

ANNOTATION

of the dissertation work for a degree of Doctor of Philosophy (Ph.D)
in specialty “6D010700 – Biotechnology”

Mashzhan Akzhigit Sembayuly

Metagenomic analysis of Kazakhstan thermophilic bacteria for obtaining promising hydrolytic enzymes

General description of the research. The dissertation is dedicated to investigating the taxonomic diversity and biotechnological potential of the hydrolytic microorganism community in geothermal springs situated on the territory of Kazakhstan. This research employs a combination of high-throughput metagenomic sequencing methods and traditional microbiological techniques.

Significance of the research. The current focus on studying the genetic diversity of microbial communities in geothermal habitats, alongside the investigation of thermophilic bacteria residing in these environments and their amino acids, reflects a dual interest from both ecological and biotechnological perspectives. The enzymes and thermozymes produced by thermophilic and hyperthermophilic microorganisms dwelling in such extreme conditions present several biotechnological benefits compared to their mesophilic counterparts. Notably, one of these advantages is their high thermal stability. This thermal stability is often complemented by resistance to various other harsh conditions, making these biomolecules particularly valuable for biotechnological applications.

The territory of Kazakhstan is abundant in diverse sources of geothermal springs. However, the microbial communities inhabiting these waters have not been comprehensively studied. This highlights the potential for discovering new extremophilic microorganisms in these geothermal waters, thereby opening avenues for biotechnological applications.

Research on the phylogenetic diversity of microorganisms in extreme habitats reveals that a significant proportion of thermophilic microorganisms have not been successfully cultivated in laboratory conditions. A primary obstacle to this lies in the fact that nearly 99% of microorganisms on Earth cannot thrive under standard laboratory conditions. This limitation makes it challenging to identify and characterize them using traditional methods of microbiology and biochemistry.

In recent years, modern molecular biology methods have provided new avenues for addressing this challenge, and one such method is metagenomics.

In recent years, advancements in molecular biology have introduced innovative solutions to address this challenge, with metagenomics standing out as one such method. Metagenomic analysis offers the opportunity to identify microorganisms and potentially discover novel beneficial enzymes from thermal water sources through culture-independent methods. Presently, high-throughput metagenome sequencing methods can complement traditional culture-based approaches, providing a comprehensive understanding of complex microbial communities.

However, the isolation of pure cultures remains essential for comprehending the physiology of microorganisms and assessing their potential for specific

environmental applications. The integration of both traditional culture-dependent and culture-independent molecular techniques has significantly broadened our knowledge of the functional and structural diversity within microbial communities. This combined approach enables a more holistic exploration of microbial ecosystems, allowing for a deeper understanding of their potential biotechnological applications.

Hence, the identification of promising hydrolytic microorganisms from geothermal sources in Kazakhstan, which have not yet been explored using modern methods, stands as an urgent task from both scientific and applied perspectives. This investigation holds the potential to unveil new species of microorganisms with distinctive hydrolytic properties, bearing significant implications for biotechnological advancements, particularly in the production of enzymes exhibiting heightened activity at high temperatures.

Moreover, such research plays a crucial role in broadening our comprehension of the biodiversity and ecology of microorganisms in extreme environments. This expanded understanding is essential for overarching scientific progress and holds promise for applications in various industries and fields.

The purpose of the research: Assessing the taxonomic diversity of thermophilic hydrolytic microorganisms within the geothermal springs of Kazakhstan.

To accomplish the set **goal**, the following tasks have been outlined:

1. To study the taxonomic diversity of microbial communities in Zharkent geothermal springs using metagenomic methods;
2. Isolate and identify new strains/species of thermophilic bacteria from the studied microbial communities;
3. Explore the hydrolytic activity of isolated strains of thermophilic bacteria;
4. Optimize the nutrient medium for the cultivation of thermophilic bacteria;
5. Conduct full genome sequencing and annotate the genomes of promising strains of thermophilic bacteria.

The research objects. The microbial communities and isolates from samples collected from Zharkent geothermal springs, situated in the Almaty region, were studied. The type strain *Polycladomyces subterraneus* KSR 13^T (collection number BCC 50740) from the International Collection of Microorganisms of Thailand was used.

Subject of study. Identification of microorganisms from geothermal springs.

Research methods. In this study, a combination of metagenomic, genomic, microbiological, and biochemical methods was employed. Next-Generation Sequencing (NGS) was carried out on Illumina platforms, including HiSeq 4000. The raw NGS data underwent bioinformatic processing, de novo assembly, and custom annotation using CLC Genomic Workbench v. 20.0.01, KBase, and RAST v. 1073. Additionally, complete genome assembly from the metagenome was performed using MaxBin2 v. 2.2.4, CONCOCT v. 1.1, and MetaBAT2 v. 1.7.

The scientific novelty of the research. The results of metagenomic sequencing reveal a bacterial community comprising three major types:

Firmicutes, *Deinococcus - Thermus*, and *Actinobacteria*, with a notable dominance of bacteria from the *Firmicutes* and *Deinococcus - Thermus* phyla.

From the metagenomes, 11 Metagenome Assembled Genomes (MAGs) were extracted, representing 10 distinct species across the following genera: *Pseudomonas*, *Fervidobacterium*, *Caloramator*, *Calditerricola*, *Sutcliffeiella*, *Geobacillus*, *Cerasibacillus*, *Paenibacillus*, *Parageobacillus*, and *Thermus*.

For the first time, thermophilic bacteria from seven genera (*Geobacillus*, *Anoxybacillus*, *Polycladomyces*, *Thermus*, *Caldicellulosiruptor*, *Caldanaerobacter*, and *Thermoanaerobacter*) were isolated and sequenced in pure culture from samples of the microbial community obtained during an expedition on the territory of Zharkent hydrothermal springs. The sequencing was performed on 16S rRNA genes. Additionally, the genomes of three thermophilic bacterial strains, belonging to the genera *Polycladomyces* spp, and *Caldanaerobacter* sp, have been sequenced.

A novel cellulolytic thermophilic bacterium, designated as *Polycladomyces zharkentensis* ZKZ2^T sp., was isolated for the first time from the geothermal springs of Zharkent. The strain is deposited as nov (=KCTC 43421, =CECT 30708). Additionally, the first subspecies of the keratolytic thermophilic bacterium *Caldanaerobacter subterraneus* subsp. *keratonaliticus* KAk was isolated and characterized. This keratolytic bacterium represents the initial identified species of the genus *Caldanaerobacter* with the ability to grow on untreated keratin, specifically chicken feathers.

New strains of aerobic bacteria with an optimal growth temperature of 75°C have been isolated and characterized, identified as *Geobacillus* 3Wak3 and *Geobacillus* 3SAk4. These strains exhibit the highest temperature resistance among currently known species of the genus *Geobacillus*.

The theoretical and practical significance of the research. The analysis of the phylogenetic diversity of microbial communities in the Zharkent hydrothermal spring (Almaty region) enhances our understanding of interspecific interactions within extremophilic microbial communities. The findings from metagenomic and microbiological studies contribute to the expansion and clarification of our knowledge regarding the microbiology of extreme environments. Within the scope of this study, a novel thermophilic bacterium, *Polycladomyces zharkentensis* ZKZ2^T, was isolated, thoroughly characterized, and deposited in the international culture collections KSTS (Korean Collection for Type Cultures) and CECT (Spanish Type Culture Collection) (=KCTC 43421, =CECT 30708).

For the first time, 54 isolates of thermophilic and hyperthermophilic bacteria exhibiting protease, cellulase, amylase, lipase, and keratinase activity were successfully isolated. These strains are valuable for biotechnology and can be used as a source of enzymes in various industries such as detergent production, chemical industry, keratin and cellulose waste processing, and other industries.

The keratinases produced by the anaerobic strain *Caldanaerobacter subterraneus* subsp. *keratonalyticus* KAk, isolated from the Zharkent geothermal spring, have been demonstrated to effectively degrade resilient β -keratin protein,

particularly chicken feathers. Additionally, the culture medium for this bacterium was optimized to enhance its growth and keratinase production.

16S rRNA sequences for 16 bacterial strains with high hydrolytic activity, as well as four metagenomes and three complete genomes were deposited in the global GenBank database.

Patent of the Republic of Kazakhstan No. 5803 dated January 29, 2021 was received for the utility model “Culture medium for the cultivation of anaerobic thermophilic microorganisms with keratinolytic activity”.

The main provisions for the defence are as follows:

1. The prokaryotic microbiota of hot geothermal springs exhibited dominance by bacterial phylotypes such as *Firmicutes*, *Actinobacteria*, and *Deinococcus - Thermus*. Furthermore, the similarity (<97%) observed between these phylotypes and culturable bacteria from the GenBank database suggests the presence of novel and unique species colonizing the studied geothermal springs.

2. *Paenibacillus* sp., *Thermus* sp., and *Geobacillus* sp. exhibited high adaptability in polysaccharide decomposition (in addition to high temperature).

3. 16 promising strains were isolated, belonging to the genera *Anoxybacillus*, *Geobacillus*, *Caldanaerobacter*, *Polycladomyces*, and *Calideclulosiraptor*. These thermophilic microorganisms in pure cultures can form the foundation for establishing a valuable culture collection.

4. The isolated thermophilic strains exhibit activity as producers of thermostable hydrolases, including cellulase, amylase, protease, keratinase, and lipase.

5. An optimized nutrient medium for cultivating the new keratinase producer, *Caldanaerobacter subterraneus* subsp. *keratonalyticus* Kak, has been developed. This medium promotes efficient biomass accumulation and enhances keratinolytic activity.

6. The strain *Polycladomyces* ZKZ2 (type Bacillota), isolated from samples of Zharkent geothermal springs, has been identified as a new species and named *Polycladomyces zharkentensis* ZKZ2^T sp.nov (=KSTS 43421, =CECT 30708).

7. A novel subspecies of the hyperthermophilic bacterium *Caldanaerobacter subterraneus* has been identified, designated as subspecies *keratonalyticus* KAk. This subspecies represents the first member of the genus *Caldanaerobacter* known for its keratinolytic activity.

Personal contribution of the author. The author personally wrote the dissertation and manuscript, developed an experimental plan, conducted the research, and also personally participated in bioinformation processing and analysis of the results obtained.

Relationship of the research with the scientific project. The dissertation work was supported by the project CPEA-LT-2017/10061 within the framework of the Eurasian Program of the Norwegian Agency for International Cooperation and Quality Improvement in Higher Education (Diku) (2017 - 2022). Additionally, funding was provided within the framework of the project AP14871683 “Biotechnology for the processing of keratin by-products using immobilized thermophilic bacteria” (2022 - 2024).

Research approbation. The results of the dissertation were presented and discussed at the following international scientific conferences:

- “International Research and Practive Conference” January 30 - February 7, 2019, Sheffield, England;;
- 23rd international scientific conference “Biology - the science of the 21st century”, April 15-19, 2019, Pushchino, Russia;
- “Current issues of organic chemistry and biotechnology” Khalykaralyk gylymi conferences, November 18-21, 2020, Ekaterinburg, Russia;
- International scientific conference of students and young scientists “Farabi Alemi”, April 6-9, 2020, Almaty, Kazakhstan;
- International scientific conference “Aspects and innovations of environmental biotechnology and bioenergy”, February 12-13, 2021, Almaty, Kazakhstan;
- “Biology and Biotechnology of Microorganisms International Conference” September 16-17, 2021, Tashkent, Uzbekistan;

Publications. The main results of the research on the dissertation topic have been published in 11 scientific papers, including 3 articles in domestic periodicals recommended by the Committee for Quality Assurance in the Field of Science and Higher Education of the MSHE RK, 2 article in scientific journals included in the Web of Science and Scopus databases, and 1 chapter in a collective monograph in Springer. Furthermore, a patent for a utility model in the Republic of Kazakhstan has been obtained.

Dissertation structure. The dissertation is comprised of 143 pages and encompasses normative references, definitions, notations, and abbreviations, introduction, literature review, materials and research methods, research results and their discussion, conclusion, and bibliography containing 222 titles. The document features 48 figures, 29 tables, and 5 appendices.