https://doi.org/10.18321/ectj823

Metagenomic Analysis Reveals Correlation Between Microbiome Structure and Leonardite Characteristics from Kazakhstan Coal Deposits

A.A. Zhubanova[1,2,3], Q. Xiaohui[3], P.S. Ualieva[2,3], G.Zh. Abdieva[2,3], K.T. Tastambek[1,2,3], G.K. Kayrmanova[2,3], N.Sh. Akimbekov[1,2,3*]

[1]Scientific Research Institute of Biology and Biotechnology Problems, 71 al-Farabi ave., Almaty, Kazakhstan
[2]Scientific Research Institute of Ecology Problems, 71 al-Farabi ave., Almaty, Kazakhstan
[3]al-Farabi Kazakh National University, 71 al-Farabi ave., Almaty, Kazakhstan

Article info

Received:
11 November 2018

Received in revised form:
14 January 2018

Accepted:
27 February 2019

Keywords
Leonardite
Coal characteristics
Microbial diversity
Microbial community
Metagenomics

Abstract

Coal microbial communities have not been well examined, despite their importance in the formation and maintenance of terrestrial ecosystems. Microorganisms are geographically versatile, exhibit wide morphological diversity and provide a rich platform for studying energy and carbon flows through different ecosystems. The coal characteristics, in turn, are important environmental factors that control the composition, structure and activity of terrestrial bio-communities through various endogenous physiological and biochemical processes. The total phylogenetic structure of prokaryotes is closely related to their functional diversity and, ultimately, to the variety of environmental conditions in oxidized coal (leonardite). Metagenomic studies in this area attempt to assess the relationship between the coal properties and its microbiome. The microbial community of the coal profiles, collected from various Kazakhstan coal deposits, have been studied in detail for the first time using high-throughput sequencing. As part of this study, a wide range of leonardites generated in various bioclimatic and geomorphological conditions are considered. A comprehensive characterization of the phylogenetic structure and diversity of coal was given on the basis of the 16S rDNA gene analysis. The revealed features of the prokaryotic composition can be used as bioindicators of the leonardite condition. In addition, metagenomic characteristics of coals of different origin can serve as valuable platform to assess the terrestrial ecosystem health.

General abbreviations

CTAB – cetyltrimethylammonium bromide
SDS – sodium dodecyl sulfate
PCR – polymerase chain reaction
FLASH – fast low angle shot
QIIME – quantitative insights into microbial ecology
OTU – operational taxonomic unit
PCA – principal component analysis
PCoA – principal coordinate analysis

Specific abbreviations

OLE – Oi-Karagai leonardite
LLE – Lenger leonardite
KLE – Kiyakty leonardite

1. Introduction

Microorganisms play a key role in the maintenance of aquatic and terrestrial ecosystems, participate in essential biogeochemical cycling of chemical elements in rocks, reservoirs, seals, etc. The recovered elements of rocks and ore deposits serve as a source of the vital energy for microorganisms. Biogeotechnology is a subset of knowledge, referring the use of microorganisms in the extraction, processing and enrichment of minerals, the recovery of metals from ores, etc.

Brown coals (lignite and leonardite) are heterogeneous organic substances composed of lignin and humic acids. The formation of brown coal occurred in Cainozoic, the Tertiary era and had undergone two phases. The first phase is a period of relatively rapid decomposition of unstable

*Corresponding author. E-mail: akimbeknur@gmail.com

substances of plant residues and the accumulation of stable compounds and products (lignin, cutin, suberin, etc.). The phase takes place under aerobic conditions due to biological activities. The second phase is the period of slow transformation of such stable compounds into even more resistant products, such as the conversion of readily soluble alkaline humic acids into insoluble humic substances of coal. This anaerobic phase is biogenous for peat and abiogenous for hard coal [1].

Despite the important role of microorganisms in the formation and deposition of lignite, there is insufficient available information regarding the occurrence, diversity and function of microbial communities in coal. In some studies the fungi with strongly pronounced enzymatic activity, such as *Phaenerochaete, Phlebia, Trametes, Bjerkandera, Nematoloma, Chrysonilia*, etc. were isolated from lignite. These fungi possess specific enzymes (laccase, peroxidase, manganese peroxidase, and hydrolases) that decompose mainly lignin, which is contained in brown coal [2, 3]. The partial degradation, i.e. depolymerization of lignite had occurred, when the above listed fungi incubated with brown coal. It has been suggested, that the microbial populations utilized humic acids present on the surface of lignite [4].

There is sporadic literature on brown coal liquefaction by *Pseudomonas* sp. [5] and anaerobic bacterial composition [6, 7]. Various aerobic bacteria groups (*Bacillus* sp., *Azotobacter* sp., *Myxobacteria*) also solubilize lignite by producing humic substances (Fig. 1) [8, 9].

Currently, the microbiology of coal is preferentially focused on the search for microorganisms with pronounced enzymatic activity, which could be used to purify low-rank coals from impurities for further application as chemical raw materials. Of equal interest are the study of microorganisms capable of transforming/converting humic acids that make up brown coal.

Only a small percentage of this microbial community can be isolated in laboratory conditions on artificial nutrient media. 90–99% of the total community composition of the prokaryotes that inhabit many diverse environments are unculturable, which means their biogeochemical and ecological functions cannot be appropriately examined by using standard microbiological techniques [10].

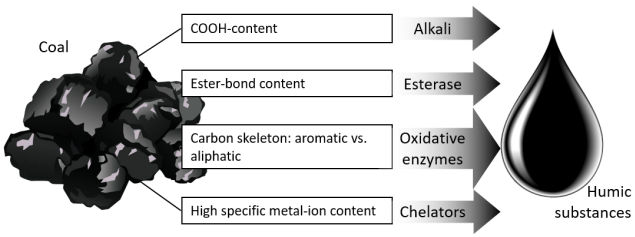

Fig. 1. Coal solubilization through various microbial agents.

Molecular-biological methods are effective for detection and assessment of the uncultivated majority of microorganisms, providing an opportunity to investigate the properties of microorganisms *in situ*, without isolation into pure cultures. Notable among them is metagenomics, the analysis of total genetic material released from the particular biological system [11]. The metagenomic approach has become possible due to the development of high-throughput sequencing, modern technologies of "reading" the nucleotide sequence of DNA, which allows analyzing large amounts of genetic information. The analysis of the 16S rDNA gene, the structure of which is based on the modern phylogenetic classification of prokaryotic organisms, is most popular in metagenomic studies.

Over the past decades, the structure and diversity of soil microbial communities and their relationship with environmental factors have been actively studied by the means of metagenomics. However, most metagenomics approaches relate only to soil microbiome. There are insufficient studies on environmental genetics (genecology) aimed at coal. Researches on ecological features of coal microbial communities and their interrelations with the properties of coal, consideration of a coal-soil profile seems reasonable and promising. The combination of modern methods of high-throughput sequencing with the classical techniques in soil science is a forward-looking way of solving global ecological problems of soil microorganisms. Especially, it is concerned with the relationship between the structure of microbial communities and the properties of terrestrial deposits, including coal.

The aim of this study is to carry out a comprehensive analysis of the phylogenetic structure and composition of microbiomes in Kazakhstan coal deposits and characterize their relation with the physical-chemical nature of leonardite.

2. Experimental

2.1. Sampling of leonardite from Kazakhstan coal deposits

Coal sampling was carried out according to ISO 18283: 2006 "Hard coal and co-sampling and ISO

13909-4: 2016 Preview Hard coal and co-mechanical sampling – Part 4: Coal – Preparation of test samples". The leonardite samples were collected from the following Kazakhstan coal deposits: the Oi-Karagai (Almaty region), which was designated in this study as "OLE", the Lenger (Turkestan region) – "LLE" and the Kiyakty (Karaganda region) – "KLE".

2.2. Ultimate and proximate analyses of leonardite samples

The elemental composition (hydrogen, carbon, nitrogen and sulfur) of leonardites was detected on a vario EL cube automatic analyzer (Elementar, Germany). The difference to 100% was assigned to the oxygen content. The following proximate characteristics were determined: moisture (W), ash content (A), calorific value (Q) and yield of volatile maters (V).

2.3. Metagenomics

2.3.1. Sequencing

Genomic DNA extraction. Total genomic DNA from samples was extracted according to CTAB and SDS protocols [12, 13]. The concentration and purity of DNA were estimated on 1% agarose gel. DNA concentration was adjusted to 1ng/μL by sterile water.

Amplicon Generation. The distinct regions of 16S rRNA genes (16SV4) were amplified using 16S V4: 515F-806R primer with the barcode. All PCR reactions were conducted with Phusion® High-Fidelity PCR Master Mix (New England Biolabs). The amplicons were generated from 25 PCR cycles. All amplicons were cleaned and sequenced according to the Illumina HiSeq 16S Metagenomic Sequencing Library Preparation protocol.

Quantification and qualification of PCR products. The same volume of 1X loading buffer (SYBR Green®) was mixed with PCR products and electrophoresis on 2% agarose gel was operated. The bright strips between 400–450 bp were selected for further procedures.

PCR Products Mixing and Purification. PCR products was compounded in equidensity ratios and purified with Qiagen Gel Extraction Kit (Qiagen, Germany).

Library preparation and sequencing. Sequencing libraries were prepared using TruSeq® DNA PCR-Free Sample Preparation Kit (Illumina, USA) following manufacturer's recommendations. Its quality was assessed on the Agilent Bioanalyzer 2100 and Qubit@ 2.0 Fluorometer (Thermo Scientific) systems. Finally, the library was sequenced on an IlluminaHiSeq2500 platform and 250 bp paired-end reads were generated.

2.4. Data analysis

Paired-end reads assembly and data split. Paired-end reads were assigned in accordance to their unique barcode by cutting off the barcode and primer sequence.

Sequence assembly. Paired-end reads were merged using FLASH [14], which was designed to merge paired-end reads.

Data Filtration. Quality filtering was performed under specific filtering conditions to obtain the high-quality clean tags [15] by means of the QIIME [16].

Chimera removal. The tags were compared with the reference database using UCHIME algorithm [17] to detect chimera sequences with the following removal [18].

OTU Production. Sequences analysis were performed using Uparse software [19]. Sequences with ≥97% similarity were assigned to the close OTUs.

Species annotation. The GreenGene Database [20] was applied for each representative sequence, based on RDP classifier algorithm to annotate taxonomic information [21].

Phylogenetic relationship construction. Multiple sequence alignment were carried out using the MUSCLE software [22] in order to evaluate phylogenetic relationship of different OTUs, as well as the difference of the dominant species in different groups.

Alpha Diversity. Alpha diversity is used in analyzing complexity of species diversity through 6 indices (Observed-species, Chao1, Shannon, Simpson, ACE, Good-coverage). All these indexes were calculated with QIIME and displayed with R software.

Beta Diversity. Beta diversity analysis was applied to assess differences of samples in species complexity. The diversity on both weighted and unweighted UniFrac were calculated by QIIME software. Cluster analysis was performed by principal component analysis (PCA), which was applied to reduce the dimension of the original variables. PCoA was used to get principal coordinates and visualize from multidimensional data.

3. Results and discussion

The proximate and ultimate analyses of OLE, LLE and KLE were carried out at the laboratory base; the results of which are presented in Table 1. The moisture content of all the coal samples studied is considerable, i.e. within the range of 9 to 12%. Such tendency suggests that coal has a high propensity to drain in natural conditions. The ash content of the OLE and KLE samples falls within a narrow range of ~11%, which makes it possible to classify these coals to the group of medium-ash, and the LLE to high-ash with 22%.

The carbon concentration in the dry ash-free mass of oxidized coal reaches 75%. The percentage of hydrogen, nitrogen and sulfur was significantly less, while oxygen content was higher on average. However, it is necessary to pay attention to the sulfur content, which in the LLE reaches 1.65%.

For the first time, a comparative metagenomic analysis of oxidized coal samples was carried out. Despite the fact that coal basins occupy vast territories of Kazakhstan, there are no studies on metagenomic analysis of terrestrial ecosystems. Characteristics of the primary data of the LLE, OLE and KLE are given in Table 2.

The bacteria communities of leonardite samples are composed primarily by the phyla of *Proteobacteria*, *Tenericutes*, *Acidobacteria*, *Firmicutes*, *Bacteroidetes*, *Nitrospirae*, *Chloroflexi*, *Gemmatimonadetes*, *Actinobacteria* and *Fusobacteria*. *Proteobacteria* phylum has the largest number,

Table 1
Results of proximate and ultimate analyses (W%) of leonardite samples*

Parameters	Designation	Samples		
		LLE	OLE	KLE
Proximate analysis				
Moisture	W	9.1	11.8	9.8
Ash	A	22.0	12.2	11.5
Volatile matter	V	40.8	35.8	41.8
Calorific value, MJ/kg	Q	7.8	15.6	21
Ultimate analysis				
	C	41.61	75.00	74.50
	H	1.60	4.81	4.12
	N	0.86	1.50	0.74
	S	1.65	0.41	0.75
	$O^{diff.}$	54.28	18.28	19.89

*Mean values ± standard deviation (n = 5)

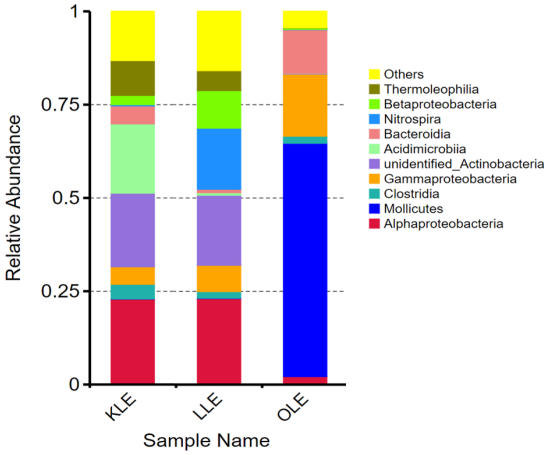

Fig. 2. Taxonomic structure of microbial communities at the level of class.

often occupying a dominant position in terrestrial microbiome. In addition to this group, a significant proportion of the KLE biome is made up of representatives of *Actinobacteria* groups. Many *Actinobacteria*, especially their mycelial species – actinomycetes – are adapted to habitats with low humidity. However, in the OLE sample, the bacteria of the phylum *Tenericutes* are superior, and *Actinobacteria* had the second highest rate. The most widespread class for the KLE and the OLE were *Alphaproteobacteria* with many important roles in various ecosystems (Fig. 2).

The most abundant family for LLE was *Nitrospiraceae*, which is the only in the phylum *Nitrospirae* and comprises the genera *Nitrospira*, *Leptospirillum* and *Thermodesulfovibrio* (Fig. 3). The family is physiologically highly diverse; however, in our case it contained mostly *Leptospirillum* (99%), which are chemolithoautotrophic aerobic

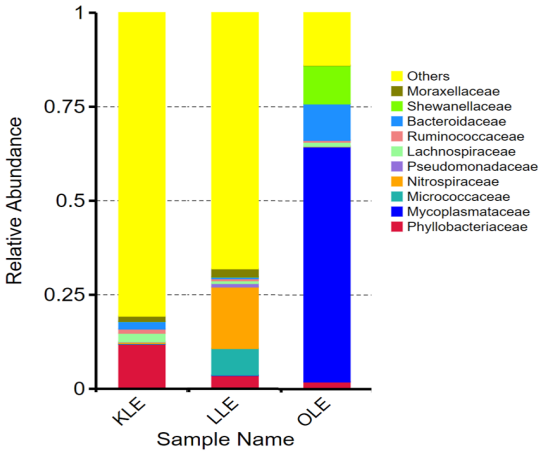

Fig. 3. Taxonomic structure of microbial communities at the level of family.

Table 2
Tags in the OTU clustering analysis

Samples	Total Tags	Taxon Tags	Unclassified Tags	Unique Tags	OTU
LLE	78590	76929	1	1660	1212
OLE	81889	80420	0	1469	682
KLE	77015	75777	40	1198	1019

and acidophilic ferrous iron oxidizers. The members of *Leptospirillum* evolved in the bioleaching of metal ores [23]. The second most abundant family identified was *Micrococcaceae* with 5 genera: *Arthrobacter* (64%), *Micrococcus* (34%), *Rothia* (1%), *Paenarthrobacter* (1%) and *Nesterenkonia* (1%). Micrococci, like many other representatives of the *Actinobacteria*, are catabolically versatile, with the potential to utilize a wide range of xenobiotics and tolerate metals. They are also involved in detoxification or biodegradation of many environmental contaminants [24].

One of the most abundant family for the OLE was *Shewanellaceae* within the order *Alteromonadales*, compose of a sole genus *Shewanella*. Facultative anaerobic. Some species are known to have unique metabolic properties, such as dissimilatory reduction of metal compounds [25].

The family *Phyllobacteriaceae* was the most abundant in the KLE samples; it mostly composed of the genus *Phyllobacterium* (80%). The members of this group were isolated from root nodules and surfaces of plants, as well as rocks and tuff stone [26].

The diversity of the prokaryotic community, according to the number of found OTE and various indices (Table 3) in given samples are almost heterogeneous due to the physical and chemical nature of leonardites.

Physical parameters and chemical contents of coal may exert influence on microbial diversity and survival. Higher levels of bacterial diversity and richness can be observed in low-rank coals, in a way that oxygen content could be one of the relevant predictor of coal bacterial diversity. Total carbon and nitrogen content might also influence the diversity of coal microbiome, including nitrogen-fixing bacteria. For instance, families *Nitrosomonadaceae* and *Phyllobacteriaceae* were detected in the KLE, as well as in the OLE.

Metagenomics revealed that the LLE held relatively high-diversity microbial communities supported by sulfur-based chemolithoautotrophy. Sulfur-sulfate reducing genera *Desulfovibrio* and H16 (*Desulfurellaceae* family) as well as sulfur-oxidizing bacteria *Thiothrix* and *Sulfurifustis* were found in this coal sample (Fig. 4). Considering the fact that the LLE contains high elemental sulfur content (1.65%), it can serve as a rich source of energy.

Principal component analysis (PC1 = 25.61%, PC2 = 37.15%) is the way to observe taxonomic relatedness between the given microbiomes (Fig. 5). The more similar the community composition of the given samples, the closer they are in the PCA plot. As observed, bacterial sample communities from the KLE clustered very close to the OLE. The LLE locality was the most distinctive

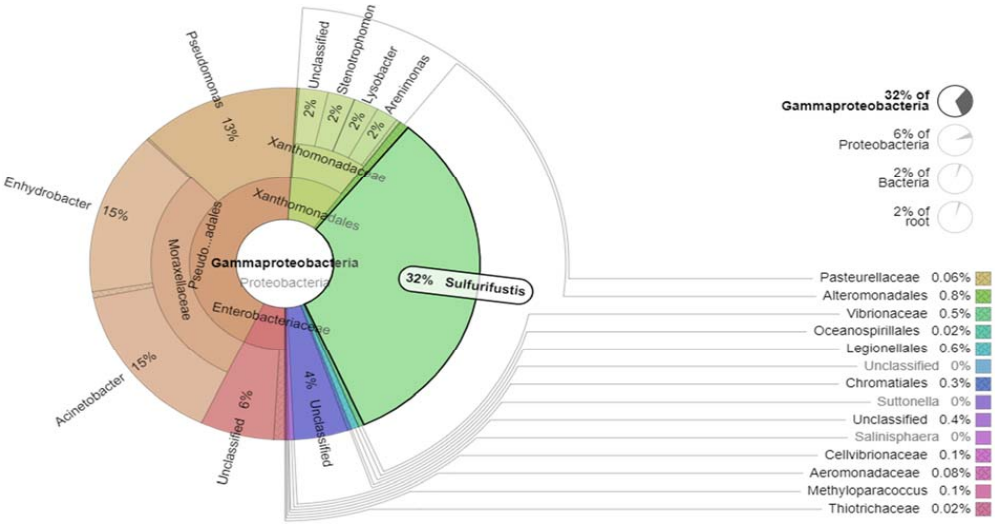

Fig. 4. Krona taxonomy web visualization illustrated by the case of Sulfurifustis in the LLE sample.

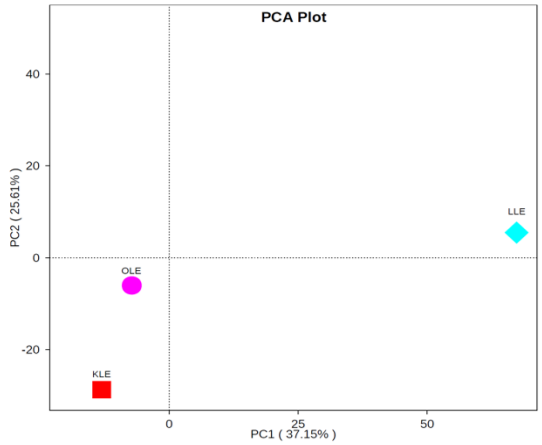

Fig. 5. Results of PCA analysis of coal samples.

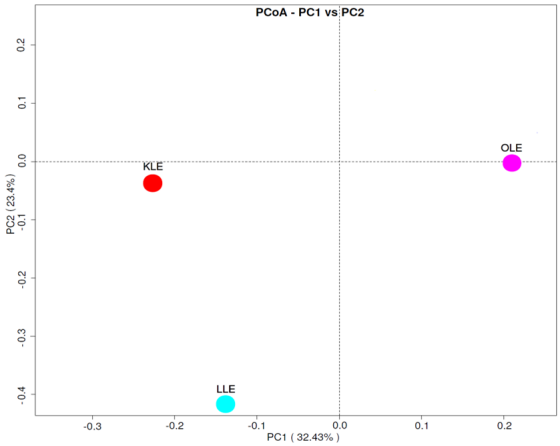

Fig. 6. PCoA on unweighted UniFrac distances between coal samples.

group, likely due to its most different parameter and reduced carbon content.

Principal coordinate analysis (PCoA) on unweighted UniFrac distances revealed that the microbial communities of the samples separated into distinct clusters (Fig. 6). The ordinate PCoA1 explains 32.43% of the variation and separates the microbiome of samples, whereas the PCoA2 explains 23.4%.

Despite the fact that all coal samples were at the same category of leonardite, PCA and PCoA were able to separate the regions into distinct clusters according to coal characteristics.

4. Conclusions

The performed analysis allowed characterizing the structure and diversity of microbial communities in accordance with coal properties and genesis. The isolates identified for three samples were assigned to 29 genera belonging to 10 families. The families *Phyllobacteriaceae*, *Nitrospiraceae* and *Mycoplasmataceae* associated with chemosynthetic and heterotrophic lifestyles were dominant microbial communities in the given leonardite samples. The observed functional diversity of microorganisms affects coal ecosystem dynamics, stability and nutritional balance. The research on non-cultivable bacteria can contribute to the examination and analysis of the entire structure and composition of coal microbiome. It is shown that samples that are more differentiated in their physicochemical properties are also more variable in composition and diversity of microbial communities. Microbiomes of the coals, according to the beta-diversity analysis, generate non-intersecting clusters when examine at the family level and lower taxonomic levels. A comparative analysis of bacteria groups in different leonardites allows one to judge about their ecological requirements as well as to provide information about their functional features by means of deep sequencing of the genomes.

References

[1]. S.L. Bend, *Fuel* 71 (1992) 851–870. DOI: 10.1016/0016-2361(92)90234-F

[2]. M.J. Ghani, M.I. Rajoka, K. Akhtar, *Biotechnol. Bioproc. E* 20 (2015) 634–642. DOI: 10.1007/s12257-015-0162-5

[3]. D.E.A. Catcheside, J.P. Ralph, Bioconversion of Coal by Fungi BT – Industrial Applications, in: H.D. Osiewacz (Ed.), Springer Berlin Heidelberg, Berlin, Heidelberg, 2002. P. 343–354. DOI: 10.1007/978-3-662-10378-4_17

[4]. R.M. Fakoussa, M. Hofrichter, *Appl. Microbiol. Biot.* 52 (1999) 25–40. DOI: 10.1007/s002530051483

[5]. A.Y. Malik, M. Ali, A. Jamal, M.I. Ali, *Geomicrobiol. J.* 34 (2017) 109–118. DOI: 10.1080/01490451.2016.1143538

[6]. A.J. Grethlein, M.K. Jain, *Trends Biotechnol.* 10 (1992) 418–423. DOI: 10.1016/0167-7799(92)90290-C

[7]. E.P. Barnhart, E.P. Weeks, E.J.P. Jones, D.J. Ritter, J.C. McIntosh, A.C. Clark, L.F. Ruppert, A.B. Cunningham, D.S. Vinson, W. Orem, M.W. Fields, *Int. J. Coal Geol.*162 (2016) 14–26. DOI: 10.1016/j.coal.2016.05.001

[8]. I. Romanowska, B. Strzelecki, S. Bielecki, *Fuel Process. Technol.* 131 (2015) 430–436. DOI: 10.1016/j.fuproc.2014.12.019

[9]. Y. David, M.G. Baylon, S.D.V.N. Pamidimarri, K.A. Baritugo, C.G. Chae, Y.J. Kim, T.W. Kim, M.S. Kim, J.G. Na, S.J. Park, *Biotechnol. Bioproc. E.* 22 (2017) 178–185. DOI: 10.1007/s12257-016-0263-9

[10]. C. Sergaki, B. Lagunas, I. Lidbury, M.L. Gifford, P. Schafer, *Front. Plant Sci.* 9 (2018) 1205. DOI: 10.3389/fpls.2018.01205
[11]. J.K. Jansson, K.S. Hofmockel, *Curr. Opin. Microbiol.* 43 (2018) 162–168. DOI: 10.1016/j.mib.2018.01.013
[12]. M.A. Lever, A. Torti, P. Eickenbusch, A.B. Michaud, T. Santl-Temkiv, B.B. Jorgensen, *Front. Microbiol.* 6 (2015) 476. DOI: 10.3389/fmicb.2015.00476
[13]. K. Minas, C.J. Newbold, N.R. McEwan, K.P. Scott, *FEMS Microbiol. Lett.* 325 (2011) 162–169. DOI: 10.1111/j.1574-6968.2011.02424.x
[14]. T. Magoc, S.L. Salzberg, *Bioinformatics* 27 (2011) 2957–2963. DOI: 10.1093/bioinformatics/btr507
[15]. N.A. Bokulich, S. Subramanian, J.J. Faith, D. Gevers, J.I. Gordon, R. Knight, D.A. Mills, J.G. Caporaso, *Nat. Methods* 10 (2013) 57–59. DOI: 10.1038/nmeth.2276
[16]. J.G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F.D. Bushman, E.K. Costello, N. Fierer, A.G. Pena, J.K. Goodrich, J.I. Gordon, *Nat. Methods* 7 (2010) 335–336. DOI: 10.1038/nmeth.f.303.
[17]. R.C. Edgar, B.J. Haas, J.C. Clemente, C. Quince, R. Knight, *Bioinformatics* 27 (2011) 2194–2200. DOI: 10.1093/bioinformatics/btr381.
[18]. B.J. Haas, D. Gevers, A.M. Earl, M. Feldgarden, D.V. Ward, G. Giannoukos, D. Ciulla, D. Tabbaa, S.K. Highlander, E. Sodergen, *Genome Res.* 21 (2011) 494–504. DOI: 10.1101/gr.112730.110.
[19]. R.C. Edgar, *Nat. Methods* 10 (2013) 996–998. DOI: 10.1038/nmeth.2604.
[20]. T.Z. DeSantis, P. Hugenholtz, N. Larsen, M. Rojas, E.L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, G.L. Andersen, *Appl. Environ. Microb.* 72 (2006) 5069–5072. DOI: 10.1128/AEM.03006-05.
[21]. Q. Wang, G.M. Garrity, J.M. Tiedje, J.R. Cole, *Appl. Environ. Microb.* 73 (2007) 5261–5267. DOI: 10.1128/AEM.00062-07.
[22]. R.C. Edgar, *Nucleic Acids Res.* 32 (2004) 1792–1797. DOI: 10.1093/nar/gkh340.
[23]. H. Daims, The Family Nitrospiraceae BT – The Prokaryotes: Other Major Lineages of Bacteria and The Archaea, in: E. Rosenberg, E.F. DeLong, S. Lory, E. Stackebrandt, et al. (Eds.), Springer Berlin Heidelberg, Berlin, Heidelberg, 2014. P. 733–749. DOI: 10.1007/978-3-642-38954-2_126.
[24]. S.G. Dastager, S. Krishnamurthi, N. Rameshkumar, M. Dharne, The Family Micrococcaceae BT – The Prokaryotes: Actinobacteria, in: E. Rosenberg, E.F. DeLong, S. Lory, E. Stackebrandt, et al. (Eds.), Springer Berlin Heidelberg, Berlin, Heidelberg, 2014. P. 455–498. DOI: 10.1007/978-3-642-30138-4_168.
[25]. M. Satomi, The Family Shewanellaceae BT – The Prokaryotes: Gammaproteobacteria, in: E. Rosenberg, E.F. DeLong, S. Lory, E. Stackebrandt, et al. (Eds.), Springer Berlin Heidelberg, Berlin, Heidelberg, 2014. P. 597–625. DOI: 10.1007/978-3-642-38922-1_226.
[26]. A. Willems, The Family Phyllobacteriaceae BT – The Prokaryotes: Alphaproteobacteria and Betaproteobacteria, in: E. Rosenberg, E.F. DeLong, S. Lory, E. Stackebrandt, et al. (Eds.), Springer Berlin Heidelberg, Berlin, Heidelberg, 2014. P. 355–418. DOI: 10.1007/978-3-642-30197-1_298.