerans</i> 809	GCCCAATGTCATGTCAGAGTA	GCCCAATGTCATGTCAGAGTA	TGTGCCAGATTGTCACATGCCC	TGTGCCAGATTGTCACATGCCC

Fig. 2. (continued).

		nrAP2awMOD		nrARI1MOD	
Clade	Organism	5' G S I C A R T G Y C A Y G T I G A R T A		T G Y R Y I G A Y T G Y C A Y A T G C C :	
L	<i>Campylobacter jejuni</i> subsp. jejuni 414	T C T C A A T G C C A T G T T G A A T A		T G T G G G A T T G T C A T A T G C C	
L	<i>Campylobacter jejuni</i> RM1221	T C T C A A T G C C A T G T T G A A T A		T G C G T G G A T T G T C A T A T G C C	
L	<i>Campylobacter coli</i> RM2228	T C A C A A T G C C A T G T T G A A T A		T G T G G G A T T G T C A T A T G C C	
L	<i>Helicobacter canadensis</i> MIT 98 5491 3	T C A C A A T G T C A T G T G G A G T A		T G T G C A G A T T G C C A C A T G C C	
L	<i>Helicobacter pullorum</i> MIT 98 5489	T C T C A A T G C C A T G T T G A A T A		T G T G C A G A T T G T C A T A T G C C	
L	<i>Helicobacter cinchaoi</i> CCUG 18818	G C C C A A T G C C A C G T G G A A T A		T G T G C G G A T T G T C A T A T G C C	
L	<i>Campylobacter lari</i> RM2100	T C T C A A T G C C A T G T T G A G T A		T G T G C A G A T T G C C A C A T G C C	
L	<i>Campylobacter fetus</i> subsp. fetus 82 40	G C C A A T G T C A C G T C G A G T A		T G C G T G A T T G T C A T A T G C C	
L	<i>Campylobacter upsaliensis</i> JY21	T C A C A G T G C C A T G T G G A G T A		T G T G C G G A T T G T C A T A T G C C	
L	<i>Helicobacter mustelae</i> L2198	A T G C A A T G C C A C G T A G A G T A		T G C T C A G A T T G T C A T A T G C C	
L	<i>Campylobacter showae</i> RM3277 2	A T G C A G T G C C A C G T C G A G T A		T G C G C A G A C T G C C A C A T G C C	
L	<i>Campylobacter rectus</i> RM3267	A T G C A G T G C C A C G T A G A A T A		T G C G G G A C T G C C A C T G C C	
L	<i>Campylobacter hominis</i> ATCC BAA 381	A T G C A A T G C C A C G T A G A A T A		T G T G T T G A T T G C C A T A T G C C	
L	<i>Caldwellia aerophila</i> DSM 14535	G C C C A A T G C C A C G T G G A G T A		T G T G C C A C T G C C A C A T G C C	
M	<i>Desulfotobacterium dehalogenans</i> ATCC 51507	G C C C A G T G C C A T G T G G A A T A		T G C G C T G A C T G C C A T A T G C C	
M	<i>Desulfotobacterium hafnense</i> DCB 2 2	G C C C A G T G T C A C G T A A C T T A		T G T G C G A C T G T C A T A T G C C	
M	<i>Desulfosporosinus merdii</i> DSM 13257	G C C A A C T G C C A T G T G G A G T A		T G T G C G G A T T G C C A C A T G C C	
M	<i>Anaerolinea thermophila</i> UN1 3	G C C A A C T G C C A T G T G G A G T A		T G C G C G A C T G C C A C A T G C C	
M	<i>Anaerolinea thermophila</i> UN1 1	G C C A A C T G C C A T G T G G A G T A		T G C G C G A C T G C C A C A T G C C	
M	<i>Dehalobacter alkaliphilus</i> AHT 1 5th	G C A C A A T G C C A C A T A A C G T A		T G T G C C A C T G T C A T A T G C C	
N	<i>Dehalobacter alkaliphilus</i> AHT 1 5	G C A C A A T G C C A C A T A A C G T A		T G T G C C A C T G T C A T A T G C C	
N	<i>Carboxydobacterium hydrogeniformans</i> Z2901 2	G C C C A G T G C C A C G T G G A A T A		T G C G T T G A T T G T C A T A T G C C	
N	<i>Symbiodinium thermophilum</i> IAM 14863	G G T C A G T G C C A C G T G G A A T A		T G T G C C A G A T T G C C A C A T G C C	
N	<i>Carboxydobacterium hydrogeniformans</i> Z2901	G C T C A A T G C C A T G T T G A A T A		T G C G C A G A T T G T C A T A T G C C	
N	<i>Desulfotobacterium dichloroethinans</i> LMGP P 21439	G C C A A C T G C C A T G T G G A A T A		T G C G C G G A T T G C C A C A T G C C	
N	Havam deep 073 Clade MN	G C C A A T G C C A T G T T G A A T A		T G T G C A G A T T G C C A T A T G C C	
N	<i>Desulfotobacterium hafnense</i> DCB2	G C C C A G T G C C A C G T G G A A T A		T G C G C G A C T G C C A C A T G C C	
N	<i>Desulfotobacterium modesticaldum</i> Ioe1	G C C C A G T G C C A C G T G G A A T A		T G C G C G A C T G C C A C A T G C C	
N	<i>Desulfotomaculum rediens</i> MI1	G C C C A G T G C C A C G T A A C C T A		T G T G C T G A T T G C C A T A T G C C	
N	<i>Desulfotomaculum carboxydovorans</i> CO1 SRB	G C C C A G T G C C A C G T A A C C T A		T G C G C T G A T T G C C A T A T G C C	
N	<i>Syntrophomonas wolfei</i> subsp. wolfei str. Goettingen	G C C C A G T G C C A C G T A A C C T A		T G C G C T G A T T G C C A T A T G C C	
O	<i>Desulfovibrio</i> sp. 3 1 syn3	G C C C A G T G C C A T G T G G A A T A		T G C G C G G A T T G C C A C A T G C C	
O	<i>Desulfovibrio desulfuricans</i> subsp. desulfuricans str. ATCC 27774	G C C C A G T G C C A C G T G G A A T A		T G C G C G G A C T G C C A C A T G C C	
O	<i>Desulfovibrio piger</i> ATCC 29098	G C C C A G T G C C A T G T G G A A T A		T G C G C G G A C T G C C A C A T G C C	
O	<i>Lawsonia intracellulare</i> FITE/UN1 00	G C A C A A T G T C A C G T A G A A T A		T G T G C T G A T T G T C A T A T G C C	
O	<i>Blasphila wadsworthia</i> 3 1 6	G C C C A G T G C C A T G T G G A A T A		T G T G C C A C T G C C A C A T G C C	
O	<i>Desulfovibrio vulgaris</i> subsp. vulgaris str. Hildenborough 2	G G T C A G T G C C A C G T G G A G T A		T G C G C G A C T G C C A C A T G C C	
O	<i>Desulfovibrio vulgaris</i> str. 'Miyazaki' F	G G C C A G T G C C A C G T G G A A T A		T G C G C C G A C T G C C A C A T G G	
O	<i>Desulfovibrio vulgaris</i> str. 'Miyazaki' F 2	G G C C A G T G C C A C G T G G A A T A		T G C G C C G A C T G C C A C A T G C	
O	<i>Desulfotomaculum baculatum</i> DSM 4028 3	G G C C A G T G C C A C G T G G A A T A		T G C G C G A C T G C C A C A T G A G	
O	<i>Desulfovibrio isopropanolius</i> Asp2 2	G G C C A G T G C C A C G T T G A A T A		T G T G C C G A C T G C C A C A T G C C	
O	<i>Desulfurespirillum indicum</i> S5	G C C C A G T G C C A C G T T G A A T A		T G T G C C G A C T G C C A C A T G C C	
O	<i>Halotheliospira halophila</i> SL1	G C C C A A T G C C A C G T G A G T A		T G C G C T G A C T G C C A C A T G C C	
O	<i>Desulfonatronospira thiodismans</i> ASO3 1	G C C C A G T G T C A T G T A G A A T A		T G T G C A G A C T G C C A T A T G C C	
P	<i>Bacillus selenitireducens</i> MLS10 2	G C G C A A T G T C A C G T G G A G T A		T G C G C C G A T T G T C A T A T G C C	
P	<i>Bacillus selenitireducens</i> MLS10	G G G C A A T G T C A T G T T G A G T A		T G T G C T G A T T G T C A T A T G C C	
P	<i>Bacillus</i> sp. INLA3E	G G G C A A T G T C A T G T T G A G T A		T G T G C T G A T T G T C A T A T G C C	
Q	<i>Selenomonas noxia</i> F0398	T G G C C A A T G C C A T A C G G A G T A		T G C G T G G A C T G C C A C A T G C C	
Q	<i>Selenomonas noxia</i> ATCC 43541	T C G C A G T G C C A T A C A G A G T A		T G C T G G A C T G C C A C A T G C C	
Q	<i>Selenomonas fluzgei</i> ATCC 43531	G C C A G T G C C A T A C G G A G T A		T G C G T G G A C T G C C A C A T G C C	
Q	<i>Selenomonas artemidis</i> F0399	G G A C A G T G C C A T A C G G A G T A		T G C G T G G A C T G C C A C A T G C C	
Q	<i>Mitsunella nitida</i> DSM 20544	G C A C A G T G C C A C A A G G A G T A		T G C G T T G A C T G C C A T A T G C C	
Q	<i>Selenomonas spuitgens</i> ATCC 35185 2	G G C C A G T G C C A T A C G G A G T A		T G C A T C G A C T G C C A C A T G C C	
Q	<i>Selenomonas spuitgens</i> ATCC 35185	G G C C A G T G C C A T A C G G A G T A		T G C A T C G A C T G C C A C A T G C C	
Q	<i>Thermosinus carboxydovorans</i> Nor1	G G C C A A T G T C A T C C G A G T A		T G C G C G A C T G C C A T A T G C C	
Q	<i>Syntrophus aciditrophicus</i> SB	G G C C A G T G C C A T G T G G A A T A		T G C G C G A C T G C C A C A T G C C	
Q	<i>Acetomaculum longum</i> DSM 6540	G C C C A A T G T C A T G T G A A T A		T G C G C G A C T G C C A C A T G C C	
Q	<i>Thermicella</i> sp. JR	G G A C A G T G C C A T G T A G A A T A		T G T T T G A C T G C C A C A T G C C	
R	<i>Anaerococcus prevotii</i> DSM 20548	G C C C A A T G T C A C G T T G A A T A		T G T A T A G A C T G T C A C A T G C C	
R	<i>Anaerococcus prevotii</i> ACS 065 V Col13	G C C C A A T G T C A C G T T G A A T A		T G T A T A G A T T G T C A T A T G C C	
R	<i>Anaerococcus lactolyticus</i> ATCC 51172	G C C A A T G C C A C G T T G A A T A		T G T A T A G A T T G C C A C A T G C C	
R	<i>Alkaliphilus oremlandii</i> OhlAs	G C C A G T G T C A C G T T A G A A T A		T G T G C C G A T T G T C A T A T G C C	
R	<i>Alkaliphilus metalliredigens</i> QYMF	G C C C A A T G T C A T G T T G A A T A		T G T A T A G A C T G C C A T A T G C C	
R	<i>Gordibacter penicillatus</i> 7101 b	G C C C A G T G C C A C A A C G A G T A		T T C G A G A A C G A A A G T O H T G	
R	<i>Eggerthella lenta</i> DSM 2243	G G C C A G T G C C A C A A C G A G T A		C A C G T G C - C A T A A C T G G G T	
R	<i>Cryptobacterium curtum</i> DSM 15641	G G T C A G T G C C A T T G C G A C T A		T A T A C A A G C C A T T A T T G G G G	
R	<i>Slackia exigua</i> ATCC 700122	G G C C A G T G C C A C T G C G A C T A		T A C A C A G C C A C T A C T G G G C A	
R	<i>Slackia helioiridireducens</i> DSM 20476 2	G G C C A G T G C C A C T G C G A C T A		T A C A C A A T C A C C A G T G G A C	
R	<i>Dehalobacter alkaliphilus</i> AHT 1	G C C C A G T G C C A C G T A A C T A		T G T A C C A G T T G T C A T A T G C C	
R	<i>Dehalobacter alkaliphilus</i> AHT 1 3	G C T C A G T G C C A C G T A A C T A		T G T G C C A T T G C C A T A T G C C	
Other	<i>Campylobacter rectus</i> RM3267 2			T G T G C G G A T T G C C A C T T G C C	
Other	<i>Thiosalkalibacterium nitratireducens</i> ONR			T G C G C C G A C T G C C A C A T G C C	

Fig. 2. (continued).

using an abbreviated phenol:chloroform protocol [5] and then followed by glycogen-enhanced recovery as described above. Final DNA concentrations (~8–10 ng/μL) were measured using Qubit 2.0 fluorometry (Invitrogen) and DNA band intensities estimated against quantitative DNA ladders following gel electrophoresis.

2.4. Optimization of PCR

All primers were HPLC-purified and obtained from IDT (Integrated DNA Technologies, Skokie, IL, USA). Stock concentrations (100 μM) of each primer were made by adding Invitrogen™ UltraPure™

		nrfAF2awMODgeo																			
		5' G S I C A R T G Y C A Y G T I A S B T A																			
Clade	Organism	G	G	T	C	A	A	T	G	T	C	A	T	G	T	G	A	G	T	T	A
D	<i>Shewanella frigidimarina</i> NCIMB 400 2	G	G	T	C	A	A	T	G	T	C	A	T	G	T	G	A	G	T	T	A
D	<i>Shewanella benthica</i> KT99	G	G	T	C	A	A	T	G	T	C	A	T	G	T	G	A	G	C	T	A
D	<i>Shewanella putrefaciens</i> 200	G	G	T	C	A	G	T	G	T	C	A	T	G	T	G	A	C	T	T	A
D	<i>Shewanella</i> sp. ANA3 2	G	G	A	C	A	A	T	G	C	C	A	C	G	T	T	A	C	C	T	A
D	<i>Shewanella sediminis</i> HAW EB3	G	G	T	C	A	A	T	G	T	C	A	T	G	T	G	A	G	C	T	A
D	<i>Shewanella halifaxensis</i> HAW EB4 2	G	G	T	C	A	G	T	G	T	C	A	T	G	T	G	A	C	T	T	A
D	<i>Shewanella pealeana</i> ATCC 700345 2	G	G	T	C	A	G	T	G	T	C	A	T	G	T	G	A	C	T	T	A
D	<i>Shewanella piezotolerans</i> WP3 5	G	G	C	C	A	A	T	G	T	C	A	T	G	T	C	A	C	T	T	A
D	<i>Shewanella loihica</i> PV4	G	G	C	C	A	G	T	G	C	C	A	T	G	T	G	A	C	C	T	A
D	<i>Ferrimonas balearica</i> DSM 9799 4	G	G	C	C	A	G	T	G	C	C	A	T	G	T	G	A	C	T	T	A
I	<i>Geobacter bemidjensis</i> Bem	G	C	C	C	A	G	T	G	C	C	A	C	G	T	T	A	C	C	T	A
I	<i>Geobacter</i> sp. M21 3	G	C	C	C	A	G	T	G	C	C	A	C	G	T	G	A	C	C	T	A
I	<i>Geobacter</i> sp. M18 3	G	C	C	C	A	G	T	G	T	C	A	C	G	T	C	A	C	C	T	A
I	<i>Geobacter lovleyi</i> SZ	G	C	C	C	A	G	T	G	C	C	A	T	G	T	C	A	G	C	T	A
I	<i>Geobacter</i> sp. M18 2	G	C	C	C	A	G	T	G	C	C	A	C	G	T	C	A	C	C	T	A
I	<i>Geobacter bemidjensis</i> Bem 2	G	C	G	C	A	A	T	G	C	C	A	C	G	T	T	A	C	C	T	A
I	<i>Geobacter</i> sp. M21 2	G	C	G	C	A	A	T	G	C	C	A	C	G	T	C	A	C	C	T	A
I	<i>Pelobacter propionicus</i> DSM 2379	G	C	T	C	A	G	T	G	T	C	A	C	G	T	T	A	C	C	T	A
I	<i>Geobacter uraniumreducens</i> Rf4	G	C	C	C	A	G	T	G	T	C	A	T	G	T	C	A	C	G	T	A
I	<i>Geobacter metallireducens</i> GS15 2	G	C	C	C	A	G	T	G	C	C	A	C	G	T	C	A	C	C	T	A
I	<i>Geobacter sulfurireducens</i> PCA 3	G	C	C	C	A	G	T	G	C	C	A	C	G	T	C	A	C	C	T	A
I	<i>Desulfuromonas acetoxidans</i> DSM 684	G	C	A	C	A	G	T	G	T	C	A	C	G	T	C	A	C	C	T	A
I	<i>Geobacter lovleyi</i> SZ 2	G	C	C	C	A	G	T	G	C	C	A	C	G	T	G	A	C	C	T	A
I	<i>Desulfovibrio africanus</i> str. Walvis Bay	G	C	C	C	A	G	T	G	C	C	A	C	G	T	G	A	C	T	T	A
I	<i>Geobacter metallireducens</i> GS15	G	C	C	C	A	G	T	G	C	C	A	C	G	T	C	A	C	C	T	A
I	<i>Pelobacter propionicus</i> DSM 2379 2	G	C	C	C	A	G	T	G	C	C	A	T	G	T	C	A	C	C	T	A
I	<i>Syntrophobacter fumaroxidans</i> MPOB 2	G	C	C	C	A	G	T	G	C	C	A	C	G	T	C	A	C	T	T	A
I	<i>Desulfovibrio magneticus</i> RS1	G	C	T	C	A	G	T	G	C	C	A	C	G	T	G	A	C	G	T	A
I	<i>Desulfovibrio</i> sp. FW1012B	G	C	C	C	A	G	T	G	C	C	A	C	G	T	C	A	C	C	T	A
I	<i>Thermodesulfovibrio yellowstonii</i> DSM 11347	G	C	C	C	A	A	T	G	T	C	A	T	G	T	T	A	C	A	T	A
I	<i>Desulfotomaculum gibsoniae</i> DSM 7213	G	C	C	C	A	G	T	G	T	C	A	T	A	A	T	A	C	T	T	A
I	<i>Ammonifex degensii</i> KC4	G	C	T	C	A	G	T	G	C	C	A	T	A	A	C	A	C	C	T	A
I	<i>Desulfotomaculum kuznetsovii</i> DSM 6115	G	C	C	C	A	G	T	G	C	C	A	T	A	A	C	A	C	C	T	A
K	Havana deep 038 Clade_K	G	G	C	C	A	G	T	G	C	C	A	T	G	T	G	A	C	C	T	A
M	<i>Dethiobacter alkaliphilus</i> AHT 1 5th	G	C	A	C	A	A	T	G	C	C	A	C	A	T	A	A	C	G	T	A
N	<i>Dethiobacter alkaliphilus</i> AHT1 5	G	C	A	C	A	A	T	G	C	C	A	C	A	T	A	A	C	G	T	A
N	<i>Carboxydotherrmus hydrogenoformans</i> Z2901 2	G	C	C	C	A	G	T	G	T	C	A	T	G	T	A	A	C	T	T	A
N	<i>Desulfotomaculum reducens</i> MI1	G	C	C	C	A	G	T	G	T	C	A	T	G	T	T	A	C	C	T	A
N	<i>Desulfotomaculum carboxydivorans</i> CO1 SRB	G	C	C	C	A	G	T	G	T	C	A	C	G	T	A	A	C	C	T	A
N	<i>Syntrophomonas wolfei</i> subsp. wolfei str. Goettingen	G	C	G	C	A	A	T	G	T	C	A	T	G	T	T	A	C	T	T	A
R	<i>Dethiobacter alkaliphilus</i> AHT 1	G	C	C	C	A	G	T	G	C	C	A	C	G	T	A	A	C	T	A	
R	<i>Dethiobacter alkaliphilus</i> AHT 1 3	G	C	T	C	A	G	T	G	C	C	A	C	G	T	C	A	A	C	T	A

Fig. 3. Alignment of primer nrfAF2awMODgeo with the target regions of *nrfA* reference sequences of members within the corresponding clades that match the primer. Shaded nucleotides match corresponding bases in the primer. Unshaded nucleotides are mismatches between primer and target site.

DNase/RNase-Free Distilled Water (Thermo Fisher Scientific Waltham, MA, USA) and subsequently diluted for use in PCR. PCR reactions were performed in 25 μ L volumes using the Takara ExTaq PCR kit (Clontech) and a MJ Research PTC-200 Gradient Thermal Cycler. The optimized reaction mixture was the following: 1X PCR buffer, 0.2 mM each deoxynucleoside triphosphate (dNTPs), 0.025U/ μ L TaKaRa ExTaq DNA polymerase, 3.2 μ M each forward and reverse primers, and ~1 ng reference template DNA. Thermocycling conditions were the following: initial denaturation step at 95 °C for 5 min, followed by 25 or 30 cycles of [95 °C for 30 sec, 56 °C for 30 sec, and 72 °C for 30 sec] and a final extension of 72 °C

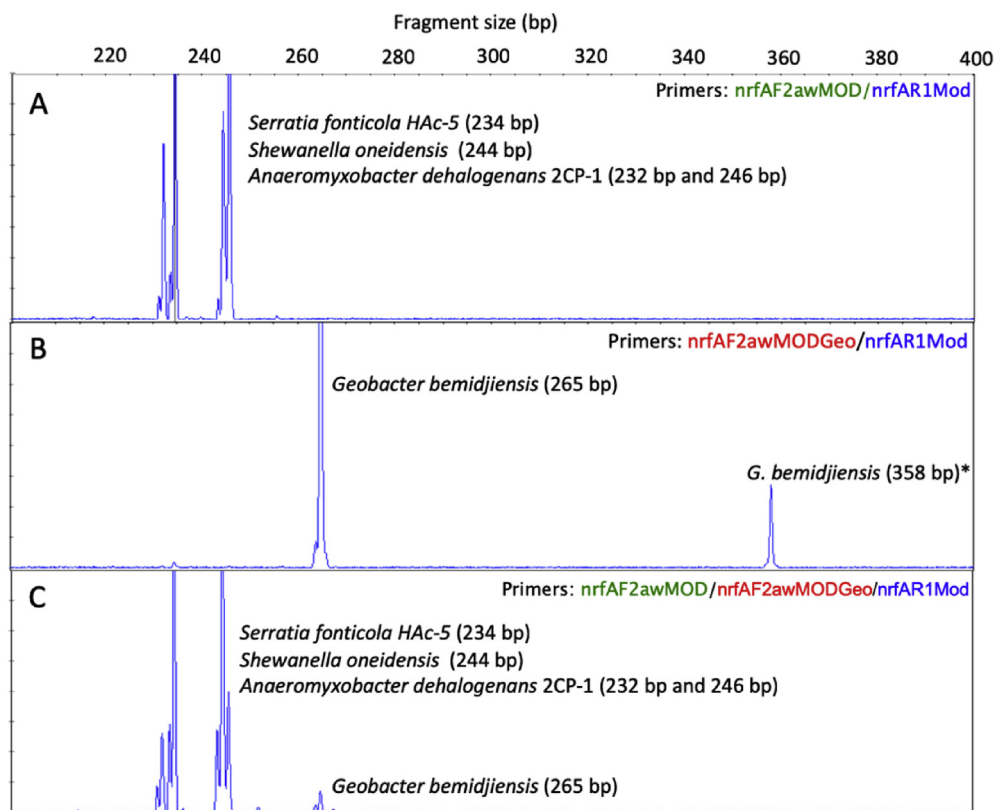


Fig. 4. Amplified fragment length polymorphism (AFLP) profiles resulting from PCR amplification of *nrfA* using a DNA pool of equivalent masses from reference organisms *A. dehalogenans* 2CP-1, *S. fonticola* HAC-5, *S. oneidensis* MR-1, and *G. bemidjiensis* using primer sets (A) 6-FAM-*nrfAF2awMOD*/*nrfAR1Mod*, (B) 6-FAM-*nrfAF2awMODGeo*/*nrfAR1Mod*, and (C) 6-FAM-*nrfAF2awMOD*/*nrfAF2awMODGeo*/*nrfAR1Mod*. Fragment sizes measured are indicated in parentheses and are consistently 1–4 bp smaller than the expected sizes (See Table 2 in [1]) due to migration characteristics during column separation. *Cross-specificity of the primer set to another heme-binding sequence homolog that is not *nrfA* from *G. bemidjiensis* yields an additional product.

for 10 min. DNA from four *nrfA* containing organisms served as positive controls to test the efficacy of the primers (Fig. 4). PCR products were resolved by gel electrophoresis using 2.5% High Resolution Agarose (fragments < 1kb) (Gold Biotechnology, Olivette, MO, USA) in 1X TBE buffer on a HU13 Midi horizontal gel unit (Scie-plas Ltd., Cambridge, UK) at 4 V/cm for 80 minutes. DNA ladders consisted of 1 μ L of Low Molecular Weight DNA Ladder (New England Biolabs Inc., Ipswich, MA, USA) and 5 μ L of Quick-Load Purple 2-Log DNA Ladder (0.1–10.0 kb).

2.5. Amplified fragment length polymorphism (AFLP) analysis

Amplified fragment length polymorphism (AFLP) analysis was used to assess the amplification efficiencies from a pool of different reference *nrfA* and to further corroborate the specificity of the forward primers *nrfAF2awMOD* and *nrfAF2awMODGeo* when paired with the reverse primer *nrfAR1MOD*. AFLP analysis was performed on amplicons generated from a mixed DNA pool (1 ng each) of reference DNA from *S. fonticola* HAC-5, *S. oneidensis* MR-1, *A. dehalogenans* 2CP-1, and *G. bemidjiensis* Bem (Fig. 4). A combined pool of both forward primers with the reverse primer was also tested against the same reference DNA to assess any inhibition that could result from competing reactions. The primer pair combinations used were 5'-(6-FAM)-*nrfAF2awMOD*/*nrfAR1Mod*, 5'-(6-FAM)-

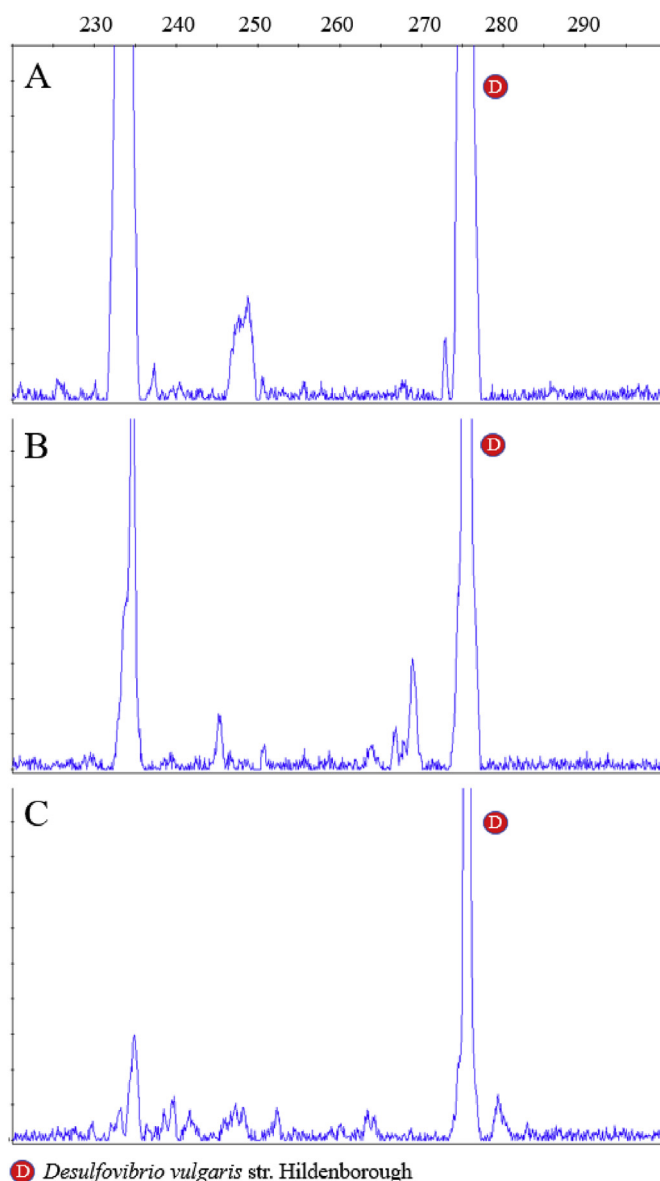


Fig. 5. AFLP results for soil sample HW1 11/12 using (A) primer set 6-FAM-nrfAF2awMOD/nrfAR1MOD and (B) primer set 6-FAM-nrfAF2awMODgeo/nrfAR1MOD and (C) AFLP results for soil sample UM2 4/2012 using 6-FAM-nrfAF2awMODgeo/nrfAR1MOD. Soil DNA samples were spiked with 0.5 ng of *Desulfovibrio vulgaris* strain Hildenborough DNA prior to PCR reactions to serve as an internal reference and positive control. Note that some amplification of the internal reference DNA spike even occurs with the Clade I-specific forward primer: nrfAF2awMODgeo.

nrfAF2awMODgeo/nrfAR1Mod, and combined forward primers 5'-(6-FAM)-nrfAF2awMOD+5'-(6-FAM)-nrfAF2awMODgeo/nrfAR1Mod. All PCR products were diluted 50-fold with ultrapure water before submitting for fragment size analysis (Roy J. Carver Biotechnology Center, University of Illinois, Urbana, IL). Fragments were sized following calibration against a MapMarker 1000 size standard and expected product sizes were accounted for in the resulting profiles. To test the application of AFLP to an

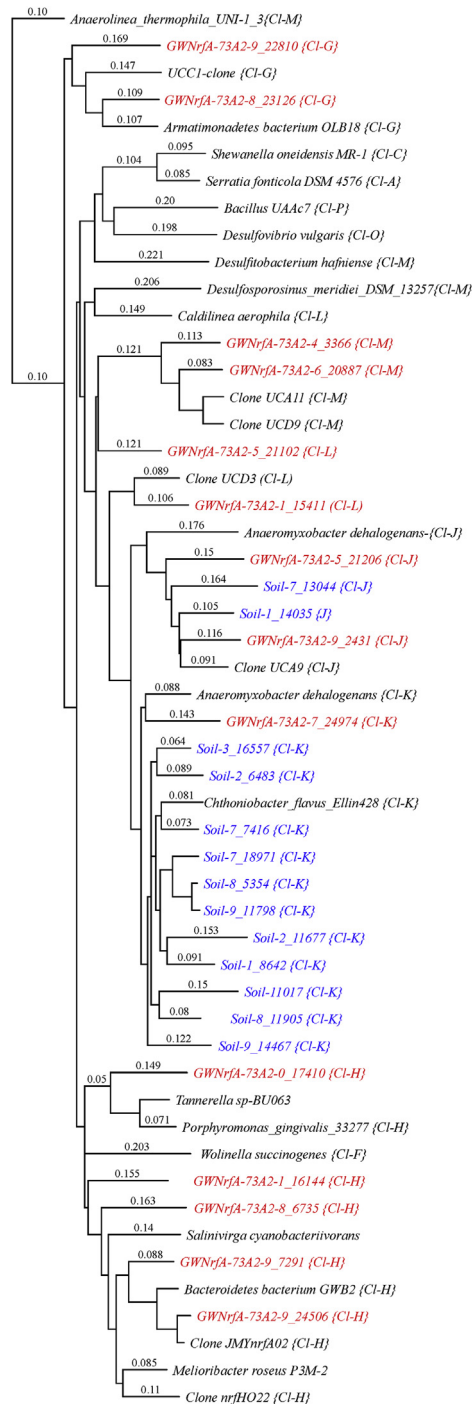


Fig. 6. Phylogenetic tree of aligned amino-acid sequences associated with translated *nrfA* gene amplicons generated from groundwater (red) and soil (blue) DNA in relation to known *NrfA* references. Clade designations are shown in parentheses. Branches labeled with "Clone" indicate *nrfA* gene sequences from soil DNA amplified in previous studies. An amino acid alignment of specific sequences representing different Clades

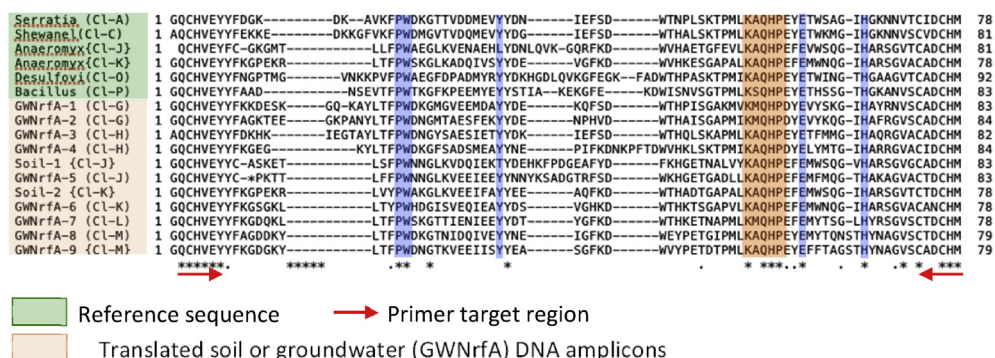


Fig. 7. Aligned amino acids from reference NrfA and representative translated *nrfA* gene amplicons from soil (Soil-) and groundwater (GWNrfA-) DNA using the new primer pair design. Highlighted regions in sequence indicate expected conserved residues associated specifically with NrfA as identified by [2]. Clade designations are shown in parentheses.

environmental sample, soil DNA were also amplified for 30 cycles using ~8–10 ng of DNA in individual reactions and PCR reaction mixtures were modified with the addition of 25 µg/mL T4 gene 32 protein (Roche Applied Science, Indianapolis, IN, USA) and spiked with 1 ng DNA from *D. vulgaris* as an internal standard (Fig. 5).

2.6. Fluidigm array and amplicon-based sequencing

To verify that the designed primers yielded actual *nrfA* gene fragments, we included the redesigned primer pair in amplicon sequencing analysis of soil, groundwater, and reference genomic DNA pools. The reference pool consisted of equal masses of genomic DNA from the known DNRA taxa *Desulfovibrio vulgaris* st. Hildenborough, *Anaeromyxobacter dehalogenans* st. 2CP-1, *Shewanella oneidensis* st. MR-1, *Geobacter bemidjensis* st. Bem, *Serratia fonticola* st. HAC5, and *Bacillus* sp. UAAc-7. Using the Fluidigm Access Array at the University of Illinois Carver Biotechnology Center, DNA from different samples (up to 48) were amplified using up to 48 primer pairs, one of which included primers *nrfAF2awMOD* and *nrfAR1MOD*. The *nrfA* primers were one set of 14 gene-specific primer sets evaluated in the Fluidigm array, allowing both an assessment of their application in multiplex PCR technology and to address the efficacy of the primer set to detect *nrfA* genes in different environmental samples. The other data from the other 13 gene amplicon sequences collected from this array was not relevant to this paper. A standard annealing temperature of 55 °C was used during PCR amplification to generate a pool of amplicons. These Fluidigm generated pooled amplicons from all PCR reactions were purified using a Qiagen™ QIAquick Gel Extraction Kit (Qiagen™, Valencia, CA, United States) according to the manufacturer's instructions. The DNA from the entire Fluidigm array was quantified and sequenced on one MiSeq flowcell for 301 cycles from each end of the fragments using a MiSeq 600-cycle sequencing kit version 3. Fastq files were generated and demultiplexed with the bcl2fastq v2.20 Conversion Software (Illumina). PhiX DNA was used as a spike-in control and removed in the data processing. Read lengths were 300 nucleotides. The raw data was sorted by the PCR-specific primers and paired end reads were obtained and demultiplexed by sample index.

Sequence data was selectively processed only for the *nrfA* gene amplicons in the reference genomic sample (no amplicons were obtained for *G. bemidjensis*), one soil DNA sample and one groundwater DNA sample. Briefly, paired end reads were stitched together and filtered to the expected amplicon length using *mothur* [6]. The resulting fasta files for each sample were shortened to a maximum of 1000 sequences and aligned using *MacVector* software. Any sequences outside the forward and reverse

from soil and groundwater used in this study are shown in Fig. 7. Summary alignment of the amino acid sequences represented in this Figure is included in supplemental material as a FASTA alignment file "*NrfA-Environ-amplicon-Translation-AA-alignment.fasta*".

primer target regions were trimmed manually. Representative OTUs of clearly different taxa were selectively translated using MacVector starting with the 5' end of the forward primer which is known to be in-frame. The resulting amino-acid sequences were separately aligned using MacVector to evaluate the predicted protein fragment for diagnostic residues expected in NrfA between the third and fourth heme-binding domains [2].

Acknowledgements

This work was funded in part by the U.S. National Science Foundation grant DEB-1656027 to WHY and RS, and by the U.S. Department of Agriculture NIFA grant 2016-67030-25211 to WHY. We wish to thank Alex Krichels for soil DNA samples used for sequencing analysis. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.dib.2019.104016>.

References

- [1] J. Cannon, R.A. Sanford, L. Connor, W.H. Yang, J. Chee-Sanford, Optimization of PCR primers to detect phylogenetically diverse *nrfA* genes associated with nitrite ammonification, *J. Microbiol. Methods* 160 (2019) 49–59. <https://doi.org/10.1016/j.mimet.2019.03.020>.
- [2] A. Welsh, J.C. Chee-Sanford, L.M. Connor, F.E. Löffler, R.A. Sanford, Refined NrfA phylogeny improves PCR-based *nrfA* gene detection, *Appl. Environ. Microbiol.* 80 (2014) 2110–2119. <https://doi.org/10.1128/AEM.03443-13>.
- [3] L.H. Orellana, J.C. Chee-Sanford, R.A. Sanford, F.E. Löffler, K.T. Konstantinidis, Year-round shotgun metagenomes reveal stable microbial communities in agricultural soils and novel ammonia oxidizers responding to fertilization, *Appl. Environ. Microbiol.* 84 (2018) 1–14. <https://doi.org/10.1128/aem.01646-17>.
- [4] Y. Tsai, B. Olson, Rapid method for direct extraction of DNA from soil and sediments, *Appl. Environ. Microbiol.* 57 (1991) 1070–1074.
- [5] R.I. Griffiths, A.S. Whiteley, A.G. O'Donnell, M.J. Bailey, Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition, *Appl. Environ. Microbiol.* 66 (2000) 5488–5491.
- [6] P.D. Schloss, S.L. Westcott, T. Ryabin, J.R. Hall, M. Hartmann, E.B. Hollister, et al., Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities, *Appl. Environ. Microbiol.* 75 (2009) 7537–7541. <https://doi.org/10.1128/aem.01541-09>.