

# Genome of a thermophilic bacterium *Geobacillus* sp. TFV3 from Deception Island, Antarctica

Xin Jie CHING<sup>1</sup>, Chui Peng TEOH<sup>1</sup>, Dexter J. H. LEE<sup>1</sup>, Marcelo GONZÁLEZ-ARAVENA<sup>2</sup>, Nazalan NAJIMUDIN<sup>3</sup>, Yoke Kqueen CHEAH<sup>4</sup>, Paris LAVIN<sup>5</sup> & Clemente Michael Vui Ling WONG<sup>1,6\*</sup>

<sup>1</sup> Biotechnology Research Institute, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia;

<sup>2</sup> Instituto Antártico Chileno, Plaza Muñoz Gamero 1055, Punta Arenas, Chile;

<sup>3</sup> School of Biological Science, Universiti Sains Malaysia, Persiaran Bukit Jambul 11900 Bayan Lepas, Penang, Malaysia;

<sup>4</sup> Department of Biomedical Science, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor Darul Ehsan, Malaysia;

<sup>5</sup> Universidad de Antofagasta, Facultad de Ciencias del Mar y Recursos Biológicos, Departamento de Biotecnología; Laboratorio de Complejidad Microbiana y Ecológica Funcional, Instituto Antofagasta, Chile;

<sup>6</sup> National Antarctic Research Centre, University of Malaya, 50603 Kuala Lumpur, Malaysia

Received on 28 September 2019; accepted on 14 January 2020; published online 12 May 2020

**Abstract** Thermophilic microorganisms have always been an important part of the ecosystem, particularly in a hot environment, as they play a key role in nutrient recycling at high temperatures where most microorganisms cannot cope. While most of the thermophiles are archaea, thermophiles can also be found among some species of bacteria. These bacteria are very useful in the fundamental study of heat adaptation, and they are also important as potential sources of thermostable enzymes and metabolites. Recently, we have isolated a Gram-positive thermophilic bacterium, *Geobacillus* sp. TFV3 from a volcanic soil sample from Deception Island, Antarctica. This project was undertaken to analyze the genes of this thermophilic Antarctic bacterium and to determine the presence of thermal-stress adaptation proteins in its genome. The genome of *Geobacillus* sp. TFV3 was first purified, sequenced, assembled, and annotated. The complete genome was found to harbor genes encoding for useful thermal-stress adaptation proteins. The majority of these proteins were categorized under the family of molecular chaperone and heat shock protein. This genomic information could eventually provide insights on how the bacterium adapts itself towards high growth temperatures.

**Keywords** temperature, 16S rDNA, genomes, *Geobacillus*, Deception Island

**Citation:** Ching X J, Teoh C P, Lee D J H, et al. Genome of a thermophilic bacterium *Geobacillus* sp. TFV3 from Deception Island, Antarctica. Adv Polar Sci, 2020, 31(2): \*\*\*-\*\*\*, doi: 10.13679/j.advps.2019.0033

## 1 Introduction

*Geobacillus* is a genus comprising of rod-shaped, Gram-positive thermophilic bacteria. Previously categorized

within the Bacillaceae, these strains were subsequently placed into the new genus *Geobacillus*, as there was a significant difference in their genotype as well as phenotype (Nazina et al., 2001). *Geobacillus* is quite abundant and has been detected in a wide range of habitats, from the sediment in the Mariana Trench, which is 10897 m below sea level (Takami et al., 2004), to the upper troposphere at an altitude

\* Corresponding author, ORCID: 0000-0003-3431-8896, E-mail: michaelw@ums.edu.my

above 10000 m (DeLeon-Rodriguez et al., 2013). Most of the members can grow at a temperature ranging from 45 °C to 75 °C, whereas some of them can extend this to as low as 35 °C or as high as 80 °C (Zeigler, 2014). Despite having high growth temperature, these thermophiles can be found not only in hot environments, but they can also be easily isolated from cool soils and cold ocean sediments (Zeigler, 2014; Studholme, 2015). Currently, there are around 14 known species in this genus as depicted in National Center for Biotechnology Information (NCBI, part of the United States National Library of Medicine (NLM)) database, the most extensively studied ones being *G. stearothermophilus* and *G. kaustophilus*. Their property of growing at high temperature allows them to be one of the great interests in the field of extremophile researches, as they can produce thermostable hydrolytic extracellular enzymes that can function optimally at a temperature higher than the majority of organisms can withstand (Weigand et al., 2013). Usually, bacteria would adopt more than one strategy to make up for their thermal adaptability. For instance, it has been found that the high thermal melting temperature of dihydrofolate reductase from *Geobacillus stearothermophilus* was attributed to several aspects, which includes additional proline residues, removal of water-accessible thermolabile residues, ligand binding, and also the presence of salt or cosolvents (Guo et al., 2014).

Even though more and more *Geobacilli* have been discovered recently, not much is known about the thermal adaptation strategies adopted by the bacteria at elevated temperature. Genome analysis is an important step in providing information on how the bacteria respond to high growth temperatures since everything gets down to the molecular level nowadays. Therefore, the objective of the project was to analyze the genome of a thermophilic Antarctic bacterium *Geobacillus* sp. TFV3 isolated from Deception Island, Antarctica, and to determine the presence of thermal-stress adaptation proteins in its genome.

## 2 Material and methods

Strain TFV3 was isolated from the volcanic soil of Deception Island (62°58'57.6"S, 60°39'52.3"W). It was grown and routinely sub-cultured in Trypticase Soy Medium at 65 °C. High-quality genomic DNA was extracted using an optimized protocol for QIAGEN DNeasy Blood and Tissue kit. PCR was carried out using primer BSF-8/20 (5'-AGAGTTTGATCCTGGCTCAG-3') and BSR-1541/20 (5'-AAGGAGGTGATCCAGCCGCA-3') to amplify the 16S rDNA. The amplicon was sequenced and aligned against sequences in NCBI GenBank using Basic Local Alignment Search Tool (BLAST) to identify the strain.

The extracted genomic DNA was then sequenced using Pacific Biosciences Sequencer RSII. The polymerase read

statistics and sub-read statistics are shown in Tables 1 and 2 respectively. The raw sequencing data were then corrected, trimmed, and assembled using Canu (Koren et al., 2017), followed by circularizing the genome assembly using Circlator (Hunt et al., 2015). The circular genome was then annotated using RAST (Rapid Annotation using Subsystem Technology) (Aziz et al., 2008).

**Table 1** Polymerase read statistics

Polymerase read statistics	Data
Polymerase read bases	1511853734
Polymerase reads	106561
Polymerase read N50	18941
Polymerase read length	14187
Polymerase read quality	0.843

**Table 2** Subread statistics

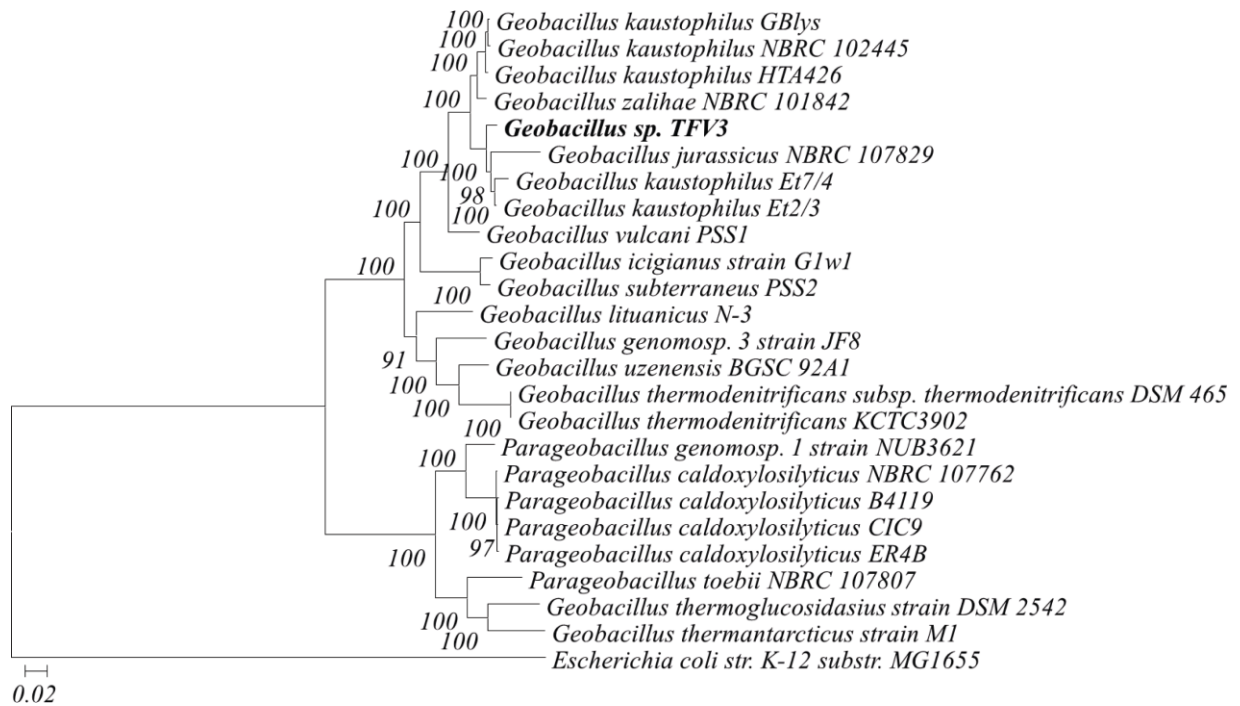
Subread statistics	Data
Subread bases	1510230322
Subreads	141606
Subread N50	15173
Average subread length	10665

A phylogenetic tree was constructed using the multiple reference genes present in the whole genome sequence of strain TFV3 instead of just using one highly conservative gene like 16S rDNA for a more reliable phylogenetic analysis. To achieve this, software PhyloSift (Darling et al., 2014) was used to compare the genome of strain TFV3 against the genome of 17 other *Geobacilli* and seven *Parageobacilli* obtained from the NCBI database. From the phylogenetic tree, a few closest related species were selected, and the thermal-stress related genes in their genomes were compared against those of strain TFV3.

The complete genome sequence of *Geobacillus* sp. TFV3 has been deposited to NCBI database, under the accession number MPSP01000000.

## 3 Results and discussion

The DNA alignment result of the 16S rDNA sequence showed that strain TFV3 has the highest similarity to the partial sequence 16S ribosomal RNA gene of *Geobacillus* sp. strain Et7/4 (99.42%), followed by *Geobacillus* sp. strain Et2/3 (99.41%), both with 99% query coverage and 0 E-value. The identification was further supported by the phylogenetic analysis using PhyloSift. As shown in Figure 1, strain TFV3 formed a distinctive cluster with *G. jurassicus* NBRC 107829, *G. kaustophilus* Et7/4, and *G. kaustophilus* Et2/3. This suggested that these 3 *Geobacilli* have the closest phylogenetic relationship with strain TFV3.



**Figure 1** Whole genome phylogenetic tree constructed by PhyloSift, using the Maximum Likelihood method based on a Generalised Time-Reversible (GTR) model. The tree shows the close relationship between TFV3 related species, while *E. coli* K-12 substr. MG1655 is included to serve as outlier.

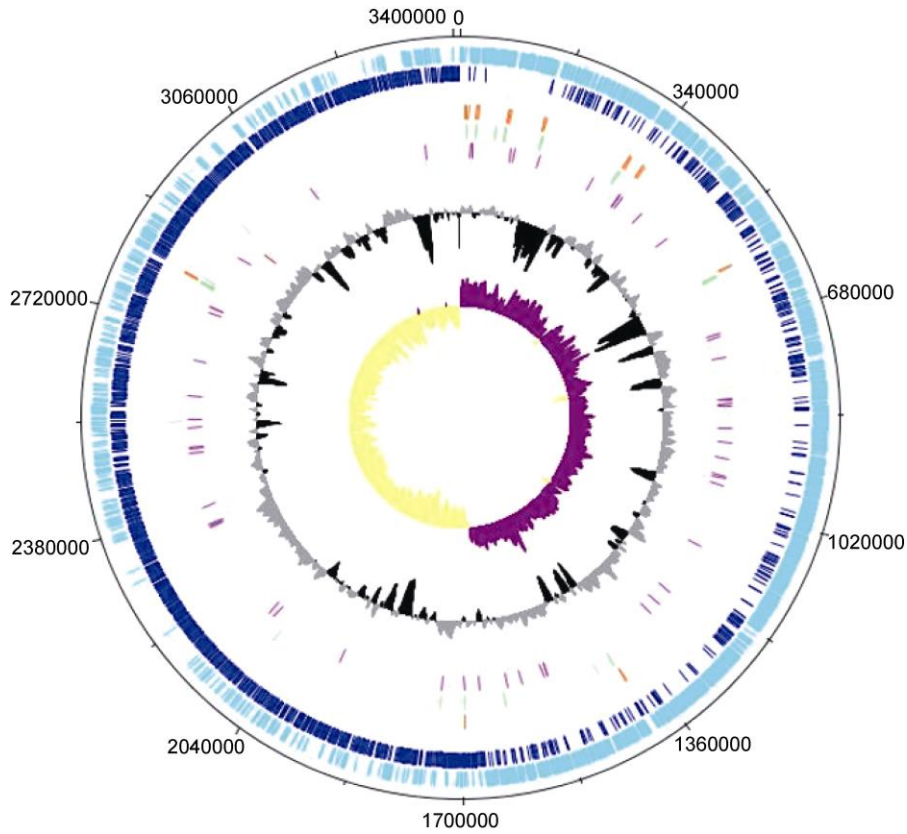
The genome properties of strain TFV3 after assembly and annotation are listed in Table 3. The genome of strain TFV3 was assembled into a single contig, suggesting that there is no plasmid in this bacterium. The genome was 3408572 bp in size, and with a G + C content of 51.9%. This falls into the 51% to 53% G + C content range of most of the *Geobacilli*. The genome has 3903 genes, including 3787 putative protein-coding sequences (CDS), 27 rRNA, 88 tRNA, and 1 tmRNA. A genome map was constructed using DNAPlotter (Carver et al., 2009) for better visualization of the position of each element in the genome (Figure 2). All the annotated genes were classified into orthologous groups based on their function using eggNOG (Jensen et al., 2008). Figure 3 shows that most of the genes fall into group S, suggesting that all the hypothetical genes and some of the genes encoding for proteins with unknown function occupied 32.11% of the total predicted genes in the genome. Aside from the hypothetical genes, the group that has the second most genes is group L (10.22%), which is responsible for DNA replication, recombination, and repair. This makes sense as strain TFV3 lived in a rather extreme environment, it is important that they protect and constantly evolve their DNA so that they adapt to the harsh living conditions. An even distribution was observed between most of the categories related to cellular metabolisms, energy production, transport, and signal transduction (~3%–6%), except for nucleotide transport and metabolism (2.31%). These genes are equally important in maintaining the routine cellular process and vitality of strain TFV3. Most of the genes encoding for thermal-stress proteins were

categorized under group O (post-translational modification, protein turnover, and chaperones) and group K (transcriptional).

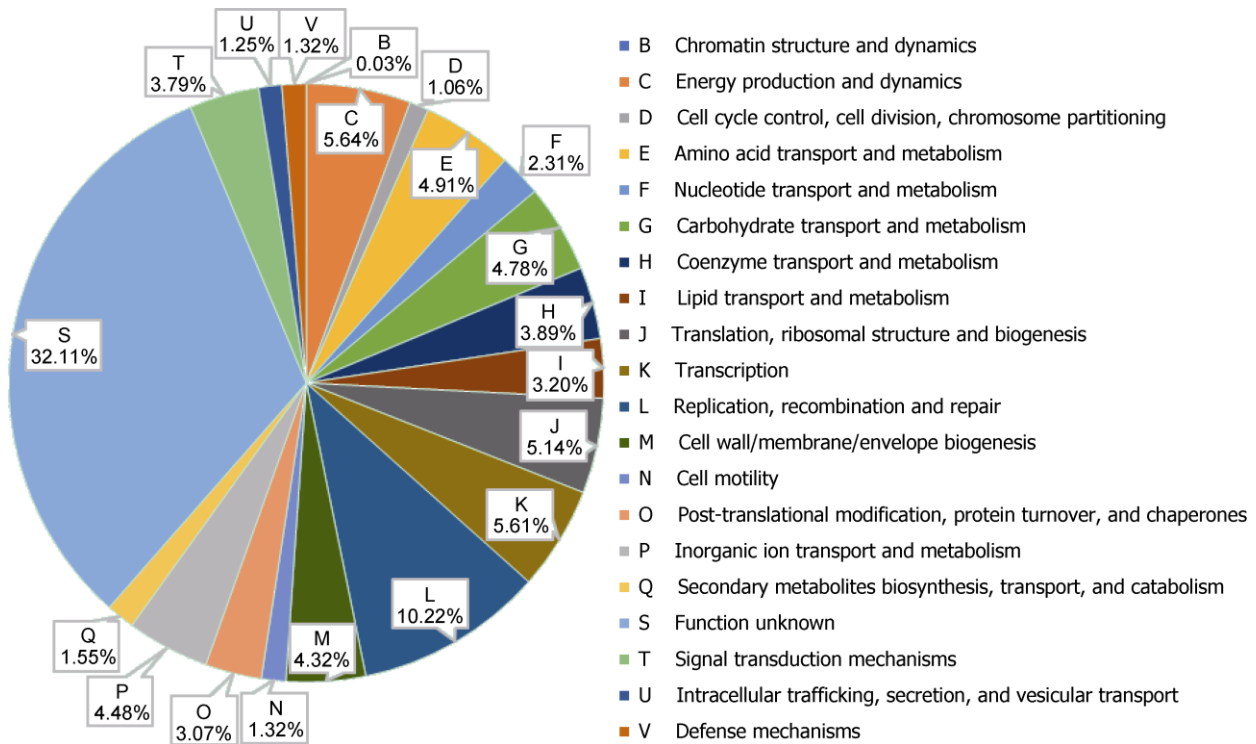
**Table 3** Genome properties of *Geobacillus* sp. TFV3

Genome properties	Data
Contigs no.	1
Genome size/bp	3408572
GC content/%	51.9
Total number of genes	3903
CDS	3787
Genes with predicted function	2621
Hypothetical genes	1166
rRNA	27
tRNA	88
tmRNA	1

Several common thermal-stress related genes were found in the genome of strain TFV3, including genes encoding for cold-shock protein (CSP), heat-shock protein (HSP), and molecular chaperone (Table 4). These genes are very important as they have a major contribution to maintaining cell viability in response to temperature fluctuations. Cold-related protein such as cold-shock protein CspD is thought to facilitate the initiation of translation as they serve as RNA chaperones to prevent the formation of secondary structure in mRNA when the temperature drops (Phadtare, 2004). Meanwhile, the other



**Figure 2** Genome map of *Geobacillus* sp. TFV3 as viewed DNAPlotter. The outermost track represents forward CDS, followed by Reverse CDS, rRNA, tRNA, miscRNA, tmRNA, GC plot, and GC skew



**Figure 3** Functional distribution of genes within *Geobacillus* sp. TFV3 genome classified by clusters of orthologous groups (COG) using eggNOG.

**Table 4** Number of gene copies for thermal stress related proteins found in each bacterial genome. (1) *Geobacillus* sp. TFV3, (2) *G. jurassicus* NBRC 107829, (3) *G. kaustophilus* Et2/3, (4) *G. kaustophilus* Et7/4, and (5) *Parageobacillus caldoxylosilyticus* ER4B as an outlier for comparison

Thermal stress related proteins	1	2	3	4	5
Antifreeze protein	0	0	0	0	0
Cold shock protein <i>CspC</i>	0	0	0	0	2
Cold shock protein <i>CspD</i>	1	1	1	1	1
Low temperature requirement B protein	0	0	0	0	1
Low temperature requirement C protein	1	1	1	1	1
Chaperone protein <i>DnaJ</i>	1	1	1	1	1
Chaperone protein <i>DnaK</i>	1	1	1	1	1
Chaperone protein <i>ClpB</i>	0	0	1	0	0
Heat shock protein 33	1	1	1	1	1
Heat shock protein 60 co-chaperone <i>GroES</i>	1	1	1	1	1
Heat shock protein 60 family chaperone <i>GroEL</i>	1	1	1	1	1
Heat shock protein class I	1	1	1	1	1
Heat shock protein <i>GrpE</i>	1	1	1	1	1
Heat-inducible transcription repressor <i>HrcA</i>	1	1	1	1	1
PBS lyase heat-like repeat domain protein	1	1	1	1	1
Ribosome-associated heat shock protein	1	1	1	1	1
RNA polymerase heat shock sigma factor SigI	1	0	1	1	1
Small heat shock protein	1	1	1	2	1

heat-related proteins such as HSPs and chaperones are useful in the event of temperature rise, as they are able to prevent protein aggregation (Whitley et al., 1999), repair damaged proteins (De Maio, 1999), and promote degradation of damaged proteins (Raboy et al., 1991).

Some thermal-stress related genes found in strain TFV3 were also compared to the three closest species and a *Parageobacillus* species (Table 4). It was found that the presence of thermal-stress related genes in the genomes of all five bacterial strains are highly similar, with *P. caldoxylosilyticus* ER4B exhibiting a relatively different pattern in terms of gene presence. Other than antifreeze-related proteins, which are absent in all the strains, cold-shock protein, *CspC* and low-temperature requirement B protein which were only present in the genome of strain ER4B, was absent in all four *Geobacilli*. As well, chaperone protein *ClpB* was also absent in strain TFV3.

The presence of proteins such as chaperone protein *DnaJ*, chaperone protein *DnaK*, heat-shock protein *GrpE*, heat-shock protein 60 family chaperone *GroEL*, and heat-shock protein 60 co-chaperone *GroES* suggested that these proteins are crucial for the *Geobacilli* to be able to survive at high temperature. These molecular chaperones are primarily responsible for maintaining the integrity of cellular proteins and preventing cellular aggregation in response to environmental stress (Ranford et al., 2000; Harrison, 2003). They do not work independently but work in conjunction with one another. For instance, *GroEL*

requires lid-like *GroES* to function properly. Binding of *GroES* to the apical domain binding site of *GroEL* would induce several changes which would lead to easier protein folding (Samali et al., 1999; Lee et al., 2002; Horwich et al., 2007). The same applies to *DnaJ* and *GrpE* which not only help to stimulate the ATPase activity of *DnaK* (Cyr et al., 1994; Hennessy et al., 2005), but also form a ternary complex with *DnaK* and polypeptide chains to prevent the aggregation of the polypeptide chains (Han and Christen, 2004). Although these heat shock proteins exist in almost every bacterium, the thermos-stability of the proteins is different in each bacterium and this influences the ability of the bacteria to grow at elevated temperature (Endo et al., 2006; Kumwenda et al., 2013).

The genome information of strain TFV3 provides insights into the genetic makeup of a thermophilic bacterium that lives in a hot environment. It harbors many genes encoding thermal-stress related proteins that might be significant for the bacterium in response to the high yet ever-changing temperature around the volcanic region in Antarctica. The genome data is important for the subsequent analysis to determine the key genes required for this bacterium to grow and survive under hot conditions. For instance, by performing a systematic gene knock-out or gene silencing study, functions of a gene related to thermal adaption can be determined. The complete genome data is also useful in comparative genomics among thermophilic bacteria as it could provide gene functional annotation by

comparing homologous thermal-stress related proteins from different bacteria. In addition, the genome of strain TFV3 also contains genes encoding for enzymes such as lipases, proteases, amylases, glycosyltransferases, RNA polymerases, DNA polymerases, and ligases with industrial applications which are most likely to be thermostable. Lastly, with the genome data, this bacterium can be manipulated to be a host to express genes at high temperatures. In conclusion, the genome of *Geobacillus* sp. TFV3 data is not only an important source of information for fundamental study on thermal-adaptation of this bacterium and potentially as a host for gene expression study but also serves as a source of thermostable enzymes.

**Acknowledgments** The funding support from the Ministry of Science, Technology, and Innovation (MOSTI), Malaysia, under the Antarctica Flagship Programme (Sub-Project 1: Grant no. FP1213E036) is gratefully acknowledged. We appreciate very much two anonymous reviewers, and guest editor for their helpful and constructive comments on the manuscript of this paper.

## References

- Aziz R K, Bartels D, Best A A, et al. 2008. The RAST server: rapid annotations using subsystems technology. *BMC Genom*, 9(1): 75, doi: 10.1186/1471-2164-9-75.
- Carver T, Thomson N, Bleasby A, et al. 2009. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics*, 25(1): 119-120, doi: 10.1093/bioinformatics/btn578.
- Han W J, Christen P. 2004. *Cis*-effect of DnaJ on DnaK in ternary complexes with chimeric DnaK/DnaJ-binding peptides. *FEBS Lett*, 563(1-3): 146-150, doi: 10.1016/S0014-5793(04)00290-X.
- Darling A E, Jospin G, Lowe E, et al. 2014. PhyloSift: phylogenetic analysis of genomes and metagenomes. *Peer J*, 2: e243, doi: 10.7717/peerj.243.
- De Maio A. 1999. Heat shock proteins: facts, thoughts, and dreams. *Shock*, 11(1): 1-12, doi: 10.1097/00024382-199901000-00001.
- Deleon-Rodriguez N, Latham T L, Rodriguez-R L M, et al. 2013. Microbiome of the upper troposphere species composition and prevalence, effects of tropical storms, and atmospheric implications. *Proc Natl Acad Sci USA*, 110(7): 2575-2580, doi: 10.1073/pnas.1212089110.
- Cyr D M, Langer T, Douglas M G. 1994. DnaJ-like proteins: molecular chaperones and specific regulators of Hsp70. *Trends Biochem Sci*, 19(4): 176-181, doi:10.1016/0968-0004(94)90281-x.
- Endo A, Sasaki M, Maruyama A, et al. 2006. Temperature adaptation of *Bacillus subtilis* by chromosomal groEL replacement. *Biosci Biotechnol Biochem*, 70(10): 2357-2362, doi: 10.1271/bbb.50689.
- Guo J N, Luk L Y P, Loveridge E J, et al. 2014. Thermal adaptation of dihydrofolate reductase from the moderate thermophile *Geobacillus stearothermophilus*. *Biochemistry*, 53(17): 2855-2863, doi: 10.1021/bi500238q.
- Harrison C. 2003. GrpE, a nucleotide exchange factor for DnaK. *Cell Stress Chaper*, 8(3): 218-224, doi: 10.1379/1466-1268(2003)008<0218:ganeff>2.0.co;2.
- Hennessy F, Nicoll W S, Zimmermann R, et al. 2005. Not all J domains are created equal: implications for the specificity of Hsp40-Hsp70 interactions. *Protein Sci*, 14(7): 1697-1709, doi: 10.1110/ps.051406805.
- Horwich A L, Fenton W A, Chapman E, et al. 2007. Two families of chaperonin: physiology and mechanism. *Annu Rev Cell Dev Biol*, 23(1): 115-145, doi: 10.1146/annurev.cellbio.23.090506.123555.
- Hunt M, Silva N D, Otto T D, et al. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol*, 16: 294, doi: 10.1186/s13059-015-0849-0.
- Jensen L J, Julien P, Kuhn M, et al. 2008. eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic Acids Res*, 36 (Database issue): D250-D254, doi: 10.1093/nar/gkm796.
- Koren S, Walenz B P, Berlin K, et al. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res*, 27(5): 722-736, doi: 10.1101/gr.215087.116.
- Kumwenda B, Litthauer D, Bishop Ö T, et al. 2013. Analysis of protein thermostability enhancing factors in industrially important *Thermus* bacteria species. *Evol Bioinform Online*, 9: 327-342, doi: 10.4137/EBO.S12539.
- Lee K H, Kim H S, Jeong H S, et al. 2002. Chaperonin GroESL mediates the protein folding of human liver mitochondrial aldehyde dehydrogenase in *Escherichia coli*. *Biochem. Biophys Res Commun*, 298(2): 216-224, doi: 10.1016/s0006-291x(02)02423-3.
- Nazina T N, Tourova T P, Poltarau A B, et al. 2001. Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, *G. th.* *Int J Syst Evol Microbiol*, 51: 433-446, doi: 10.1099/00207713-51-2-433.
- Phadtare S. 2004. Recent developments in bacterial cold-shock response. *Curr Issues Mol Biol*, 6(2): 125-136.
- Raboy B, Sharon G, Parag H A, et al. 1991. Effect of stress on protein degradation: role of the ubiquitin system. *Acta Biologica Hungarica*, 42(1-3): 3-20.
- Ranford J C, Coates A R M, Henderson B. 2000. Chaperonins are cell-signalling proteins: the unfolding biology of molecular chaperones. *Expert Rev Mol Med*, 2(8): 1-17, doi: 10.1017/S1462399400002015.
- Samali A, Cai J, Zhivotovsky B, et al. 1999. Presence of a pre-apoptotic complex of pro-caspase-3, Hsp60 and Hsp10 in the mitochondrial fraction of jurkat cells. *EMBO J*, 18(8): 2040-2048, doi: 10.1093/emboj/18.8.2040.
- Studholme D J. 2015. Some (bacilli) like it hot: genomics of *Geobacillus* species. *Mircob Biotechnol*, 8(1): 40-48, doi: 10.1111/1751-7915.12161.
- Takami H, Nishi S, Lu J, et al. 2004. Genomic characterization of thermophilic *Geobacillus* species isolated from the deepest sea mud of the Mariana Trench. *Extremophiles*, 8(5): 351-356, doi: 10.1007/s00792-004-0394-3.

- Weigand S, Rabausch U, Chow J, et al. 2013. Complete genome sequence of *Geobacillus* sp. strain GHH01, a thermophilic lipase-secreting bacterium. *Genome Announc*, 1(2): e00092-13, doi: 10.1128/2FgenomeA.00092-13.
- Whitley D, Goldberg S P, Jordan W D. 1999. Heat shock proteins: a review of the molecular chaperones. *J Vasc Surg*, 29(4): 748-751, doi: 10.1016/s0741-5214(99)70329-0.
- Zeigler D R. 2014. The *Geobacillus* paradox: why is a thermophilic bacterial genus so prevalent on a mesophilic planet? *Microbiology*, 160(1): 1-11, doi: 10.1099/mic.0.071696-0.