Panda et al. Big Data Analytics (2017) 2:7
DOI 10.1186/s41044-017-0022-8

Big Data Analytics

RESEARCH Open Access

Bacterial diversity analysis of Yumthang hot spring, North Sikkim, India by Illumina sequencing

Amrita Kumari Panda[1*], Satpal Singh Bisht[1], Bodh Raj Kaushal[1], Surajit De Mandal[2], Nachimuthu Senthil Kumar[2] and Bharat C. Basistha[3]

* Correspondence: itu.linu@gmail.com
[1]Department of Zoology, Kumaun University, Nainital, Uttarakhand 263002, India
Full list of author information is available at the end of the article

Abstract

Background: Hot springs harbor rich bacterial diversity that could be the source of commercially important enzymes, antibiotics and many more products. Most of the hot springs present in Northeast of India are unexplored and their microbial diversity analysis could be of great interest to facilitate various industrial, agricultural and medicinal applications. The present study is an attempt to analyze the comprehensive bacterial diversity of Yumthang hot spring, Sikkim located at an altitude of 11, 800 ft. with a close proximity of Tibet 27° 47′ 30″ N 88° 42′ E using culture independent approach i.e. 16S rRNA gene amplicon metagenomic sequencing.

Results: The temperature and pH of the hot spring was recorded as 39^0–41^0 C and 8 respectively. Metagenome comprised of 1, 381,343 raw sequences with a sequence length of 151 bp and 55.62% G + C content. Metagenome sequence information is submitted at NCBI, SRA database under accession no. SRP057072. A total of 9, 95, 955 pre-processed reads were clustered into 1, 999 representative OTUs (operational taxonomical units) phylogenetically comprising of 17 bacterial phyla including unknown phylum indicating 99 families. Hot spring bacterial community is dominated by *Proteobacteria* (54.33%), *Actinobacteria* (32.19%), *Firmicutes* (6.03%), *Bacteroidetes* (2.87%) and unclassified bacteria (2.91%) respectively out of the total reads.

Conclusions: Several bacterial and archaeal sequences remained taxonomically unclassified, indicating potentially novel microorganisms in this hot spring ecosystem. Metagenomics of this habitat will facilitate identification of microorganisms possessing industrially relevant traits.

Keywords: Bacterial diversity, Illumina, Sikkim, Hot springs, 16S rDNA

Background

Geologists have spotted and studied many thermal springs in the various regions of Indian subcontinent [1–3]. However, their microbial diversity has not been fully explored by employing modern molecular phylogenetic techniques. The present study reveals information on the bacterial community structure of Yumthang hot water spring, Sikkim, India. Illumina platform was used to sequence V3 hyper-variable region of 16S rDNA from microbial mat metagenome to profile the microbial community of this Northeastern hot spring of India. T1

The thermal springs of Sikkim are scattered in the Himalayan geothermal province. There are numerous natural hot springs at Sikkim; located at various locations like Polok, Reshi, Borong, Takrum, Yume Samdong, Yumthang, Zee, Shagyong Phedok and Tholug Kang of Sikkim [4]. Hot water springs are sign of geological activity and represent extreme environment. The Yumthang hot spring (27° 47′ 30″ N 88° 42′ E) is one of the less explored hot spring in Sikkim, India (Fig. 1). This hot spring is located at the Lachung river bed resulting in mixing of river water as soon as the hot water comes to the surface. This causes difficulty in measuring the hot water temperature at the site of emergence. Microbial mat along with water and sediment samples were collected in March 2014 using a hand trowel and pooled into sterile tubes, frozen in dry ice and transported to the laboratory for further analysis. Prior to sampling, the temperature and pH was measured at the precise sampling locations and recorded as 39°–41 °C and 8–8.5 respectively. The lithology of Yumthang, Sikkim is mostly composed of high grade gneisses Darjeeling gneiss and Kanchendzonga gneiss. The gneisses dominantly comprises of quartz, feldspar and biotite with minor amounts of other minerals.

Methods

Metagenomic DNA extraction

In the present investigation, the total community DNA was isolated from the microbial mat samples of the spring using FastDNA spin kit (MP Biomedicals, LLC, USA) according to the manufacturer's protocol with some modifications to increase the yield and purity of the extracted DNA sample. The modifications included the addition of 300 μl of PPS (protein precipitation solution) to remove the protein contamination and

Fig. 1 Location of Yumthang hot spring, Sikkim, India (Indo-Tibetan plateu)

the binding matrix pellet was re-suspended in 25 μl of DES (DNase/ Pyrogen-Free Water) to avoid over-dilution of the purified DNA. Final DNA concentration was quantified by microplate reader (BMG Labtech, Jena, Germany).

Illumina sequencing

The V3 hypervariable region of the 16S rRNA gene was amplified using 341F/ 518R primer combination 5'CCTACGGGAGGCAGCAG 3′ and 5'ATTACCGCGGCTGCTGG 3′ [5]. PCR Master Mix will contain 2 μL each primers, 0.5 μL of 40 mM dNTP (NEB, USA), 5 μL of 5X Phusion HF reaction buffer (NEB, USA), 0.2 μL of 2 U/μL F-540Special Phusion HS DNA Polymerase (NEB, USA), 5 ng input DNA and water to make up the total volume to 25 μL. Cycling conditions for the PCR reaction were 98 °C for 30 s, followed by 30 cycles of 98 °C for 10 s and 72 °C for 30 s, with a 5 s elongation step at 72 °C followed by 4 °C hold. Amplicon was excised and purified by QIA quick Gel Extraction kit (Qiagen, Valencia, CA, USA) according to the manufacturer's manual. Purified amplicon was paired-end sequenced (2 X 151 base pairs) on an Illumina Mi-Seq platform at Scigenome India Pvt. Ltd., Cochin, India.

Phylogenetic analysis

QIIME data analysis package was used for 16S rRNA data analysis [6]. Quality check on raw sequences was performed as per base quality score distributions, average base content per read and GC distribution in the reads. Singletons, the unique OTU that did not cluster with other sequences, were removed as it might be a consequence of sequencing errors and can result in spurious OTUs. Chimeras were also removed using UCHIME, pre- processed consensus V3 sequences were grouped into operational taxonomic units (OTUs) using the clustering program UCLUST at a similarity threshold of 0.97 [7, 8]. All the pre-processed reads were used to identify the OTUs using QIIME program for constructing a representative sequence for each OTU. The representative sequence was finally aligned against Greengenes core set of sequences using PyNAST program [9]. Representative sequence for each OTU was classified using RDP classifier and Greengenes OTUs database and the sequences those not classified were categorized as unknown.

Results and discussion

Phylum abundance results showed *Proteobacteria* to be abundant in the amplicon library. In the sample studied, 1, 999 OTUs represented 17 distinct phyla dominated by *Proteobacteria* (54.33%), *Actinobacteria* (32.19%), *Firmicutes* (6.03%), *Bacteroidetes* (2.87%) and unclassified bacteria (2.91%) respectively based on the total no. of reads (Table 1, Fig. 2a). Out of the total of 9, 95,955 pre-processed reads 49.80% are from the class *Betaproteobacteria* (Fig. 2b). The order *Rhodocyclales* of *Betaproteobacteria* constitutes 40.81% of the sequence reads followed by 40.81% of the family *Rhodocyclaceae*. Members under this family were mainly aerobic or denitrifying rod-shaped bacteria with diverse metabolic activities and survive in oligotrophic conditions such as aquatic habitats using photoautotrophic carbon fixation process [10, 11].

The phylum to genus level bacterial diversity identified in the high throughput 16S rRNA libraries from YM1 (Yumthang) hot spring is presented in Fig. 3. *Proteobacteria* constituted 671 OTUs i.e. 33.56% of total OTUs and 5, 41,190 reads which are 54.33%

Table 1 Top ten OTU's based on total read count number in the hot spring sample

OTU Table Id	Read count	Phylum	Class	Order	Family	Genus
Denovo 131	383,815	*Proteobacteria*	*Betaproteobacteria*	*Rhodocyclales*	*Rhodocyclaceae*	—
Denovo 817	167,211	*Actinobacteria*	*Actinobacteria*	*Actinomycetales*	*Nocardiaceae*	*Rhodococcus*
Denovo 992	67,431	*Actinobacteria*	*Actinobacteria*	*Actinomycetales*	—	—
Denovo 1092	66,706	*Actinobacteria*	*Actinobacteria*	*Actinomycetales*	—	—
Denovo 419	40,749	*Firmicutes*	*Bacilli*	*Bacillales*	*Bacillaceae*	—
Denovo 2491	30,883	*Proteobacteria*	*Betaproteobacteria*	*Hydrogenophilales*	*Hydrogenophilaceae*	*Thiobacillus*
Denovo 797	19,836	*Proteobacteria*	*Betaproteobacteria*	*Rhodocyclales*	*Rhodocyclaceae*	—
Denovo 949	17,357	*Proteobacteria*	*Betaproteobacteria*	*Thiobacterales*	—	—
Denovo 1957	15,452	*Proteobacteria*	*Betaproteobacteria*	*Burkholderiales*	*Oxalobacteraceae*	—
Denovo 378	11,730	*Bacteroidetes*	*Bacteroidia*	*Bacteroidales*	GZKB119	—

of the total reads. There are studies supporting and substantiating findings of the present study that bacterial communities in the hot springs located on the high altitude are dominated by *Proteobacteria* [12]. Similar observations have been made by [13, 14] that geographic locations play a significant role in bacterial community structure. *Proteobacteria* has also been reported from many studies based on the 16SrRNA analysis of hot springs with moderately high and very high temperatures (44–110 °C) at various geographical locations, including India [15–17]. Since, *Proteobacteria* have been found in other hot spring studies including Indian hot springs, and was also one of the abundant taxa in this study, it appears to be indigenous to this region.

The other dominant OTUs within *Betaproteobacteria* were OTU 395, 2102 and 501 classified under the order *Hydrogenophilales, Thiobacterales* and *Burkholderiales* respectively (data not shown). Two represented genus under the order *Hydrogenophilales* were *Hydrogenophilus* and *Thiobacillus*. *Hydrogenophilus* sp. are thermophilic, growing around 50 °C and obtaining their energy from oxidizing hydrogen. However they were previously isolated from an ice layer covering Lake Vostok in Antarctica which indicates the possibility of a geothermal system exists beneath the cold water body. The other genus *Thiobacillus* can oxidise the sulfur to sulfuric acid and widely used as a pesticide against pest potato scabs [18]. *Thiobacillus* sp., belonging to phylum *Proteobacteria* observed as abundant, has been reported from Yumthang in an earlier study [19]. Three thousand seventy two V3 16S rDNA reads from the amplicon library share

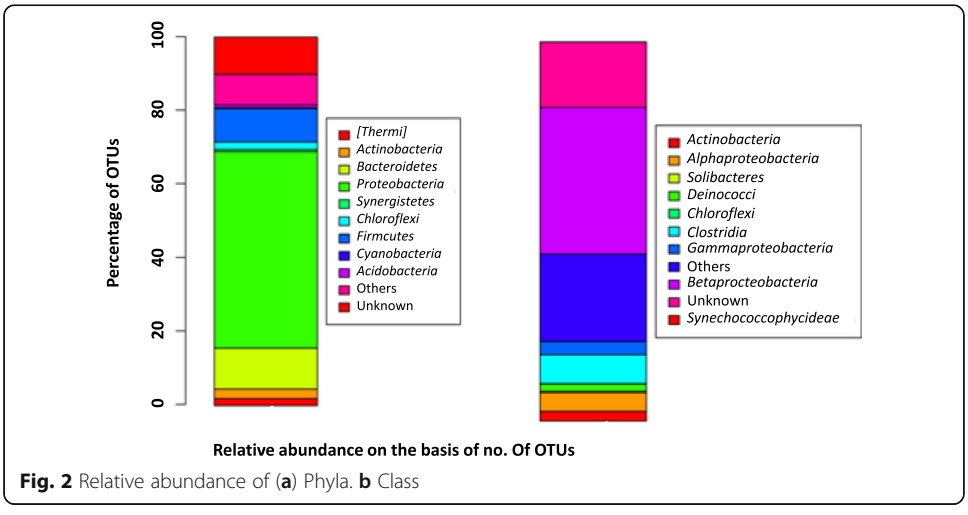

Fig. 2 Relative abundance of (**a**) Phyla. **b** Class

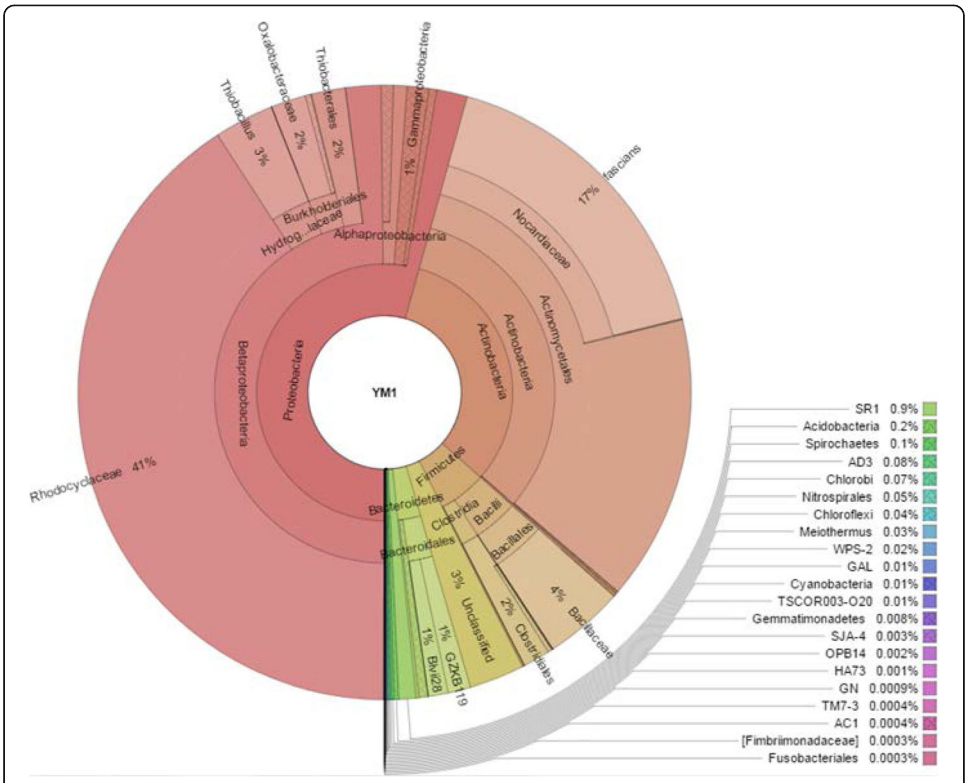

Fig. 3 Microbial community composition of Yumthang hot spring, Sikkim, India by 16S amplicon library sequencing

more than 97% identity with members of sulfate reducing *Desulfomicrobium* sp. isolated from low temperature anaerobic enrichment culture from oil reservoir production water, China. This study also identified few *Roseococcus* and *Alteromonas* reads those are poorly described from other thermal environments. *Alteromonas* is reported to be a candidate genus for exopolysaccharide production [20, 21]. *Thiovirga sulfuroxydans* gen. Nov., sp. nov., a chemolithoautotrophic sulfur-oxidizing bacterium isolated from a microaerobic waste-water biofilm [22]. This study also identified few *Thiovirga* OTUs in amplicon library. It is the first description worldwide in association with hot springs. The genus *Psychrobacter*, a member of the class Gammaproteobacteria, is predominantly isolated from cold and/or saline environments, such as Arctic permafrost, Antarctic ice pack, estuaries, and marine fish, including Korean fermented seafood [23–29]. The occurrence of *Psychrobacter* OTUs in the amplicon library is of interest as there are no reports from hot springs all over the World.

Actinobacteria constitutes 351 OTUs i.e. 17.55% of total OTUs and 3, 20, 616 reads (32.19%) of total reads whereas 536 OTUs i.e. 26.81% of total OTUs belong to the unknown phylum. *Actinobacteria* edged over other microbes as a prolific producer of antibiotics and other biopharmaceuticals [30]. Thermophilic *actinobacteria* are biotechnologically important producers of several enzymes such as DNA polymerases, pullulanases, amylases, xylanases, lipases and proteases [31]. However, little is known about the distribution and biogeography of *Actinobacteria* in hot springs. The present findings revealed the presence of large number of sequence reads of *Rhodococcus* from bacterial phylum *Actinobacteria* which may embody many novel species within this industrially important genus [32, 33].

The representatives of this genus have been reported from varied environments viz. soil, sewage treatment plants, polluted and unpolluted water bodies etc. [32], *Rhodococcus* has been also reported from alkaline hot springs of the world [34]. Strains of *Rhodococcus* are well known microbes carrying out biologically relevant reactions such as desulfurization of fossil fuels, degradation of pollutants, biosurfactants and bioflocculants etc. [33].

Conclusions

The Yumthang hot spring of Indo-Tibetan plateau is home to many possibly unknown and novel microbes as indicated by the presence of 26.81% unknown OTUs out of 1, 999 OTUs.

Acknowledgements
The authors are grateful to the Head Department of Zoology, Kumaun University, Nainital, India for providing basic infrastructural facility and administrative assistance.

Funding
Financial support was provided by SERB, Govt. of India Vide Project SB/FT/LS-335/2012 for the design of the study, sample collection and analysis of data.

Availability of data and materials
Sequence data that support the findings of this study have been deposited in the Sequence Read Archive (SRA) service of the National Centre for Biotechnology Information (NCBI) database under the accession number SRP057072 https://www.ncbi.nlm.nih.gov/sra/?term=SRP057072.

Authors' contributions
AKA conceived organized and wrote the paper. AKA and SDM analyze the data; SSB, BRK and NSK critically analyzed the study and helped in drafting the article as well as edited the manuscript. AKA obtained funding for the original project idea. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

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

Author details
[1]Department of Zoology, Kumaun University, Nainital, Uttarakhand 263002, India. [2]Department of Biotechnology, Mizoram University, Aizawl, Mizoram 796004, India. [3]Sikkim State Council of Science & Technology, Vigyan Bhawan, Deorali, Gangtok, Sikkim 737102, India.

Received: 25 October 2016 Accepted: 6 August 2017
Published online: 20 August 2017

References

1. Chatterjee GC, Guha SK. The problems of origin of high temperature springs of India, In: abstracts of the 23rd international geological congress, vol. 17; 1968. p. 141–9.
2. Gupta ML, Narain H, Saxena VK. Geochemistry of thermal waters from various geothermal provinces of India. In: Abstracts of the *Grenoble Symposium of the International Association of Hydrological Sciences*. Wallingford: Institute of Hydrology, IAHS Press; 1975. p. 47–58.
3. Ghosh D, Bal B, Kashyap VK, Pal S. Molecular phylogenetic exploration of bacterial diversity in a Bakreshwar (India) hot spring and culture of *Shewanella*-related thermophiles. Appl Environ Microbiol. 2003;69(7):4332–6.
4. Sherpa MT, Das S, Thakur N. Physicochemical analysis of hot water springs of Sikkim-Polok Tatopani, Borong Tatopani and Reshi Tatopani. Recent Res Sci Technol. 2013;5(1):63–7.
5. Li H, Zhang Y, Li D, Xu H, Chen G, Zhnag C. Comparisons of different hypervariable regions of rrs genes for fingerprinting of microbial communities in paddy soils. Soil Biol Biochem. 2009;41:954–8.
6. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon J.I et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 2010; 7: 335-336.
7. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26(19):2460–1.
8. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics. 2011;27(16):2194–200.
9. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol. 2006;72:5069–72.

10. Brenner DJ, Krieg NR, editors. Bergey's manual® of systematic bacteriology, The Proteobacteria, vol. Volume Two. Berlin: Springer Science & Business Media; 2006.
11. Badhai J, Ghosh TS, Das SK. Taxonomic and functional characteristics of microbial communities and their correlation with physicochemical properties of four geothermal springs in Odisha. India Frontiers in microbiology. 2015;6:1166. doi:10.3389/fmicb.2015.01166.
12. Huang Q, Dong CZ, Dong RM, Jiang H, Wang S, et al. Archaeal and bacterial diversity in hot springs on the Tibetan plateau, China. Extremophiles. 2011;15:549–3.
13. Kumar M, Nath Yadav A, Tiwari R, Prasanna R, Saxena AK. Evaluating the diversity of culturable thermotolerant bacteria from four hot springs of India. J Biodivers Biopros Dev. 2014;1:127. doi:10.4172/2376-0214.1000127.
14. Whitaker RJ, Grogan DW, Taylor JW. Geographic barriers isolate endemic populations of hyperthermophilic archaea. Science. 2003;301:976–8.
15. Bowen De Leon K, Gerlach R, Peyton BM, Fields MW. Archaeal and bacterial communities in three alkaline hot springs in heart Lake Geyser Basin, Yellowstone national park. Front Microbiol. 2013;4:330. doi:10.3389/fmicb.2013.00330.
16. Chan CS, Chan KG, Tay YL, Chua YH, Goh KM. Diversity of thermophiles in a Malaysian hot spring determined using 16SrRNA and shot gun metagenome sequencing. Front Microbiol. 2015;6:177. doi:10.3389/fmicb.2015.00177.
17. Ghelani A, Patel R, Mangrola A, Dudhagara P. Cultivation- independent comprehensive survey of bacterial diversity in Tulsi Shyam Hot Springs, India. Genome Data. 2015;4:54–6. doi:10.1016/j.gdata.2015.03.003.
18. Bulat SA, Alekhina IA, Blot M, Petit JR, De Angelis M, Wagenbach D, Lipenkov VY, Vasilyeva LP, Wloch DM, Raynaud D, Lukin VV. DNA signature of thermophilic bacteria from the aged accretion ice of Lake Vostok, Antarctica: implications for searching for life in extreme icy environments. Int J Astrobiol. 2004;3(01):1–12.
19. Panda AK, Bisht SS, Mandal S, Kumar NS. Bacterial and archeal community composition in hot springs from indo-Burma region, north-east India. AMB Express. 2016;6(1):111.
20. Nicolaus B, Kambourova M, Oner ET. Exopolysaccharides from extremophiles: from fundamentals to biotechnology. Environ Technol. 2010;31(10):1145–58.
21. Mata JA, Bejar V, Bressollier P, et al. Characterization of exopolysaccharides produced by three moderately halophilic bacteria belonging to the family Alteromonadaceae. J Appl Microbiol. 2008;105(2):521–8.
22. Ito T, Sugita K, Yumoto I, Nodasaka Y, Okabe S. *Thiovirga sulfuroxydans* gen. Nov., sp. nov., a chemolithoautotrophic sulfur-oxidizing bacterium isolated from a microaerobic waste-water biofilm. Int J Syst Evol Microbiol. 2005;55:1059–64.
23. Azevedo JS, Correia A, Henriques I. Molecular analysis of the diversity of genus *Psychrobacter* present within a temperate estuary. FEMS Microbiol Ecol. 2013. doi:10.1111/1574-6941.12075.
24. Bakermans C, Ayala-del-Rio HL, Ponder MA, Vishnivetskaya T, Gilichinsky D, Thomashow MF, Tiedje JM. *Psychrobacter cryohalolentis* sp. nov. and *Psychrobacter arcticus* sp. nov., isolated from Siberian permafrost. Int J Syst Evol Microbiol. 2006;56:1285–91.
25. Borsodi AK, Kiss RI, Cech G, Vajna B, Toth EM, Marialigeti K. Diversity and activity of cultivable aerobic planktonic bacteria of a saline lake located in Sovata. Romania Folia Microbiol. 2010;55:461–6.
26. Shivaji S, Reddy GS, Suresh K, Gupta P, Chintalapati S, Schumann P, Stackebrandt E, Matsumoto GI. Psychrobacter vallis sp. nov. and Psychrobacter aquaticus sp. nov., from Antarctica. Int J Syst Evol Microbiol. 2005;55:757–62.
27. Yoon JH, Kang KH, Park YH. *Psychrobacter jeotgali* sp. nov., isolated from jeotgal, a traditional Korean fermented seafood. Int J Syst Evol Microbiol. 2003;53:449–54.
28. Yoon JH, Lee CH, Kang SJ, Oh TK. *Psychrobacter celer* sp. nov., isolated from sea water of the South Sea in Korea. Int J Syst Evol Microbiol. 2005;55:1885–90.
29. Yoon JH, Lee CH, Yeo SH, Oh TK. *Psychrobacter aquimaris* sp. nov. and *Psychrobacter namhaensis* sp. nov., isolated from sea water of the South Sea in Korea. Int J Syst Evol Microbiol. 2005;55:1007–13.
30. Bull AT, Ward AC, Goodfellow M. Search and discovery strategies for biotechnology: the paradigm shift. Microbiol Mol Biol Rev. 2000;64(3):573–606.
31. Mahajan GB, Balachandran L. Sources of antibiotics: hot springs. Biochem Pharmacol. 2016. http://dx.doi.org/10.1016/j.bcp.2016.11.021.
32. Oren A. The family Rhodocyclaceae. In: Rosenberg Eugene, DeLong EF, Lory S, Stackebrandt E, Thompson F, editors. The prokaryotes-Alphaproteobacteria and Betaproteobacteria. Heidelberg, Berlin: Springer-Verlag; 2014.
33. Bell KS, Philp JC, Aw DWJ, Christofi N. The genus *Rhodococcus*. J Appl Microbiol. 1998;85:195–210.
34. Khalilova EA, Nuratinov RA, Kotenko SC, Islammagomedova EA. Hydrocarbon-oxidizing microorganisms of hot springs and their significance in the assessment of the biodiversity of microbial communities. Arid Ecosyst. 2014;4(1):25–30.