BMC Genomic Data

DATA NOTE Open Access

Complete genome sequence of thermophilic PHA-producing *Hydrogenophilus thermoluteolus* strain KS0102 isolated from hot spring sediments in Thailand

Pattarawan Ruangsuj[1†], Wariya Yamprayoonswat[1†], Manassanan Phatcharaharikarn[1], Thunwarat Songngamsuk[1], Parweenuch Santaweesuk[1], Marut Tangwattanachuleeporn[2], Kanokporn Srisucharitpanit[2] and Montri Yasawong[1,3*]

Abstract

Objectives Thermophilic bacteria are valuable biotechnological resources. A thermophilic bacterium isolated from the Thai hot spring sediments was subjected to complete genome sequencing. Genome data will facilitate bacterial identification and comparative genomics within the *Hydrogenophilus* genus and provide insights into polyhydroxybutyrate biosynthesis pathways for industrial applications.

Data description The *H. thermoluteolus* KS0102 genome comprises a single circular chromosome containing 2,136,028 bp and 61.70% GC. The genome was sequenced using Illumina and Oxford Nanopore technologies and assembled using Unicycler. It contained 1,848 CDS, 9 rRNAs, 49 tRNAs, and 1 CRISPR region. This dataset enables the investigation of polyhydroxybutyrate biosynthesis pathways and comparative genomic studies of thermophilic adaptation mechanisms in *Hydrogenophilus* species.

Keywords Biodegradable polymers, Hot spring sediments, Hybrid assembly, *phaC1*, *phaC2*, PHB synthase, Polyhydroxybutyrate

Objective

Thermophilic bacteria offer significant biotechnological advantages through heat tolerance, including reduced contamination risks, lower cooling costs, and enhanced process efficiency [1]. Polyhydroxybutyrate (PHB) represents a promising biodegradable alternative to petroleum-based plastics, addressing concerns related to plastic pollution [2]. Thermophilic PHB production provides natural sterile conditions, simplified product recovery, and reduced energy requirements [3].

Hot springs host various thermophilic bacteria with distinctive metabolic functions [4]. The thermophilic bacterium *Schlegelella thermodepolymerans* has demonstrated exceptional PHA production capabilities,

†Pattarawan Ruangsuj and Wariya Yamprayoonswat are contributed equally to this work.

*Correspondence:
Montri Yasawong
montri@cgi.ac.th
[1]Programme on Environmental Toxicology, Chulabhorn Graduate Institute, Bangkok 10210, Thailand
[2]Faculty of Allied Health Sciences, Burapha University, 169 Long- Hard Bangsean Road, Saensuk Sub-district, Mueang Chon Buri District, Chon Buri Province 20130, Thailand
[3]Center of Excellence on Environmental Health and Toxicology (EHT), OPS, MHESI, Bangkok 10400, Thailand

Table 1 Overview of data files and datasets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	Paired-end Illumina reads	Sequence file (.fastq)	Sequence Read Archive (http://identifiers.org/insdc.sra:SRR34586846) [7]
Data file 2	ONT long reads	Sequence file (.fastq)	Sequence Read Archive (http://identifiers.org/insdc.sra:SRR34586845) [8]
Data file 3	Complete genome sequence of *H. thermoluteolus* KS0102	Sequence file (.gbk)	Nucleotide Sequence Database (http://identifiers.org/insdc:CP196346) [9]
Data file 4	Circular map of *H. thermoluteolus* KS0102 genome	Portable Document Format file (.pdf)	Figshare (https://doi.org/10.6084/m9.figshare.29643509) [10]
Data file 5	Genome quality assessment metrics for *H. thermoluteolus* KS0102 based on CheckM2 analysis	MS Excel file (.xlsx)	Figshare (https://doi.org/10.6084/m9.figshare.29643563) [11]
Data file 6	Digital DNA-DNA hybridisation results	MS Excel file (.xlsx)	Figshare (https://doi.org/10.6084/m9.figshare.29643692) [12]
Data file 7	Average nucleotide identity analysis results for *H. thermoluteolus* KS0102	MS Excel file (.xlsx)	Figshare (https://doi.org/10.6084/m9.figshare.30296377) [13]
Data file 8	Phylogenomic tree of strain KS0102 and related bacterial species	Portable Document Format file (.pdf)	Figshare (https://doi.org/10.6084/m9.figshare.29643860) [14]
Data file 9	Nucleotide sequence of the PHA synthase gene (*phaC1*) from *H. thermoluteolus* KS0102	Text file (.txt)	Figshare (https://doi.org/10.6084/m9.figshare.29643965) [15]
Data file 10	Nucleotide sequence of the PHA synthase gene (*phaC2*) from *H. thermoluteolus* KS0102	Text file (.txt)	Figshare (https://doi.org/10.6084/m9.figshare.29643974) [16]
Data file 11	Deduced amino acid sequence of PHA synthase (PhaC1) from *H. thermoluteolus* KS0102	Text file (.txt)	Figshare (https://doi.org/10.6084/m9.figshare.29643980) [17]
Data file 12	Deduced amino acid sequence of PHA synthase (PhaC2) from *H. thermoluteolus* KS0102	Text file (.txt)	Figshare (https://doi.org/10.6084/m9.figshare.29643989) [18]
Data file 13	Intact prophage region identified in *H. thermoluteolus* KS0102 genome	Portable Document Format file (.pdf)	Figshare (https://doi.org/10.6084/m9.figshare.29644004) [19]
Data file 14	Prophage analysis results for *H. thermoluteolus* KS0102	MS Excel file (.xlsx)	Figshare (https://doi.org/10.6084/m9.figshare.29644010) [20]
Data file 15	Carbohydrate-active enzyme analysis results for *H. thermoluteolus* KS0102	MS Excel file (.xlsx)	Figshare (https://doi.org/10.6084/m9.figshare.30296389) [21]
Data file 16	COG functional classification of *H. thermoluteolus* KS0102	Portable Document Format file (.pdf)	Figshare (https://doi.org/10.6084/m9.figshare.30296920) [22]

achieving yields of up to 80 wt% cell dry weight from xylose substrates at an optimal growth temperature of 55 °C [5]. Thermophilic growth conditions help reduce the risk of microbial contamination and facilitate non-sterile biotechnological processes [6]. Hot springs and hot pools in Thailand remain largely unexplored for their biotechnological applications. Strain KS0102, isolated from the Mae Kasa hot spring sediments in Tak Province, exhibited optimal growth at 50–55 °C and demonstrated PHB synthase activity, indicating its potential for biodegradable polymer production.

Complete genome sequencing provides insights into thermostability and biopolymer synthesis pathways. Genome data enable taxonomic identification, comparative genomics within *Hydrogenophilus*, and biotechnological development. This study presents the complete genome sequence of *H. thermoluteolus* KS0102 to support taxonomic identification, enable comparative genomics within the genus *Hydrogenophilus*, and provide genomic resources for understanding thermophilic PHB production pathways.

Data description

KS0102 was isolated from hot spring sediment samples collected from the Mae Kasa hot spring in Tak Province, Thailand (16°53'08.3"N, 98°38'09.1"E) using Mueller Hinton agar at 50 °C. Genomic DNA was extracted using Quick-DNA HMW™ MagBeads Kit (Zymo Research, USA) from overnight cultures grown in Mueller-Hinton broth at 50 °C with 200 rpm agitation. Short-read libraries were prepared using the Nextera XT DNA library preparation kit (Illumina, USA) and sequenced on the NextSeq 550, generating 150-bp paired-end reads. These high-quality reads, which are available in the SRA under accession number SRR34586846 (Table 1, Data file 1), were subjected to quality-trimming with fastp v0.23.4 [23]. Long-read libraries were prepared using the SQK-RAD004 rapid barcoding sequencing kit (Oxford Nanopore Technologies, UK) and sequenced on the Oxford Nanopore FLO-MIN106 flow cell, generating reads with an average length of 6,443 bp. These high-quality reads were processed using NanoFilt v1.1.0 [24], which are available in the SRA under accession number

SRR34586845 (Table 1, Data file 2). A hybrid assembly integrating both short and long reads was performed using Unicycler v0.5.0 [25]. The assembled genome, functionally annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) [26], exhibits a single-replicon architecture with high sequencing depth (617-fold average coverage), consisting of a 2,136,028-bp closed circular genome available under accession number CP196346 (Table 1, Data file 3). A circular genome map depicting the chromosomal replicon of KS0102 was generated using Bakta v1.10.4 [27] visualisation module, illustrating key genomic features and their distributions (Table 1, Data file 4). Genome completeness was validated using CheckM2 v1.0.1 [28], revealing 99.96% completeness with no contamination (Table 1, Data file 5). The taxonomic classification using the Type (Strain) Genome Server (TYGS) [29] revealed that strain KS0102 belongs to *H. thermoluteolus* with a dDDH value of 96.3% compared with the type strain NBRC 14,978^T (Table 1, Data file 6). Average nucleotide identity (ANI) analysis was performed using JSpeciesWS [30] with default parameters for ANIb and ANIm and EzBioCloud [31] with default parameters for OrthoANIu. Strain KS0102 showed 97.72% OrthoANIu, 97.74% ANIb, and 97.82% ANIm identity with *H. thermoluteolus* NBRC 14,978^T (Table 1, Data file 7), confirming species-level identity (>95% ANI threshold). The minimal difference in the G + C content (0.19%) supports reliable species identification. Phylogenomic analysis showed that KS0102 closely clustered with *H. thermoluteolus* NBRC 14,978^T with 86% bootstrap support, confirming taxonomic placement (Table 1, Data file 8). Comprehensive genomic annotation using Bakta v1.10.4 [27] identified two PHB biosynthesis genes in the KS0102 genome, comprising two PHB synthase genes: *phaC1* (Table 1, Data file 9) and *phaC2* (Table 1, Data file 10), which encode the key enzyme for polyhydroxybutyrate production. BLAST was used to validate protein sequences against the UniProt database with default parameters to confirm functional annotations and assess evolutionary relationships. The *phaC1* gene product (597 amino acids) (Table 1, Data file 11) demonstrated 99.3% sequence identity and 99.5% similarity to Class I poly(R)-hydroxyalkanoic acid synthase from *H. thermoluteolus* (UniProt: A0A2Z6DZ60), confirming its classification as a Class I PHA synthase. The *phaC2* gene product (570 amino acids) (Table 1, Data file 12) showed 99.1% sequence identity and 100% similarity to poly-beta-hydroxybutyrate synthase from *H. thermoluteolus* (UniProt: A0A2Z6DWT1), validating its functional role in PHB biosynthesis. Prophage identification using PHASTEST [32] revealed one intact prophage region (33.6 kb; positions 961,422–995,034) comprising 35 phage-related proteins, including 7 hypothetical proteins (Table 1, Data file 13 and 14). The region contains key structural genes (terminase, head, capsid, protease, tail, and plate proteins) with the highest number of gene matches to BcepMu and phiE255. The prophage GC content (66.20%) exceeded that of the host genome (61.8%) (Table 1, Data file 14). Secondary metabolite biosynthetic gene cluster analysis was performed using antiSMASH version 7.0 [33] with default parameters. No notable biosynthetic gene clusters (BGCs) were detected within the KS0102 genome. Identification of carbohydrate-active enzyme (CAZyme) was performed using dbCAN3 [34] with default parameters. A total of 35 genes, which encode 36 distinct CAZyme domains, were identified, including 19 glycosyl transferases (GT), 13 glycoside hydrolases (GH), 2 carbohydrate esterases (CE), and 2 auxiliary activities (AA) (Table 1, Data file 15). Functional annotation and classification of anticipated proteins were executed using COG (Clusters of orthologous groups) analysis employing COGclassifier v1.2.0 [35] with standard parameters. A total of 2,066 proteins were allocated to 26 functional categories (Table 1, Data file 16). The most abundant categories were translation and ribosomal structure (category J, 200 proteins), energy production and conversion (category C, 183 proteins), amino acid transport and metabolism (category E, 177 proteins), cell wall/membrane/envelope biogenesis (category M, 160 proteins), and signal transduction mechanisms (category T, 140 proteins).

Limitations

Although this study provides a complete genome sequence of *H. thermoluteolus* strain KS0102, several limitations exist. Functional gene identification solely relied on computational prediction and annotation. Future studies should focus on experimental validation to confirm gene functions and understand their roles in thermophilic adaptation and biodegradable polymer synthesis.

Abbreviations

ANI	Average nucleotide identity
CAZyme	Carbohydrate-active enzyme
dDDH	Digital DNA-DNA hybridisation
ONT	Oxford Nanopore Technology
PGAP	NCBI Prokaryotic Genome Annotation Pipeline
PHB	Polyhydroxybutyrate
SRA	Sequence read archive
TYGS	Type (Strain) Genome Server

Acknowledgements

This research project was supported by the Chulabhorn Graduate Institute (Fundamental Fund by National Science Research and Innovation Fund (NSRF): fiscal year 2025, FRB680079/0518 Project code 209042), the Center of Excellence on Environmental Health and Toxicology (EHT), OPS, Ministry of Higher Education, Science, Research and Innovation, and the research unit for applied microbiology, Faculty of Allied Health Sciences, Burapha University.

Author contributions

MY conceptualised and designed the research framework; PR, WY, MP, TS and PS conducted laboratory experiments and performed computational

genomic analyses; WY and PR contributed essential research materials; MY secured funding for the project; MT and KS prepared the initial manuscript draft; and MY critically reviewed and revised the manuscript for intellectual content.

Funding
Chulabhorn Graduate Institute (Fundamental Fund by National Science Research and Innovation Fund (NSRF): fiscal year 2025, FRB680079/0518 Project code 209042).

Data availability
The complete genome sequence and associated data described in this Data note have been deposited in GenBank under the BioProject accession number PRJNA493966. Supplementary analytical results, annotations, and comparative genomic data are freely accessible through the Figshare repository. Table 1 provides comprehensive access information and direct links to all datasets.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 31 July 2025 / Accepted: 29 October 2025
Published online: 24 November 2025

References

1. Sitara A, Hocq R, Horvath J, Pflügl S. Industrial biotechnology goes thermophilic. Thermoanaerobes as promising hosts in the circular carbon economy. Bioresour Technol. 2024;408:131164.
2. Adnan M, Siddiqui AJ, Ashraf SA, et al. Characterization and process optimization for enhanced production of polyhydroxybutyrate (PHB)-based biodegradable polymer from *Bacillus flexus* isolated from municipal solid waste landfill site. Polym (Basel). 2023;15(6):1407.
3. Levett L, Birkett G, Davies N, et al. Techno-economic assessment of poly-3-hydroxybutyrate (PHB) production from methane - The case for thermophilic bioprocessing. J Environ Chem Eng. 2016;4(4):3733–42.
4. Burkhardt C, Baruth L, Neele Meyer-Heydecke, et al. Mining thermophiles for biotechnologically relevant enzymes: evaluating the potential of European and Caucasian hot springs. Extremophiles. 2023;28(1):5.
5. Kourilova X, Pernicova I, Sedlar K, et al. Production of polyhydroxyalkanoates (PHA) by a thermophilic strain of *Schlegelella thermodepolymerans* from xylose rich substrates. Bioresour Technol. 2020;315:123885.
6. Musilova J, Kourilova X, Bezdicek M, et al. First complete genome of the thermophilic polyhydroxyalkanoates-producing bacterium *Schlegelella thermodepolymerans* DSM 15344. Genome Biol Evol. 2021;13(4):evab007.
7. NCBI. Paired-end Illumina reads. 2025. http://identifiers.org/insdc.sra:SRR34586846.
8. NCBI. ONT long reads. 2025. http://identifiers.org/insdc.sra:SRR34586845.
9. NCBI. Complete genome sequence of *H. thermoluteolus* KS0102. 2025. http://identifiers.org/insdc:CP196346.
10. Yasawong M. Circular map of *H. thermoluteolus* KS0102 genome. 2025. https://doi.org/10.6084/m9.figshare.29643509.
11. Yasawong M. Genome quality assessment metrics for *H. thermoluteolus* KS0102 based on CheckM2 analysis. 2025. https://doi.org/10.6084/m9.figshare.29643563.
12. Yasawong M, Digital. DNA-DNA hybridisation results. 2025. https://doi.org/10.6084/m9.figshare.29643692.
13. Yasawong M. Average nucleotide identity analysis results for *H. thermoluteolus* KS0102. 2025. https://doi.org/10.6084/m9.figshare.30296377.
14. Yasawong M. Phylogenomic tree of strain KS0102 and related bacterial species. 2025. https://doi.org/10.6084/m9.figshare.29643860.
15. Yasawong M. Nucleotide sequence of the PHA synthase gene (*phaC1*) from *H. thermoluteolus* KS0102. 2025. https://doi.org/10.6084/m9.figshare.29643965.
16. Yasawong M. Nucleotide sequence of the PHA synthase gene (*phaC2*) from *H. thermoluteolus* KS0102. 2025. https://doi.org/10.6084/m9.figshare.29643974.
17. Yasawong M. Deduced amino acid sequence of PHA synthase (PhaC1) from *H. thermoluteolus* KS0102. 2025. https://doi.org/10.6084/m9.figshare.29643980.
18. Yasawong M. Deduced amino acid sequence of PHA synthase (PhaC2) from *H. thermoluteolus* KS0102. 2025. https://doi.org/10.6084/m9.figshare.29643989.
19. Yasawong M. Intact prophage region identified in *H. thermoluteolus* KS0102 genome. 2025. https://doi.org/10.6084/m9.figshare.29644004.
20. Yasawong M. Prophage analysis results for *H. thermoluteolus* KS0102. 2025. https://doi.org/10.6084/m9.figshare.29644010.
21. Yasawong M. Carbohydrate-active enzyme analysis results for *H. thermoluteolus* KS0102. 2025. https://doi.org/10.6084/m9.figshare.30296389.
22. Yasawong M. COG functional classification of *H. thermos luteolus* KS0102. 2025. https://doi.org/10.6084/m9.figshare.30296920.
23. Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using Fastp. Imeta. 2023;2(2):e107.
24. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics. 2018;34(15):2666–9.
25. Wick RR, Judd LM, Gorrie CL, Holt KE, Unicycler. Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13(6):e1005595.
26. Tatusova T, DiCuccio M, Badretdin A, et al. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 2016;44(14):6614–24.
27. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. Microb Genom. 2021;7(11):000685.
28. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. Nat Methods. 2024;21(4):735.
29. Meier-Kolthoff JP, Göker M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat Commun. 2019;10(1):2182.
30. Richter M, Rosselló-Móra R, Oliver Glöckner F, et al. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics. 2016;32(6):929–31.
31. Yoon SH, Ha SM, Lim J, et al. A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek. 2017;110(10):1281–6.
32. Wishart DS, Han S, Saha S, et al. PHASTEST: faster than PHASTER, better than PHAST. Nucleic Acids Res. 2023;51(W1):W443–50.
33. Blin K, Shaw S, Augustijn HE, et al. AntiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. Nucleic Acids Res. 2023;51(W1):W46–50.
34. Zheng J, Ge Q, Yan Y, et al. dbCAN3: automated carbohydrate-active enzyme and substrate annotation. Nucleic Acids Res. 2023;51(W1):W115–21.
35. Shimoyama Y, COGclassifier. A tool for classifying prokaryote protein sequences into COG functional category. 2022. https://github.com/moshi4/COGclassifier.

Publisher's note