Gene Reports 10 (2018) 162–166

Contents lists available at ScienceDirect

Gene Reports

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

# Draft genome sequence of *Geobacillus yumthangensis* AYN2 sp. nov., a denitrifying and sulfur reducing thermophilic bacterium isolated from the hot springs of Sikkim

Ishfaq Nabi Najar, Mingma Thundu Sherpa, Sayak Das, Nagendra Thakur*

*Department of Microbiology, School of Life Sciences, Sikkim University, 6th Mile, Samdur, Tadong, Gangtok 737102, Sikkim, India*

ARTICLE INFO

*Keywords:*
*Geobacillus*, whole genome sequencing
Hot springs, Sikkim
Sulfur reducing bacteria

ABSTRACT

We are reporting the draft genome sequence of the *Geobacillus yumthangensis* AYN2 sp. nov. with the genome size of (~3.4 Mb) corresponding to 3712 predicted genes with (G + C) content of 42.28%. This bacterium possesses motility and chemotaxis genes, sulfur and denitrifying reductase gene clusters. This thermophilic bacterium was isolated from the Yumthang Hot Spring located in the North district of Sikkim, India.

## 1. Introduction

The search for new microflora constituting thermophilic and hyperthermophilic bacteria and archaea from subterranean ecosystems such as hot spring or deep thermal vents have attracted the attention of researchers due the high thermostability and various other contributions of these microflora in industrial and biotechnological sectors (Nazina et al., 2000). The biocatalytic potential of thermophiles and their enzymes such as protease, lipase and polymerase degrading enzymes such as cellulases, chitinases and amylases have been reviewed by various researchers (Sharma et al., 2017). Also at high temperature there is a significant improvement in the solubility of many reaction components. Other factors such as high metabolic rates, physically (i.e., thermally stable) and chemically stable enzymes and cells, facilitated end product recovery and the least risk of contamination (which may cause undesired complications) are very important characteristics of thermophilic bacteria (Lamed and Zeikus, 1980).

Subsequent analysis and research done by several authors, led to the discovery of new genus, phylogenetically distinct, physiologically and morphologically consistent taxon, for which they have submitted the validly-described genus name of *Geobacillus* (Nazina et al., 2001). With high levels of 16S rRNA sequence similarity ranging from 98.5 to 99.2%, the *Geobacillus* species comprise a phenotypically and phylogenetically cogent group of thermophilic bacilli (*Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermoglucosidasius* and *Bacillus thermodenitrificans*).

The genus *Geobacillus* comprises as many as 10 validated species i.e. *G. stearothermophilus*, *G. thermocatenulatus*, *G. thermoleovorans*, *G. kaustophilus*, *G. thermoglucosidasius*, *G. thermo-denitrificans*, *G. subterraneus*, *G. uzenensis*, *G. caldoxylosilyticus* and *G. toebii*. These species are usually Gram-positive, rod shaped thermophilic bacteria (Nazina et al., 2001). *Geobacillus* species are ubiquitous and can be found from hot geothermal locations to cold regions on earth and these bacteria possess a great industrial and biotechnological potential due to their thermostable nature (McMullan et al., 2004). A less genomic study has been done between mesophilic bacillus and thermophilic bacillus related species (Takami et al., 2004).

The study of thermophily in the prokaryotic cells which often survive by modifying their metabolic pathway or other mechanism the thermostability of their proteins mainly rely on the genomic information of the considered genome (Takami et al., 2017). Comparative genomics plays an essential role in mining the contender genes concomitant with thermophily (Tatusov et al., 2001). Comparing genomes based on genomic information between closely related organisms including both mesophiles and thermophiles is also an effective approach for understanding thermoadaptation with respect to evolution. Although the sequences from an appropriate set of organisms such as genus *Geobacillus* are needed, but have not yet been obtained (Takami et al., 2017). Also the study of the genomes of *Geobacillus* species led to the discovery, cloning and exploitation of natural products (Studholme, 2014). Several genomic sequences potentially related to thermostable homologues of useful enzymes have been reported. Also many genome sequences have been used to clone and express the genes of interest and

*Abbreviations:* Mbp, Mega Base Pairs; Bp, base pairs; G., *Geobacillus*; sp., species; ATCC, American Type Culture Collection; MTCC, Microbial Type Culture Collection and Gene Bank; DNA, deoxyribonucleic acid; tRNA, transfer ribonucleic acid; rRNA, ribosomal ribonucleic acid

* Corresponding author.
*E-mail address:* nthakur@cus.ac.in (N. Thakur).

https://doi.org/10.1016/j.genrep.2017.12.007
Received 29 November 2017; Received in revised form 25 December 2017; Accepted 29 December 2017
Available online 30 December 2017

**Table 1**
Genome characteristics and resources of *Geobacillus yumthangenesis* AYN2 nov. sp.

| | |
|---|---|
| NCBI BIOPROJECT ID | PRJNA407404 |
| NCBI BIOSAMPLE ID | SAMN07653191 |
| NCBI GENOME ACCESSION NUMBER | NWUZ00000000 |
| SEQUENCING PLATFORM | ILLUMINA HISEQ 4000 |
| SEQUENCING MODULE | PAIRED END |
| TOTAL NUMBER OF READS | 6,856,386,000 (~6856 Mbp) |
| READ LENGTH | 101 bp |
| AVERAGE COVERAGE | 99.94% |
| ESTIMATED GENOME SIZE | 3,409,966 (~3.4 Mb) |
| GC CONTENT | 42.11% |
| PROTEIN CODING GENES | 3631 |
| tRNA CODING GENES | 71 |
| rRNA Coding genes | 5 |
| PLASMID SEQUENCES | NONE |

to characterize the enzyme for biotechnological potential. For example the genome of *Geobacillus kaustophillus* HTA426 was recently extracted for members of the glycoside hydrolase family 1 (Suzuki and Okazaki, 2013). Also for the first time, the nitrous oxide reductase gene from a Gram-positive and a novel thermophilic long chain alkane monooxygenase from NG80-2 genome sequence was discovered (Feng et al., 2007). Thus complete genomic studies may lead to new insights and will provide much information related to various aspects of a bacterial cell such as difference in metabolism of such bacteria and functionality and thermostability of various proteins and enzymes at molecular level. Here we report the draft genome sequence of a novel *Geobacillus* species - .*Geobacillus yumthangenesis* AYN2.

## 2. Isolation of the bacterium

The *Geobacillus yumthangensis* AYN2 sp. nov. was isolated from the Yumthang Hot spring of North District of Sikkim, India. The bacterium was grown in Thermus medium (ATCC medium: 697 Thermus medium) [Peptone 8 g $L^{-1}$, Yeast Extract 4 g $L^{-1}$and NaCl 2 g $L^{-1}$] at 60 °C for 24 h.

## 3. Genomic DNA isolation, sequencing and data assembly

The genomic DNA of *Geobacillus yumthangensis* AYN2 sp. nov. was extracted from the bacteria, which was grown in Thermus medium at 60 °C using Qiagen QIAamp DNA Mini Kit (50) as per the guide lines of manufacturer. The whole genome sequencing was performed by using Illumina Hiseq 4000 sequencing technology with a paired end sequencing module (Table 1). It produced a total of 68,563,860 bp (68.56 Million) paired end reads with a maximum read length of 101 bp. For high quality data and genome assembly, the data was filtered by employing Next Generation Sequencing Quality Control (NGSQC) Toolkit and SQIT (Patel and Jain, 2012). The total QC passed high quality reads were 52,028,822 (52.03 Million). The overall quality of the data was good with more than 75.88% high quality reads. The primary genome assembly was carried out by Velvet (V 1.2.10) (Zerbino and Birney, 2008). The primary genome assembly statistics revealed the K-mer length of 71 with total number of 454 contigs. The average contig length was 7487.7 bp (~0.07 Mb) with N50 contig size of 24,353 (~0.02 Mb).

Based on paired-end directional information, the genome was assembled into 264 scaffolds, with $N_{50}$ length 27,853 bp (0.03 Mb) and average scaffold length 12,863.51 bp (~0.01 Mb) using SSPACE v3.0 scaffolder (Boetzer et al., 2011). The obtained draft genome was assembled resulting in a total genome size of ~3.3 Mb with (G + C) content of 42.28%. The final genome draft consists of 124 scaffolds with average scaffold length of 27,499.73 bp (~0.02 Mb) and N50 contig size of 2,988,775 bp (~3.0 Mb), constituting 3,409,966 bp (~3.4 Mb) of the genome with (G + C) content of 42.11%.

Bowtie2 (v2.2.2) (Langmead and Salzberg, 2012) was used for de novo genome validation and quality control. By using this tool the reads mapped to the assembly concordantly and discordantly were 19,386,432 and 1,141,476 respectively. The overall alignment was 98.79%. The scaffolds covered were 264 with the average depth (scaffold level in x) of 1494.37 × and average coverage (scaffold level in %) of 99.94%.

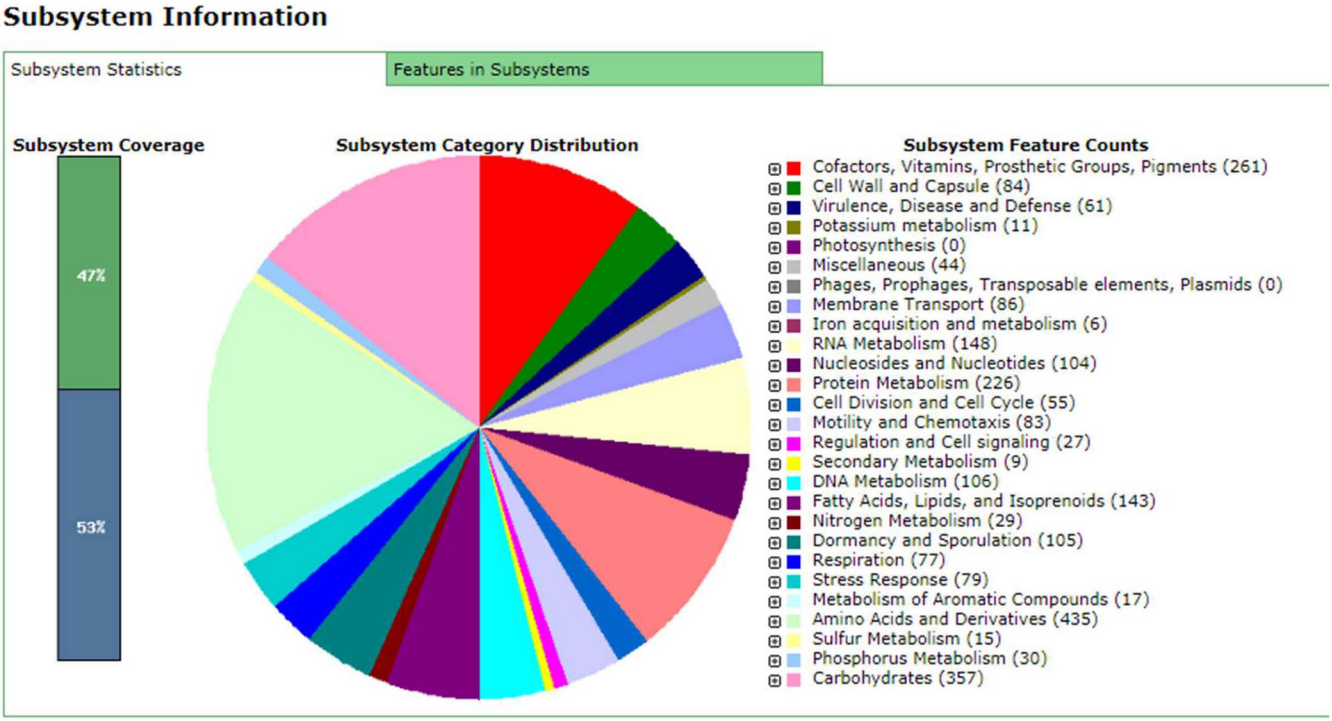


**Fig. 1.** Subsystem category distribution of major protein coding genes of *Geobacillus yumthangenesis* AYN2 nov. sp. as annotated by the RAST annotation server. The bar chart shows the subsystem coverage in percentage (blue bar corresponds to percentage of proteins included). The pie chart shows percentage distribution of the 27 most abundant subsystem categories. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)