UC Merced
UC Merced Previously Published Works

Title

High-quality draft genome sequence of Sedimenticola selenatireducens strain AK4OH1T, a gammaproteobacterium isolated from estuarine sediment

Permalink

https://escholarship.org/uc/item/6s23t6p5

Journal

Environmental Microbiome, 11(1)

ISSN

1944-3277

Authors

Louie, Tiffany S
Giovannelli, Donato
Yee, Nathan
et al.

Publication Date

2016

DOI

10.1186/s40793-016-0191-5

Peer reviewed

Standards in
Genomic Sciences

SHORT GENOME REPORT

Open Access

CrossMark

High-quality draft genome sequence of *Sedimenticola selenatireducens* strain AK4OH1[T], a gammaproteobacterium isolated from estuarine sediment

Tiffany S. Louie[1], Donato Giovannelli[2,3,4], Nathan Yee[5], Priya Narasingarao[1], Valentin Starovoytov[6], Markus Göker[7], Hans-Peter Klenk[7,8], Elke Lang[7], Nikos C. Kyrpides[9,10], Tanja Woyke[9], Elisabetta Bini[11,12] and Max M. Häggblom[1*]

Abstract

Sedimenticola selenatireducens strain AK4OH1[T] (= DSM 17993[T] = ATCC BAA-1233[T]) is a microaerophilic bacterium isolated from sediment from the Arthur Kill intertidal strait between New Jersey and Staten Island, NY. *S. selenatireducens* is Gram-negative and belongs to the *Gammaproteobacteria*. Strain AK4OH1[T] was the first representative of its genus to be isolated for its unique coupling of the oxidation of aromatic acids to the respiration of selenate. It is a versatile heterotroph and can use a variety of carbon compounds, but can also grow lithoautotrophically under hypoxic and anaerobic conditions. The draft genome comprises 4,588,530 bp and 4276 predicted protein-coding genes including genes for the anaerobic degradation of 4-hydroxybenzoate and benzoate. Here we report the main features of the genome of *S. selenatireducens* strain AK4OH1[T].

Keywords: *Sedimenticola selenatireducens*, *Gammaproteobacteria*, Anaerobe, Selenate respiration, 4-hydroxybenzoate

Introduction

Selenium (Se) is an intriguing element in that microbes actively metabolize it through reduction, oxidation, methylation and demethylation reactions, using some of these to conserve energy. Of particular interest is the process of dissimilatory Se reduction, where the Se oxyanion, selenate [Se(VI)], is sequentially reduced to selenite [Se(IV)] and further to insoluble elemental Se(0). The ability to respire selenate/selenite is comparatively rare, nonetheless, is found in phylogenetically diverse anaerobes [1]. SeRB display a tremendous phylogenetic diversity, and yet the metabolic function seems to be conserved (or alternatively horizontally dispersed) in these unrelated groups. Furthermore, the physiologies of the known selenate-respiring bacteria appear to vary greatly. For example, they are able to couple growth to a wide range of electron acceptors such as arsenate, [2, 3] cobalt oxide (Co(III)) [4], and tellurite [5] to name a few. SeRB have been isolated from a variety of different locations. A few examples are: in California in the San Joaquin Valley [6], from estuarine sediment in NJ [7], from a glass manufacturing plant in Japan [8], and from the dead sea [9].

Sedimenticola selenatireducens type strain AK4OH1[T] (= DSM 17993[T] = ATCC BA-1233[T]) is a member of the *Gammaproteobacteria* isolated from estuarine sediment for its unique ability to couple the oxidation of aromatic acids to selenate respiration. The genus *Sedimenticola* currently includes seven cultivated strains of which two species have been named and described: *S. selenatireducens* strain AK4OH1[T], the type strain of the type species for this genus [10], *S. selenatireducens* strain CUZ [11], *S. thiotaurini* strain SIP-G1 [12], *Sedimenticola* sp. strain Ke4OH1 [7], and *Sedimenticola* sp. strain NSS [11]. Here we summarize the physiological features of

* Correspondence: haggblom@sebs.rutgers.edu
[1]Department of Biochemistry and Microbiology, School of Environmental and Biological Sciences, Rutgers, The State University of New Jersey, New Brunswick, NJ, USA
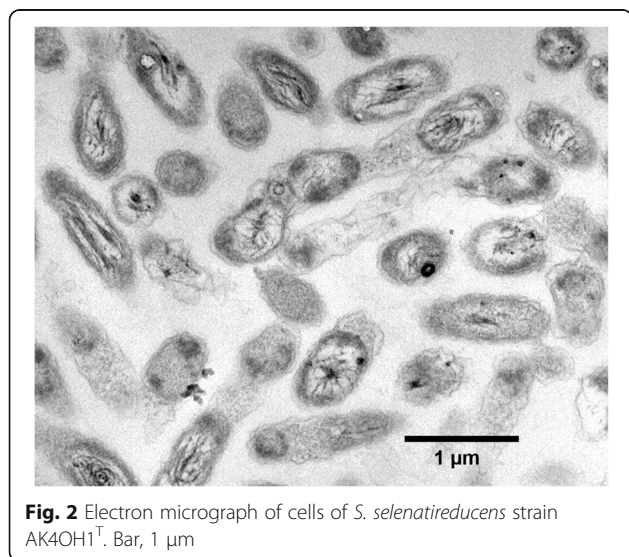Full list of author information is available at the end of the article

Louie *et al. Standards in Genomic Sciences* (2016) 11:66

Page 2 of 7
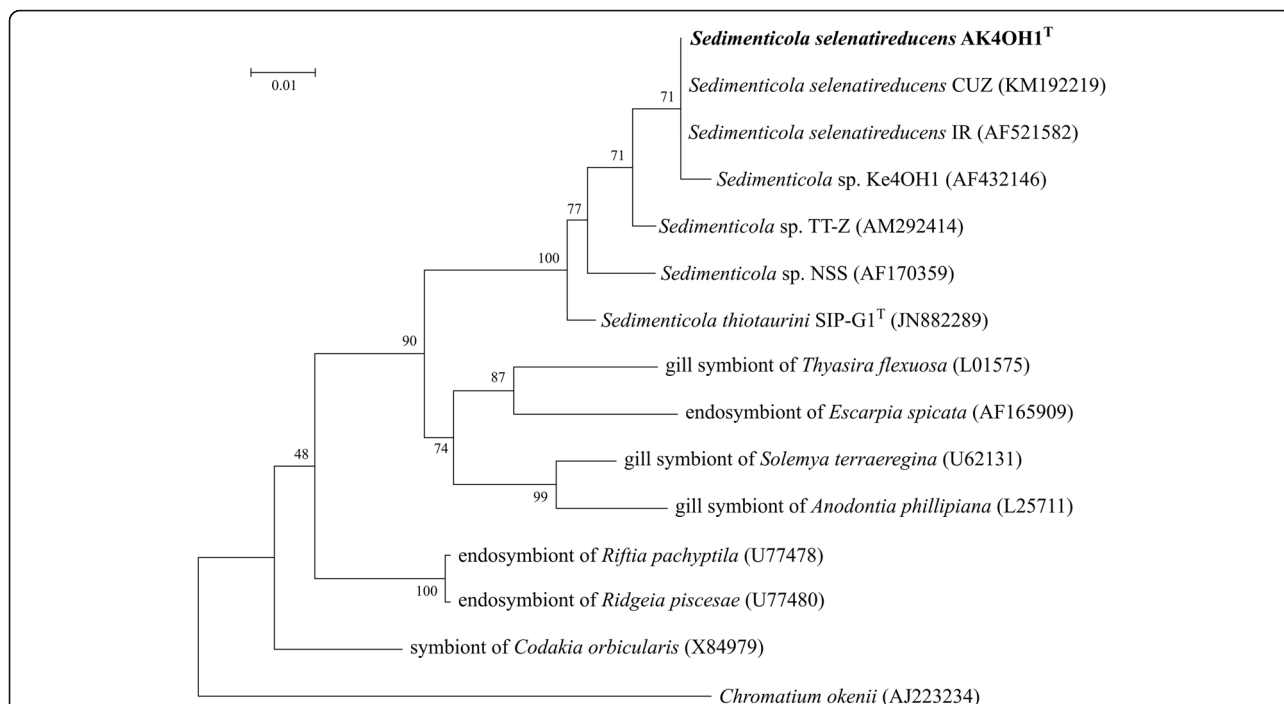
Fig. 1 Phylogenetic analysis highlighting the position of *Sedimenticola selenatireducens* strain AK4OH1[T] relative to its closest neighbors based on the 16S rRNA gene. The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model [29]. The tree with the highest log likelihood (-3985.1130) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 15 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 1276 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 [30]. The strains and their corresponding GenBank accession numbers for 16S rRNA genes are listed in parentheses. The genome accession number and locus tag of strain AK4OH1[T] are NZ_ATZE00000000.1 and A3GODRAFT_03746. (T = type strain). Bar: 0.01 substitutions per nucleotide position. *C. okenii* was used as an outgroup

Fig. 2 Electron micrograph of cells of *S. selenatireducens* strain AK4OH1[T]. Bar, 1 μm

Sedimenticola selenatireducens AK4OH1[T] and provide a description of its genome.

Organism information

Classification and features

S. selenatireducens strain AK4OH1[T] was isolated from estuarine sediment in the New York-New Jersey harbor estuary (40°586′N, 74°207′E) [10]. The position of strain AK4OH1[T] relative to its phylogenetic neighbors is shown in Fig. 1. *S. selenatireducens* strain CUZ [11] is the closest relative to strain AK4OH1[T] with a 16S rRNA gene similarity of 100 %, yet interestingly, it has not been found to respire selenate. In addition to these two, there are five other cultivated strains of the genus *Sedimenticola*: *S. thiotaurini* strain SIP-G1[T] [12], *Sedimenticola* sp. strain NSS [11], and *Sedimenticola* sp. strain Ke4OH1 [7]. The isolate TT-Z (accession number AM292414) [13] groups among the *Sedimenticola* strains (Fig. 1) suggesting that it is part of the *Sedimenticola* genus. The isolate IR (accession number AF521582) groups closely with strain AK4OH1[T] and strain CUZ, and its position in the phylogenetic tree suggests that it

Table 1 Classification and general features of *Sedimenticola selenatireducens* strain AK4OH1[T] according to the MIGS recommendations [18]

MIGS ID	Property	Term	Evidence code[a]
	Classification	Domain *Bacteria*	TAS [31]
		Phylum *Proteobacteria*	TAS [32]
		Class *Gammaproteobacteria*	TAS [33, 34]
		Genus *Sedimenticola*	TAS [10, 35]
		Species *Sedimenticola selenatireducens*	TAS [10, 35]
		Type strain: AK4OH1[T]	
	Gram stain	negative	TAS [10]
	Cell shape	rod (1.5 μm long, 0.5 μm wide)	TAS [10]
	Motility	motile at some growth stages	TAS [12]
	Sporulation	none	TAS [10]
	Temperature range	mesophile	TAS [10]
	Optimum temperature	28 °C	TAS [10]
	pH range; Optimum	7	TAS [10]
	Carbon source	benzoate, 3-hydroxybenzoate, 4-hydroxybenzoate, acetate, formate, pyruvate, methyl-pyruvate, L-lactate, D- and L-malate, propionate, fumarate, succinate, methyl-succinate, bromo-succinate, p-hydroxyphenylacetic acid, cysteine	TAS [10, 12]
MIGS-6	Habitat	estuarine sediment	TAS [10]
MIGS-6.3	Salinity	1.1-2.3 % NaCl (w/v)	TAS [10]
MIGS-22	Oxygen requirement	anaerobe-microaerophile	TAS [10, 12]
MIGS-15	Biotic relationship	free-living	TAS [10]
MIGS-14	Pathogenicity	unknown	NAS
MIGS-4	Geographic location	Hudson River estuary, Arthur Kill, intertidal strait NY/NJ, USA	TAS [10]
MIGS-5	Sample collection	1995	TAS [10]
MIGS-4.1	Latitude	40°586′N	TAS [10]
MIGS-4.2	Longitude	74°207′E	TAS [10]
MIGS-4.3	Depth	surface sediment	TAS [10]
MIGS-4.4	Altitude	sea level	TAS [10]

[a] Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [36]

is a member of the *Sedimenticola selenatireducens* species.

Cells of strain AK4OH1[T] are Gram-negative and rod-shaped [10] (Fig. 2 and Table 1). The strain can grow heterotrophically or lithoautotrophically under hypoxic and anaerobic conditions [12]. Motility is observed during early to mid-exponential growth on liquid MB2216 medium, but not in late exponential phase, and cell morphology varies depending on growth conditions [10, 12].

Strain AK4OH1[T] is able to utilize benzoate, 3-hydroxybenzoate, 4-hydroxybenzoate, acetate, formate, fumarate, L-lactate, D- and L-malate, pyruvate, methyl-pyruvate, propionate, succinate, methyl-succinate, bromo-succinate, p-hydroxyphenylacetic acid, α-ketoglutaric acid, arabinose, lyxose, ribose, xylose, D-galactonic acid-γ-

lactone, α-hydroxy-glutaric acid-γ-lactone, L-alanine, L-glutamic acid, L-serine, tyramine, and phenylethylamine [10, 12].

Chemotaxonomic data
The predominant cellular fatty acids in strain AK4OH1[T] are $C_{16:0}$ (61.9 %), $C_{16:1}$ ω7c (14.4 %), $C_{18:0}$ (8.4 %), and $C_{18:1}$ ω7c (7.2 %) [10].

Genome sequencing information
Genome project history
S. selenatireducens strain AK4OH1[T] was selected for sequencing in 2011 based on its phylogenetic position [14, 15] and is part of the study Genomic Encyclopedia of Type Strains, Phase I: the one thousand microbial

Louie *et al. Standards in Genomic Sciences* (2016) 11:66

Page 4 of 7

Table 2 Project information

MIGS ID	Property	Term
MIGS 31	Finishing quality	Level 2: High-Quality Draft
MIGS-28	Libraries used	Illumina std PE IIOC
MIGS 29	Sequencing platforms	Illumina
MIGS 31.2	Fold coverage	273×
MIGS 30	Assemblers	ALLPATHS v. R37654
MIGS 32	Gene calling method	Prodigal 2.5
	Locus Tag	A3GO
	Genbank ID	ATZE00000000.1
	GenBank Date of Release	06/18/14
	GOLD ID	Gp0013295
	BIOPROJECT ID	PRJNA165429
MIGS 13	Source Material Identifier	AK4OH1[T]
	Project relevance	Bioremediation, environmental, biogeochemical cycling of Se, Genomic Encyclopedia of Bacteria and Archaea (GEBA)

genomes project (KMG-I) [16]. The goal of the KMG-I study was to increase the coverage of sequenced reference microbial genomes [17]. The Quality Draft (QD) assembly and annotation were made available for public access on June 18, 2014. Table 2 presents the project information and its association with MIGS version 2.0 compliance [18]. The NCBI accession number for the Bioproject is PRJNA165429. The genome accession number is ATZE00000000.1 consisting of 41 contigs (ATZE01000001-ATZE01000041) and 37 scaffolds.

Table 3 Genome statistics

Attribute	Value	% of Total[a]
Genome size (bp)	4,588,530	100.00
DNA coding (bp)	4,041,165	88.07
DNA G + C (bp)	2,597,447	56.61
DNA scaffolds	37	100.00
Total genes[b]	4331	100.00
Protein coding genes	4276	98.73
RNA genes	55	1.27
Genes with function prediction	3440	79.43
Genes assigned to COGs	2832	65.39
Genes with Pfam domains	3595	83.01
Genes with signal peptides	383	8.84
Genes with transmembrane helices	1143	26.39
CRISPR repeats	1	-

[a] The total is based on either the size of the genome in base pairs or the total number of protein coding genes in the annotated genome
[b] no pseudogenes found

Growth conditions and genomic DNA preparation

S. selenatireducens strain AK4OH1[T] was grown in mineral salt medium at 28 °C with 10 mM Na_2SeO_4 as electron acceptor and 250 μM 4-hydroxybenzoate as carbon source, as previously described [10]. Genomic DNA was isolated from 0.5 g of cell paste using JetFlex Genomic DNA Purification Kit (GENOMED) as recommended by the manufacturer.

Genome sequencing and assembly

Sequencing was achieved using an Illumina [19] platform using a std paired-end library obtaining 273× fold coverage. The sequencing was done at the DOE Joint Genome Institute. ALLPATHS assembly software [20] was used to obtain 41 final contigs. Quality check and assembly statistics were performed at JGI. The raw sequences were screened against contaminants and 0.1 % of the reads were removed.

Genome annotation

Gene calling was performed using Prodigal 2.5 [21]. The genome sequence was analyzed using the Joint Genome Institute IMG system [22]. Ribosomal RNAs were predicted based upon sequence similarity, using BLAST, against the non-redundant nucleotide database and/or using Infernal and Rfam models. tRNA genes were found using tRNAscan-SE [23]. The predicted CDS were searched using the NCBI non-redundant protein database. The major metabolic pathways and predicted protein set were searched using KEGG, SwissProt, COG, Pfam, and InterPro protein databases implemented in the IMG. Additional gene prediction analysis and manual functional annotation were performed within IMG and using Artemis software (release 13.0, Sanger Institute).

Genome properties

The high quality draft genome sequence consists of 37 scaffolds that account for a total of 4,588,530 bp with a 56.6 % G + C content. In total, 4331 genes were predicted, 4276 of which are protein-coding genes, 55 RNA genes, and no pseudogenes. The majority of the predicted genes (79 %) were assigned a predicted function. The properties and statistics of the genome are summarized in Table 3 and Table 4.

Insights from the genome sequence

The respiratory flexibility of anaerobic prokaryotes allowing them to employ different terminal electron acceptors for respiration enables these organisms to thrive in dynamic redox environments. Among the enzymes that catalyze oxidation-reduction reactions of metals and metalloids are those that are highly conserved and belong to the DMSO reductase family [24]. Key members

Louie *et al. Standards in Genomic Sciences* (2016) 11:66

Page 5 of 7

Table 4 Number of genes associated with general COG functional categories

Code	Value	%age	Description
J	205	6.48	Translation, ribosomal structure and biogenesis
A	1	0.03	RNA processing and modification
K	180	5.69	Transcription
L	117	3.70	Replication, recombination and repair
B	2	0.06	Chromatin structure and dynamics
D	41	1.30	Cell cycle control, Cell division, chromosome partitioning
V	66	2.09	Defense mechanisms
T	244	7.71	Signal transduction mechanisms
M	160	5.06	Cell wall/membrane biogenesis
N	120	3.79	Cell motility
U	49	1.55	Intracellular trafficking and secretion
O	207	6.54	Posttranslational modification, protein turnover, chaperones
C	339	10.71	Energy production and conversion
G	116	3.67	Carbohydrate transport and metabolism
E	244	7.71	Amino acid transport and metabolism
F	57	1.80	Nucleotide transport and metabolism
H	166	5.24	Coenzyme transport and metabolism
I	148	4.68	Lipid transport and metabolism
P	187	5.91	Inorganic ion transport and metabolism
Q	76	2.40	Secondary metabolites biosynthesis, transport and catabolism
R	211	6.67	General function prediction only
S	175	5.53	Function unknown
-	1499	34.61	Not in COGs

The total is based on the total number of protein coding genes in the genome
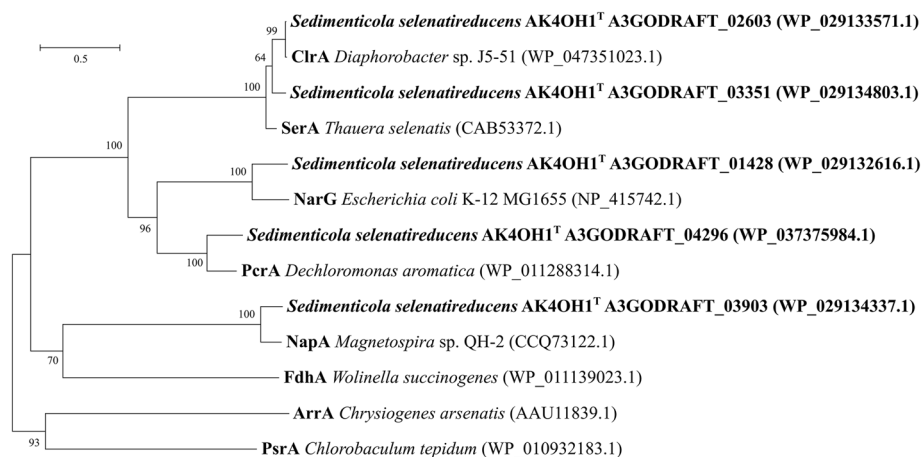
Fig. 3 Phylogenetic analysis highlighting the relation of *Sedimenticola selenatireducens* strain AK4OH1^T genes to known DMSO reductases by Maximum Likelihood method. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model [37]. The tree with the highest log likelihood (-17325.9218) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 13 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 724 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 [30]. GenBank accession numbers are listed in parentheses. Bar = 0.5 substitutions per nucleotide position

Louie et al. Standards in Genomic Sciences (2016) 11:66

Page 6 of 7

of the DMSO family of reductases, which transfer electrons to a variety of substrates that act as terminal electron acceptors for energy generation, are nitrate reductases (Nar, Nap, Nas), arsenate reductase (Arr), selenate reductase (Ser), and chlorate reductase (Clr), among others.

S. selenatireducens strain AK4OH1[T] can use nitrate, nitrite and selenate as the terminal electron acceptors for anaerobic growth, while using the electron donors acetate, lactate, pyruvate, benzoate, 3-hydroxybenzoate, and 4-hydroxybenzoate [10]. Chlorate and perchlorate can be used as electron acceptors when peptone is used as an energy source [12]. (Micro-)aerobic growth with oxygen as electron-acceptor and peptones as electron-donor is also detected [12]

Within the AK4OH1[T] genome, there are several likely DMSO reductases. Figure 3 shows the grouping of AK4OH1[T] genes with closely matching, known, DMSO reductases. A3GODRAFT_03903 groups closely with the NapA, from *Magnetospira* sp. QH-2. A3GOD-RAFT_01428 clusters together with the NarG of *Escherichia coli* K-12 MG1655. Both of these genes are organized in gene clusters similar to known *nap* and *nar* operons [25]. BLAST searches of the AK4OH1[T] genome using arsenate reductases showed no genes with significant similarity. This agrees with strain AK4OH1's inability to respire arsenate [10]. A3GODRAFT_02603 and A3GODRAFT_03351 from strain AK4OH1[T] cluster closely with the chlorate reductase from *Diaphorobacter* sp. J5-51 and with the selenate reductase from *Thauera selenatis*. A3GODRAFT_02603, which groups closest with ClrA, resembles the gene organization of a *clr* operon [26]. While the only well-studied respiratory selenate reductase, *serA*, is from *Thauera selenatis*, A3GODRAFT_03351 and its neighboring genes follow the same organization as found with *serABDC* [27]. Gene A3GODRAFT_04296 clusters together with the perchlorate reductase from *Dechloromonas aromatica*, and appears to have the same gene organization as a *pcr* operon [28].

Conclusions

The complete genome of the estuarine bacterium *Sedimenticola selenatireducens* AK4OH1[T] provides a stronger foundation from which to learn more about the process of dissimilatory selenate reduction. As AK4OH1[T] was the first organism isolated capable of coupling the respiration of selenate to the oxidation of benzoic acids, its genome also provides a starting point for learning more about this unique capability.

Abbreviations
DMSO: Dimethyl sulfoxide; SeRB: Selenate reducing bacteria;

Authors' contributions
MMH, EB and NY designed the research. PN carried out initial strain characterization. VS provided the electron micrograph. MG, H-PK, EL, NCK and TW sequenced, assembled and annotated the genome. TSL, DG, EB, NY and MMH performed the research. TSL and DG analyzed the data. TSL, DG, EB, NY and MMH wrote the manuscript. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Biochemistry and Microbiology, School of Environmental and Biological Sciences, Rutgers, The State University of New Jersey, New Brunswick, NJ, USA. [2]Institute of Earth, Ocean, and Atmospheric Science, School of Environmental and Biological Sciences, Rutgers, The State University of New Jersey, New Brunswick, NJ, USA. [3]Institute of Marine Science, ISMAR, National Research Council of Italy, CNR, Ancona, Italy. [4]Institute for Advanced Studies, Program in Interdisciplinary Studies, Princeton, NJ, USA. [5]Department of Environmental Sciences, School of Environmental and Biological Sciences, Rutgers, The State University of New Jersey, New Brunswick, NJ, USA. [6]Department of Cell Biology and Neuroscience, Rutgers, The State University of New Jersey, Piscataway, NJ, USA. [7]Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany. [8]Newcastle University, School of Biology, Newcastle upon Tyne, UK. [9]Department of Energy Joint Genome Institute, Genome Biology Program, Walnut Creek, CA, USA. [10]Department of Biological Sciences, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia. [11]Pharmacy Practice and Administration, Ernest Mario School of Pharmacy, Rutgers, The State University of New Jersey, Piscataway, NJ, USA. [12]Present address: Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, MN, USA.

References
1. Nancharaiah YV, Lens PNL. Ecology and biotechnology of selenium-respiring bacteria. Microbiol Mol Biol Rev. 2015;79:61–80.
2. Laverman AM, Blum JS, Schaefer JK, Phillips E, Lovley DR, Oremland RS. Growth of strain SES-3 with arsenate and other diverse electron acceptors. Appl Environ Microbiol. 1995;61:3556–61.
3. Rauschenbach I, Posternak V, Cantarella P, McConnell J, Starovoytov V, Häggblom MM. Seleniivibrio woodruffii gen. nov., sp. nov., a selenate- and arsenate-respiring bacterium in the Deferribacteraceae. Int J System Evol Microbiol. 2013;63:3659–65.
4. Knight V, Blakemore R. Reduction of diverse electron acceptors by Aeromonas hydrophila. Arch Microbiol. 1998;169:239–48.
5. Baesman SM, Stolz JF, Kulp TR, Oremland RS. Enrichment and isolation of Bacillus beveridgei sp. nov., a facultative anaerobic haloalkaliphile from Mono Lake, California, that respires oxyanions of tellurium, selenium, and arsenic. Extremophiles. 2009;13:695–705.
6. Macy J, Rech S, Auling G, Dorsch M, Stackebrandt E, Sly L. Thauera selenatis gen. nov., sp. nov., a member of the beta subclass of Proteobacteria with a novel type of anaerobic respiration. Int J System Bacteriol. 1993;43:135.
7. Knight VK, Nijenhuis I, Kerkhof LJ, Häggblom MM. Degradation of aromatic compounds coupled to selenate reduction. Geomicrobiol J. 2002;19:77–86.

Louie *et al. Standards in Genomic Sciences* (2016) 11:66

Page 7 of 7

8. Yamamura S, Yamashita M, Fujimoto N, et al. *Bacillus selenatarsenatis* sp. nov., a selenate- and arsenate-reducing bacterium isolated from the effluent drain of a glass-manufacturing plant. Int J System Evol Microbiol. 2007;57:1060–4.

9. Blum JS, Stolz JF, Oren A, Oremland RS. *Selenihalanaerobacter shriftii* gen. nov., sp. nov., a halophilic anaerobe from Dead Sea sediments that respires selenate. Arch Microbiol. 2001;175:208–19.

10. Narasingarao P, Häggblom MM. *Sedimenticola selenatireducens*, gen. nov., sp. nov., an anaerobic selenate-respiring bacterium isolated from estuarine sediment. Syst Appl Microbiol. 2006;29:382–8.

11. Carlström CI, Loutey DE, Wang O, et al. Phenotypic and genotypic description of *Sedimenticola selenatireducens* strain CUZ, a marine (per)chlorate-respiring gammaproteobacterium, and its close relative the chlorate-respiring *Sedimenticola* strain NSS. Appl Environ Microbiol. 2015;81:2717–26.

12. Flood BE, Jones DS, Bailey JV. *Sedimenticola thiotaurini* sp. nov., a sulfide-oxidizing bacterium isolated from salt marsh sediments, and emended description of the genus *Sedimenticola* and *Sedimenticola selenatireducens*. Int J Syst Evol Microbiol. 2015;65:2522–30.

13. Alain K, Harder J, Widdel F, Zengler K. Anaerobic utilization of toluene by marine alpha- and gammaproteobacteria reducing nitrate. Microbiology. 2012;158:2946–57.

14. Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, et al. A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. Nature. 2009;462:1056–60.

15. Göker M, Klenk HP. Phylogeny-driven target selection for large-scale genome sequencing (and other) projects. Stand Genomic Sci. 2013;8:360–74.

16. Kyrpides NC, Woyke T, Eisen JA, Garrity G, Lilburn TG, Beck BJ, et al. Genomic encyclopedia of type strains, phase I: the one thousand microbial genomes (KMG-I) project. Stand Genomic Sci. 2013;9:628–6234.

17. Kyrpides NC, Hugenholtz P, Eisen JA, Woyke T, Göker M, Parker CT, et al. Genomic encyclopedia of Bacteria and Archaea: sequencing a myriad of type strains. PLoS Biol. 2014;8:e1001920.

18. Field D, Garrity G, Gray T, et al. The minimum information about a genome sequence (MIGS) specification. Nat Biotech. 2008;26:541–7.

19. Bennett S. Solexa Ltd. Pharmacogenomics J. 2004;5:433–8.

20. Butler J, MacCallum I, Kleber M, et al. ALLPATHS: De novo assembly of whole-genome shotgun microreads. Genome Res. 2008;18:810–20.

21. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119.

22. Markowitz VM, Chen I-MA, Palaniappan K, et al. IMG 4 version of the integrated microbial genomes comparative analysis system. Nucl Acids Res. 2014;42:D560–7.

23. Lowe TM, Eddy SR. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucl Acids Res. 1997;25:0955–64.

24. Rothery RA, Workun GJ, Weiner JH. The prokaryotic complex iron–sulfur molybdoenzyme family. BBA-Biomembranes. 2008;1778:1897–929.

25. Richardson DJ, Berks BC, Russell DA, Spiro S, Taylor CJ. Functional, biochemical and genetic diversity of prokaryotic nitrate reductases. Cell Mol Life Sci. 2001;58:165–78.

26. Lindqvist MH, Nilsson T, Sundin P, Rova M. Chlorate reductase is cotranscribed with cytochrome c and other downstream genes in the gene cluster for chlorate respiration of *Ideonella dechloratans*. FEMS Microbiol Lett. 2015;362:1–6.

27. Krafft T, Bowen A, Theis F, Macy JM. Cloning and sequencing of the genes encoding the periplasmic-cytochrome B-containing selenate reductase of *Thauera selenatis*. DNA Seq. 2000;10:365–77.

28. Bender KS, Shang C, Chakraborty R, Belchik SM, Coates JD, Achenbach LA. Identification, characterization, and classification of genes encoding perchlorate reductase. J Bacteriol. 2005;187:5090–6.

29. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 1993;10:512–26.

30. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol. 2013;30:2725–9.

31. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. PNAS. 1990;87:4576–9.

32. Garrity GM, Bell JA, Lilburn T. Phylum XIV. Proteobacteria *phyl. nov.* In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. Bergey's manual of systematic bacteriology. Volume 2, Part B. New York: Springer; 2005. p. 1.

33. Garrity GM, Bell JA, Lilburn T, Class III. Gammaproteobacteria *class. nov.* In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. Bergey's manual of systematic bacteriology. Volume 2, Part B. New York: Springer; 2005. p. 1.

34. Euzéby J. Validation of publication of new names and new combinations previously effectively published outside the IJSEM. List no. 106. Int J Syst Evol Microbiol. 2005;55:2235–8.

35. Euzéby J. List of new names and new combinations previously effectively, but not validly, published. List no. 112. Int J Syst Evol Microbiol. 2006;56:2507–8.

36. Ashburner M, Ball CA, Blake JA, et al. Gene Ontology: tool for the unification of biology. Nat Genet. 2000;25:25–9.

37. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci. 1992;8:275–82.