Data Article

Whole-genome sequence data and analysis of type strains 'Pusillimonas nitritireducens' and 'Pusillimonas subterraneus' isolated from nitrate- and radionuclide-contaminated groundwater in Russia

Denis S. Grouzdev [a], Tatiyana P. Tourova [b], Tamara L. Babich [b], Margarita A. Shevchenko [c], Diyana S. Sokolova [b], Ruslan R. Abdullin [b], Andrey B. Poltaraus [d], Stepan V. Toshchakov [c], Tamara N. Nazina [b,e,*]

[a] Institute of Bioengineering, Research Center of Biotechnology, Russian Academy of Sciences, Moscow, Russian Federation
[b] Winogradsky Institute of Microbiology, Research Center of Biotechnology, Russian Academy of Sciences, Moscow, Russian Federation
[c] Immanuel Kant Baltic Federal University, Kaliningrad, Russian Federation
[d] Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russian Federation
[e] V.I. Vernadsky Institute of Geochemistry and Analytical Chemistry of Russian Academy of Sciences, Moscow, Russian Federation

ARTICLE INFO

ABSTRACT

Two strains, 'Pusillimonas nitritireducens' JR1/69-2-13[T] and 'Pusillimonas subterraneus' JR1/69-3-13[T], of aerobic, motile, Gram-negative, non-spore-forming, organotrophic, psychrotolerant bacteria were isolated from a sample of nitrate- and radionuclide-contaminated groundwater in Russia. Here we describe the draft genomes of these strains. The sequenced and annotated genome of the strain JR1/69-2-13[T] contained 4.3 Mbp with 4108 protein-coding genes. The genome of the strain JR1/69-3-13[T] contained 4.5 Mbp with 4260 protein-coding genes. Genome analysis of both strains provides an insight into the genomic basis of their resistance to nitrate, heavy metals and metalloids. The draft genome sequences of strains 'Pusillimonas nitritireducens' JR1/69-2-13[T] and

* Corresponding author at: V.I. Vernadsky Institute of Geochemistry and Analytical Chemistry of Russian Academy of Sciences, Moscow, Russian Federation.
E-mail address: nazina@inmi.ru (T.N. Nazina).

‘*Pusillimonas subterraneus*’ JR1/69-3-13[T] are available at DDBJ/ EMBL/GenBank under the accession nos. https://www.ncbi.nlm. nih.gov/nuccore/PDNV00000000 and https://www.ncbi.nlm.nih. gov/nuccore/PDNW00000000, respectively.

Specifications table

Subject area	Biology
More specific subject area	Microbiology and genomics.
Type of data	Genome sequencing data, table, text file, image and figure.
How data was acquired	Transmission electron microscope (JEOL JEM-1010, Japan), Shotgun draft genome DNA sequencing using MiSeq platform (Illumina) and bioinformatics applications.
Data format	Draft genome sequences and gene prediction.
Experimental factors	Isolation and characterization of strains JR1/69-2-13[T] and JR1/69-3-13[T]. Genomic DNA extraction and sequencing procedure.
Experimental features	Draft genome sequencing was performed according to Illumina sequencing protocols for DNA-seq followed by annotation and gene description.
Data source location	Strains were isolated from nitrate- and radionuclide-contaminated groundwater sample (Ozyorsk town, South Urals, Russia). Latitude: 55°38′ N and Longitude: 60°47′ E.
Data accessibility	The draft genome sequences of strains JR1/69-2-13[T] and JR1/69-3-13[T] are available at DDBJ/EMBL/GenBank under the accession nos. PDNV00000000 and PDNW00000000, respectively. (https://www.ncbi.nlm.nih.gov/nuccore/PDNV00000000, https://www.ncbi.nlm.nih.gov/nuccore/PDNW00000000)

Value of the data

- Draft genome sequences of ‘*P. nitritireducens*’ and ‘*P. subterraneus*’ will create an opportunity for comparative taxonomic studies of the closely related *Pusillimonas – Candidimonas* group.
- Draft genomes and identified genes of new strains will provide insights into the molecular mechanisms by which these strains transform nitrate or nitrite to dinitrogen gas and survive in the nitrate- and radionuclides-contaminated environment.
- This data set will be useful for the scientific community, working in the area of functional and phylogenetic diversity of microorganisms in environments contaminated with radioactive waste, and for development of biotechnologies for *in situ* bioremediation of groundwater by nitrate removal.

1. Data

In the present work, we describe the draft genome sequences and gene prediction of two strains of aerobic organotrophic bacteria, JR1/69-2-13[T] and JR1/69-3-13[T] (Fig. 1), isolated from nitrate- and radionuclides-contaminated groundwater collected near the industrial reservoir for liquid radioactive waste (Ozyorsk town, South Urals, Russia) [1]. Functional and phylogenetic diversity of the microorganisms
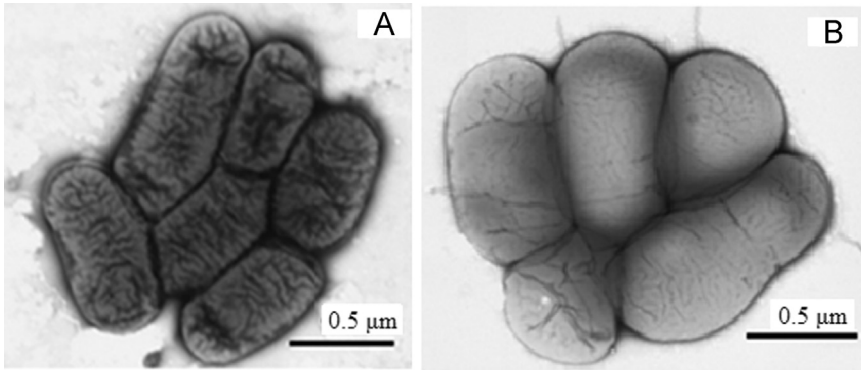
Fig. 1. Transmission electron micrographs of 'Pusillimonas nitritireducens' JR1/69-2-13[T] (A) and Pusillimonas subterraneus' JR1/69-3-13[T] (B). Scale bars are 0.5 μm.

Table 1
Genomes features of 'Pusillimonas nitritireducens' JR1/69-2-13[T] and 'Pusillimonas subterraneus' JR1/69-3-13[T].

Features	JR1/69-2-13[T]		JR1/69-3-13[T]	
	Value	% of Total	Value	% of Total
Genome size (bp)	4,310,404	100.00	4,544,755	100.00
DNA coding (bp)	3,920,667	90.96	4,139,718	91.09
DNA G+C (bp)	2,467,295	57.24	2,632,477	57.92
DNA scaffolds	38	100.00	68	100.00
Total genes	4122	100.00	4270	100.00
Protein coding genes	3981	96.58	4118	96.44
RNA genes	51	1.24	54	1.26
Pseudo genes	90	2.18	98	2.30
Genes in internal clusters	1312	31.46	1321	30.59
Genes with function prediction	3547	85.04	3686	85.34
Genes assigned to COGs	3194	76.58	3286	76.08
Genes with Pfam domains	3626	86.93	3778	87.47
Genes with signal peptides	431	10.33	447	10.35
Genes with transmembrane helices	997	23.90	1044	24.17
CRISPR repeats	–		–	

inhabiting the subterranean water-bearing horizons associated with the plants processing uranium and other radionuclides remain insufficiently studied [2,3]. Two strains, JR1/69-2-13[T] and JR1/69-3-13[T], were chosen for genome sequencing with the goal of determination of their taxonomic position and identification of the genetic determinants providing for their occurrence in the environment contaminated with radioactive waste. Based on the 16S rRNA gene phylogeny data, both strains were found to belong to the domain Bacteria, the phylum Proteobacteria, the class Betaproteobacteria, the order Burkholderiales, the family Alcaligenaceae [4–8]. The 98.7% similarity level between the 16S rRNA gene sequences of these two strains indicated that they could probably belong to different species. The 16S rRNA gene sequences of the strains JR1/69-2-13[T] and JR1/69-3-13[T] were phylogenetically closely related to those of the species of the genera Pusillimonas and Candidimonas [9–15]. As was demonstrated in previous studies, differentiation between the genera of the family Alcaligenaceae on the basis of their phenotypic characteristics is not possible [9,10,14,16]. Genomic features of strains Pusillimonas nitritireducens' JR1/69-2-13[T] and 'Pusillimonas subterraneus' JR1/69-3-13[T] are presented in Table 1. Numerous genes responsible for heavy metal tolerance and detoxification were identified in their genomes. In the JR1/69-3-13[T] genome, the denitrification gene
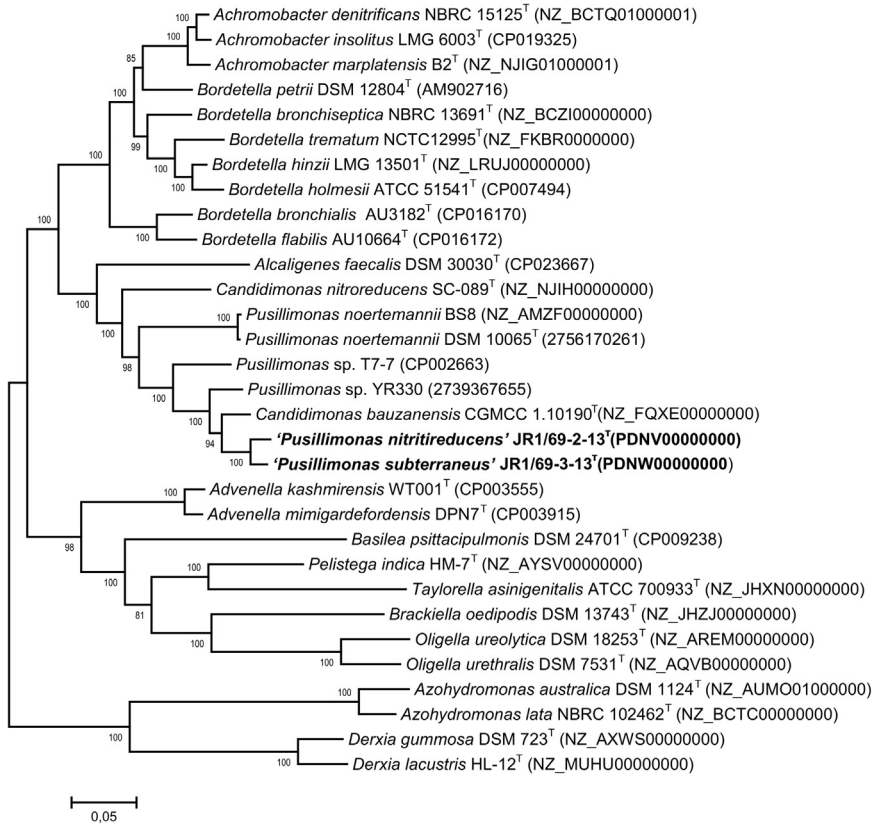
Fig. 2. Phylogenetic tree inferred from the comparison of 492 concatenated *Alcaligenaceae* core protein-coding genes (167141 aa) showing the position of strains JR1/69-2-13^T and JR1/69-3-13^T. Bootstrap values are indicated at branch nodes.

cluster included the genes predicted to encode the enzymes for sequential reduction of nitrate to dinitrogen gas, which was consistent with the complete denitrification phenotype of the strain. In the JR1/69-2-13^T genome, the genes encoding nitrite reduction to N_2 were revealed. *In silico* DNA-DNA hybridization (dDDH) values against reference genomes of *Pusillimonas* and *Candidimonas* strains were within the range of 19.3–28.8, and were below the 70% threshold to differentiate bacterial species [17]. These data and position of strains JR1/69-2-13^T and JR1/69-3-13^T on the tree of 492 concatenated core protein-coding genes (Fig. 2) testify to their belonging to two novel species within the *Pusillimonas* genus with proposed names 'Pusillimonas nitritireducens' and 'Pusillimonas subterraneus', respectively. The draft genome sequences of strains JR1/69-2-13^T and JR1/69-3-13^T are deposited in DDBJ/EMBL/Genbank under the accession nos. PDNV00000000 and PDNW00000000, respectively. The versions described in this paper are the first versions of the genome of each strain.

2. Experimental design, materials, and methods

2.1. Isolation of the strains JR1/69-2-13^T and JR1/69-2-13^T

Strains JR1/69-2-13^T and JR1/69-2-13^T were isolated from nitrate- and radionuclide-contaminated groundwater sample collected from the depth of 44 m at a distance 3.2 km from the Karachai Lake

(Ozyorsk town, South Urals, Russia) [1,18]. Two strains were purified by successive transfer from the liquid TEG medium containing bacto-trypton ($5.0\,g\,L^{-1}$), yeast extract ($2.5\,g\,L^{-1}$), glucose ($1.0\,g\,L^{-1}$), and distilled water (1 L, pH 7.0) to solid TEG medium with agar-agar ($15.0\,g\,L^{-1}$). Bacteria were incubated at 22–30 °C. Strains JR1/69-2-13[T] and JR1/69-3-13[T] were deposited in the Russian Collection of Microorganisms as VKM в-3222[T] and VKM в-3223[T], respectively.

2.2. DNA isolation and sequencing

Strains JR1/69-2-13[T] and JR1/69-3-13[T] were grown in TEG liquid medium at 30 °C for 72 h. The cells were harvested by centrifugation. The cell integrity was accessed by transmission electron microscopy (JEOL JEM-1010, Japan) of the cells negatively stained with 1% phosphotungstic acid (Fig. 1A and B). DNA of each strain was isolated from the biomass by the phenol-chloroform-based method as described previously [19]. NEBNext Ultra DNA library preparation kit (New England Biolabs, USA) was used to prepare fragment libraries for genome sequencing. Next-generation shotgun-sequencing was performed on MiSeq (Illumina Inc., USA) at the Immanuel Kant Baltic Federal University (Kaliningrad, Russian Federation).

2.3. Genome assembly

Raw sequence reads were subjected to stringent quality filtering with CLC Genomics Workbench 10.0 (Qiagen, Germany). After filtering, sequencing adapters were trimmed with SeqPrep tool (https:// github.com/jstjohn/SeqPrep). Finally, 2,802,807 and 2,451,909 read pairs were used for *de novo* assembly of JR1/69-2-13[T] and JR1/69-3-13[T] genomes, respectively. The reads were assembled with SPAdes 3.10.0 [20] and refined by length and coverage using CLC Genomics Workbench 10.0 software (Qiagen, Germany).

2.4. Genome annotation

Identification of protein-coding sequences and primary annotation was performed using the NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) [21]. Additional gene prediction and functional annotation were performed in the Rapid Annotation using Subsystems Technology (RAST) server [22] and Integrated Microbial Genome-Expert Review pipeline [23], respectively. Analysis of metal resistance was performed with the Antibacterial Biocide and Metal Resistance Genes Database (BacMet) [24]. *In silico* DNA-DNA hybridization (dDDH) was carried out with the online genome-to-genome calculator (GGDC 2.0) provided by the DSMZ [17].

Transparency document.   Supplementary materia

Transparency document associated with this article can be found in the online version at https:// doi.org/10.1016/j.dib.2018.10.060.

References

[1] A.P. Novikov, S.N. Kalmykov, S. Utsunomiya, R.C. Ewing, F. Horreard, A. Merkulov, S.B. Clark, V.V. Tkachev, B.F. Myasoedov, Colloid transport of plutonium in the far-field of the Mayak Production Association, Russia, Science 314 (5799) (2006) 638–641. https://doi.org/10.1126/science.1131307.

[2] J.D. Wall, L.R. Krumholz, Uranium reduction, Annu. Rev. Microbiol. 60 (2006) 149–166. https://doi.org/10.1146/annurev. micro.59.030804.121357.

[3] J.R. Lloyd, J.C. Renshaw, Bioremediation of radioactive waste: radionuclide–microbe interactions in laboratory and field-scale studies, Curr. Opin. Biotechnol. 16 (2005) 254–260. https://doi.org/10.1016/j.copbio.2005.0.

[4] C.R. Woese, O. Kandler, M.L. Wheelis, Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya, Proc. Natl. Acad. Sci. USA (1990) 4576–4579. https://doi.org/10.1073/pnas.87.12.4576.

[5] G.M. Garrity, J.A. Bell, T. Lilburn, Phylum XIV. Proteobacteria phyl. nov, in: G.M. Garrity, D.J. Brenner, N.R. Krieg, J.T. Staley (Eds.), 2nd ed.,Bergey's Manual of Systematic Bacteriology, vol. 2, Part B, Springer, New York, 2005, p. 1.

[6] G.M. Garrity, J.A. Bell, T. Lilburn, Class II. Betaproteobacteria class. nov, in: G.M. Garrity, D.J. Brenner, N.R. Krieg, J.T. Staley (Eds.), 2nd ed.,Bergey's Manual of Systematic Bacteriology, vol. 2, part B, Springer, New York, 2005, p. 1.

[7] G.M. Garrity, J.A. Bell, T. Lilburn, Order I. Burkholderiales ord. nov, in: G. Garrity, D. Brenner, N. Krieg, J. Staley (Eds.), 2nd ed.,Bergey's Manual of Systematic Bacteriology, 2, Springer, New York, 2005.

[8] J. De Ley, P. Segers, K. Kersters, W. Mannheim, A. Lievens, Intra- and intergeneric similarities of the Bordetella ribosomal ribonucleic acid cistrons: proposal for a new family, Alcaligenaceae, Int. J. Syst. Bacteriol. 36 (1986) 405–414. https://doi. org/10.1099/00207713-36-3-405.

[9] A. Stolz, S. Burger, A. Kuhm, P. Kampfer, H.-J. Busse, Pusillimonas noertemannii gen. nov., sp. nov., a new member of the family Alcaligenaceae that degrades substituted salicylates, Int. J. Syst. Evol. Microbiol. 55 (2005) 1077–1081. https://doi. org/10.1099/ijs.0.63466-0.

[10] S. Srinivasan, M.K. Kim, G. Sathiyaraj, Y.-J. Kim, D.-C. Yang, Pusillimonas ginsengisoli sp. nov., isolated from soil of a ginseng field, Int. J. Syst. Evol. Microbiol. 60 (2010) 1783–1787. https://doi.org/10.1099/ijs.0.018358-0.

[11] M. Lee, S.-G. Woo, M. Chae, L.N. Ten, Pusillimonas soli sp. nov., isolated from farm soil, Int. J. Syst. Evol. Microbiol. 60 (2010) 2326–2330. https://doi.org/10.1099/ijs.0.020404-0.

[12] M.S. Park, Y.-J. Park, J.Y. Jung, S.H. Lee, W. Park, K. Lee, C.O. Jeon, Pusillimonas harenae sp. nov., isolated from a sandy beach, and emended description of the genus Pusillimonas, Int. J. Syst. Evol. Microbiol. 61 (2011) 2901–2906. https://doi.org/ 10.1099/ijs.0.029892-0.

[13] L. Jin, S.-R. Ko, Y. Cui, C.S. Lee, H.-M. Oh, C.-Y. Ahn, H.-G. Lee, Pusillimonas caeni sp. nov., isolated from a sludge sample of a biofilm reactor, Antonie Van Leeuwenhoek 110 (2017) 124–132. https://doi.org/10.1007/s10482-016-0782-6.

[14] I. Vaz-Moreira, V. Figueira, A.R. Lopes, E. De Brandt, P. Vandamme, O.C. Nunes, C.M. Manaia, Candidimonas nitroreducens gen. nov., sp. nov. and Candidimonas humi sp. nov., isolated from sewage sludge compost, Int. J. Syst. Evol. Microbiol. 61 (2011) 2238–2246. https://doi.org/10.1099/ijs.0.021188-0.

[15] D.-C. Zhang, H.-J. Busse, C. Wieser, H.-C. Liu, Y.-G. Zhou, F. Schinner, R. Margesin, Candidimonas bauzanensis sp. nov., isolated from soil, and emended description of the genus Candidimonas Vaz-Moreira et al. 2011, Int. J. Syst. Evol. Microbiol. 62 (2012) 2084–2089. https://doi.org/10.1099/ijs.0.036400-0.

[16] P. Kämpfer, K. Denger, A.M. Cook, S.-T. Lee, U. Jäckel, E.B.M. Denner, H.-J. Busse, Castellaniella gen. nov., to accommodate the phylogenetic lineage of Alcaligenes defragrans, and proposal of Castellaniella defragrans gen. nov., comb. nov. and Castellaniella denitrificans sp. nov, Int. J. Syst. Evol. Microbiol. 56 (2006) 815–819. https://doi.org/10.1099/ijs.0.63989-0.

[17] J.P. Meier-Kolthoff, A.F. Auch, H.-P. Klenk, M. Göker, Genome sequence-based species delimitation with confidence intervals and improved distance functions, BMC Bioinforma. 14 (2013) 60. https://doi.org/10.1186/1471-2105-14-60.

[18] I.N. Solodov, A.V. Zotov, A.D. Khoteev, A.P. Mukhamet-Galeev, B.R. Tagirov, J.A. Apps, Geochemistry of natural and con-taminated subsurface waters in fissured bed rocks of the Lake Karachai area, Southern Urals, Russia, Appl. Geochem. 13 (8) (1998) 921–939. https://doi.org/10.1016/S0883-2927(98)00025-0.

[19] O.A. Podosokorskaya, E.A. Bonch-Osmolovskaya, A.V. Beskorovaynyy, S.V. Toshchakov, T.V. Kolganova, I.V. Kublanov, Mobilitalea sibirica gen. nov., sp. nov., a halotolerant polysaccharide-degrading bacterium, Int. J. Syst. Evol. Microbiol. 64 (2014) 2657–2661. https://doi.org/10.1099/ijs.0.057109-0.

[20] A.D. Prjibelski, I. Vasilinetc, A. Bankevich, A. Gurevich, T. Krivosheeva, S. Nurk, S. Pham, A. Korobeynikov, A. Lapidus, P. A. Pevzner, ExSPAnder: a universal repeat resolver for DNA fragment assembly, Bioinformatics 30 (12) (2014) i293–i301. https://doi.org/10.1093/bioinformatics/btu266.

[21] T. Tatusova, M. DiCuccio, A. Badretdin, V. Chetvernin, S. Ciufo, W. Li, Prokaryotic Genome Annotation Pipeline. The NCBI Handbook [Internet], 2nd ed., NCBI, Bethesda, MD, 2013 ⟨http://www.ncbi.nlm.nih.gov/books/NBK174280⟩.

[22] R.K. Aziz, D. Bartels, A.A. Best, M. DeJongh, T. Disz, R.A. Edwards, K. Formsma, S. Gerdes, E.M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A.L. Osterman, R.A. Overbeek, L.K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G.D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, O. Zagnitko, The RAST server: rapid annotations using subsystems tech-nology, BMC Genom. 9 (2008) 75. https://doi.org/10.1186/1471-2164-9-75.

[23] V.M. Markowitz, K. Mavromatis, N.N. Ivanova, I.M.A. Chen, K. Chu, N.C. Kyrpides, IMG ER: a system for microbial genome annotation expert review and curation, Bioinformatics. 25 (17) (2009) 2271–2278. https://doi.org/10.1093/bioinformatics/ btp393.

[24] C. Pal, J. Bengtsson-Palme, C. Rensing, E. Kristiansson, D.G.J. Larsson, BacMet: antibacterial biocide and metal resistance genes database, Nucl. Acids Res. 42 (2014) D737–D743. https://doi.org/10.1093/nar/gkt1252.