

DATA NOTE

Open Access



# Genome sequence and assembly of the amylolytic *Bacillus licheniformis* T5 strain isolated from Kazakhstan soil

Arman Mussakhmetov<sup>1</sup>, Assel Kiribayeva<sup>1</sup>, Asset Daniyarov<sup>2,4</sup>, Aitbay Bulashev<sup>3</sup>, Ulykbek Kairov<sup>2\*</sup> and Bekbolat Khassenov<sup>1\*</sup>

## Abstract

**Objectives** The data presented in this study were collected with the aim of obtaining the complete genomes of specific strains of *Bacillus* bacteria, namely, *Bacillus licheniformis* T5. This strain was chosen based on its enzymatic activities, particularly amylolytic activity. In this study, nanopore sequencing technology was employed to obtain the genome sequences of this strain. It is important to note that these data represent a focused objective within a larger research context, which involves exploring the biochemical features of promising *Bacillus* strains and investigating the relationship between enzymatic activity, phenotypic features, and the microorganism's genome.

**Data description** In this study, the whole-genome sequence was obtained from one *Bacillus* strain, *Bacillus licheniformis* T5, isolated from soil samples in Kazakhstan. Sample preparation and genomic DNA library construction were performed according to the Ligation sequencing gDNA kit (SQK-LSK109) protocol and NEBNext module. The prepared library was sequenced on a MinION instrument (Oxford Nanopore Technologies nanopore sequencer with a maximum throughput of up to 30 billion nucleotides per run and no limit on read length), using a flow cell for nanopore sequencing FLO-MIN106D. The genome de novo assembly was performed using the long sequencing reads generated by MinION Oxford Nanopore platform. Finally, one circular contig was obtained harboring a length of 4,247,430 bp with 46.16% G + C content and the mean contig 428X coverage. *B. licheniformis* T5 genome assembly annotation revealed 5391 protein-coding sequences, 81 tRNAs, 51 repeat regions, 24 rRNAs, 3 virulence factors and 53 antibiotic resistance genes. This sequence encompasses the complete genetic information of the strain, including genes, regulatory elements, and noncoding regions. The data reveal important insights into the genetic characteristics, phenotypic traits, and enzymatic activity of this *Bacillus* strain.

The findings of this study have particular value to researchers interested in microbial biology, biotechnology, and antimicrobial studies. The genomic sequence offers a foundation for understanding the genetic basis of traits such as endospore formation, alkaline tolerance, temperature range for growth, nutrient utilization, and enzymatic activities. These insights can contribute to the development of novel biotechnological applications, such as the production of enzymes for industrial purposes.

\*Correspondence:

Ulykbek Kairov  
ulykbek.kairov@nu.edu.kz  
Bekbolat Khassenov  
khassenov@biocenter.kz

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Overall, this study provides valuable insights into the genetic characteristics, phenotypic traits, and enzymatic activities of the *Bacillus licheniformis* T5 strain. The acquired genomic sequences contribute to a better understanding of this strain and have implications for various research fields, such as microbiology, biotechnology, and antimicrobial studies.

**Keywords** *Bacillus licheniformis* strain, Oxford Nanopore sequencing, Genome assembly, Enzymatic activities, Kazakhstan

## Objective

The objective of this study was to utilize Oxford Nanopore sequencing technology to obtain the complete genome sequences of a specific *Bacillus* strain, *Bacillus licheniformis* T5, which is known for its amylolytic activity. This strain exhibits significant potential for industrial applications [1]. *Bacillus licheniformis* T5 produces a thermostable  $\alpha$ -amylase with high pH stability. The utilization of Oxford Nanopore sequencing allows long read sequencing and rapid acquisition of genomic data, enabling a comprehensive analysis of the genetic factors underlying the enzymatic activities of this strain.

Bacilli species are renowned for their diverse enzymatic capabilities, making them valuable in various industries, such as biocatalysis, hydrolysis of proteins [2], degradation of plant polymers [3], biofuel production, and starch food processing [4]. The amylases derived from this *Bacillus* strain have found extensive industrial applications. *B. licheniformis* T5 with specific enzymatic activity holds immense potential for optimization and utilization in these sectors [5].

The comprehensive analysis of the complete genome of *Bacillus licheniformis* T5 offers valuable insights into the genetic basis of its enzymatic activity. This knowledge can be utilized for genetic engineering approaches and optimization strategies, ultimately enhancing the industrial applications of this strain.

In summary, this study focuses on the application of Oxford Nanopore sequencing technology to obtain the complete genome sequence of *Bacillus licheniformis* T5. This strain possesses amylolytic activity and demonstrates significant potential for industrial use. The genomic information acquired through this study contributes to our understanding of the genetic factors underlying the enzymatic capabilities, facilitating further research and applications in diverse industrial sectors.

## Data description

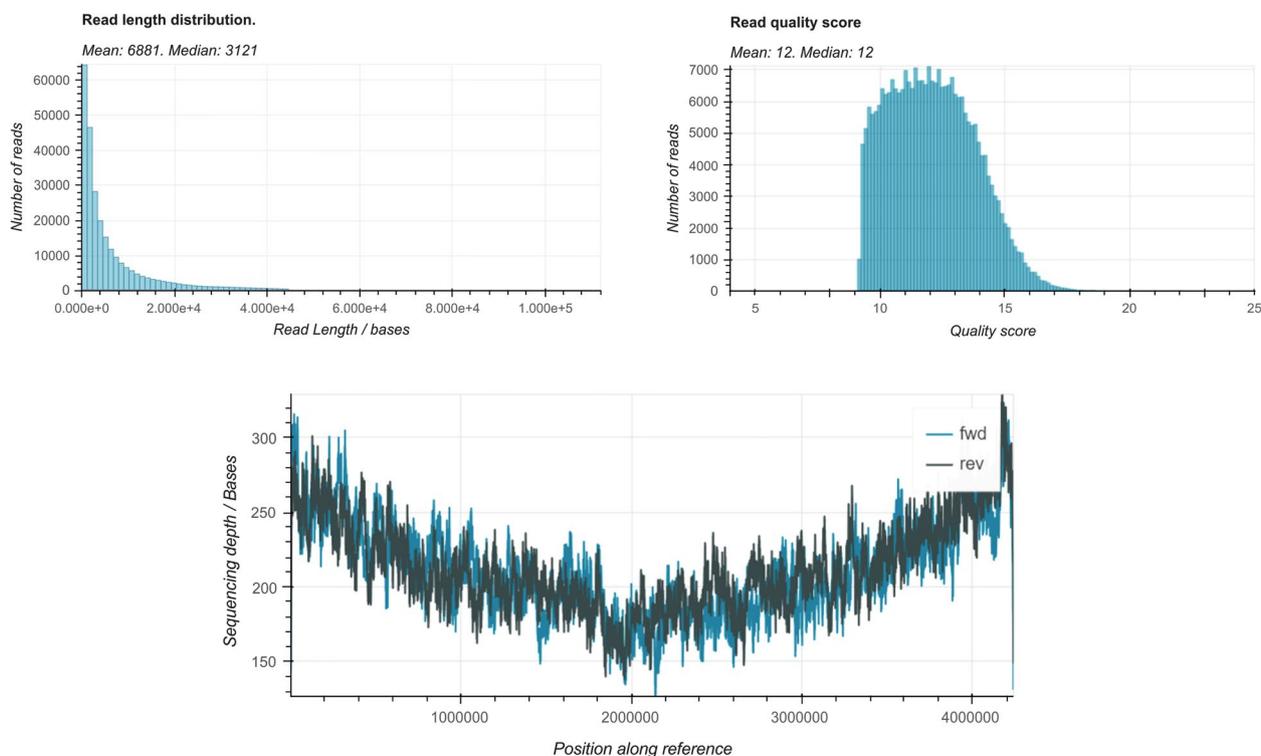
Genomic sequences obtained from *Bacillus licheniformis* strain T5 were acquired in this study. This strain was isolated from soil samples collected in Kazakhstan. The data

revealed that the strain has the ability to form endospores and can thrive within a temperature range of 30–60 °C. It exhibited growth on various nutrient media, including nutrient broth/agar, Luria–Bertani medium, and Mueller–Hinton agar. Additionally, the *B. licheniformis* T5 strain was found to be alkaline-tolerant and capable of growing within a pH range of 5.5–8.0. Furthermore, this strain displayed sensitivity to several antibiotics, including clindamycin, rifampicin, erythromycin, ciprofloxacin, tobramycin, tetracycline, penicillin, gentamicin, ampicillin, kanamycin, streptomycin, and chloramphenicol [6].

The genomic sequences provide a comprehensive representation of the genetic information present in *Bacillus licheniformis* T5. This sequence encompasses the complete set of genes, regulatory elements, and noncoding regions that constitute their genomes.

During a 5-day culture period on Difco sporulation medium (Sigma-Aldrich, UK), Arret-Kirshbaum sporulation agar, and modified nutrient agar, the strain was observed to develop endospores. This strain possesses distinct enzymatic characteristics. *B. licheniformis* T5 displays amylolytic activity, producing thermostable  $\alpha$ -amylase when grown on starch medium, with maximum activity observed at 80 °C and pH 6.0. The  $\alpha$ -amylase isolated from *B. licheniformis* T5 has a number of features that distinguish it favorably from  $\alpha$ -amylases isolated from related strains:  $\alpha$ -amylase retains 100% activity after preliminary incubation of the enzyme for 10 h in buffers with pH from 6 to 12; retains 100% activity in the presence of 1%  $\beta$ -mercaptoethanol and is not inhibited by SDS at concentrations of 10 and 20 mM. These indicators allow us to consider  $\alpha$ -amylase as a promising enzyme for industrial use, and the *B. licheniformis* T5 strain as a producer strain.

To process the data, the collected strain was cultured in 10 mL of nutrient broth (Himedia, India) in a shaker incubator at 37 °C and 150 rpm for 18 h. Following culture, cells were collected via centrifugation at 6000 $\times$ g, and genomic DNA was isolated using a Genomic DNA Purification Kit (Promega, USA) according to the manufacturer's protocol. The quality of the isolated DNA was assessed through spectrophotometry and agarose gel electrophoresis.



**Fig. 1** *Bacillus licheniformis* T5 strain ONT sequencing read quality control and genome coverage representation

Construction of genomic DNA libraries was performed using the Oxford Nanopore Technologies (ONT) sequencing kit (SQK-LSK109). This involved DNA fragmentation followed by adapter ligation [7]. The resulting libraries were quantified using Qubit 2.0 (Invitrogen, USA) and subjected to sequencing on a MinION platform (<https://nanoporetech.com/>) with a FLO-MIN106D flow cell (R9). Raw sequence files were processed using Guppy v3.4.1 to call the reads, and low-quality reads were removed from further analysis. A total of 262,436 reads with a mean read length of 3121 bp and a mean read quality of 12.11 were obtained [8] (Fig. 1). The Epi2me “epi2me-labs/wf-bacterial-genomes” pipeline was implemented to perform further analysis and assembly. The genome de novo assembly was performed using Flye v2.9.1 [9] designed for the long sequencing reads generated by ONT. One circular contig was obtained harboring a length of 4,247,430 bp with 46.16% G + C content and high mean 428X contig coverage. The resulting sequences from ONT were annotated using Prokka rapid prokaryotic genome annotation (Prokka v1.13.7) [10] and DNA Features Viewer software v.3.1.2. Additional genome annotation and genome circular visualization [8] were performed using the PATRIC [11] which is now readily accessible

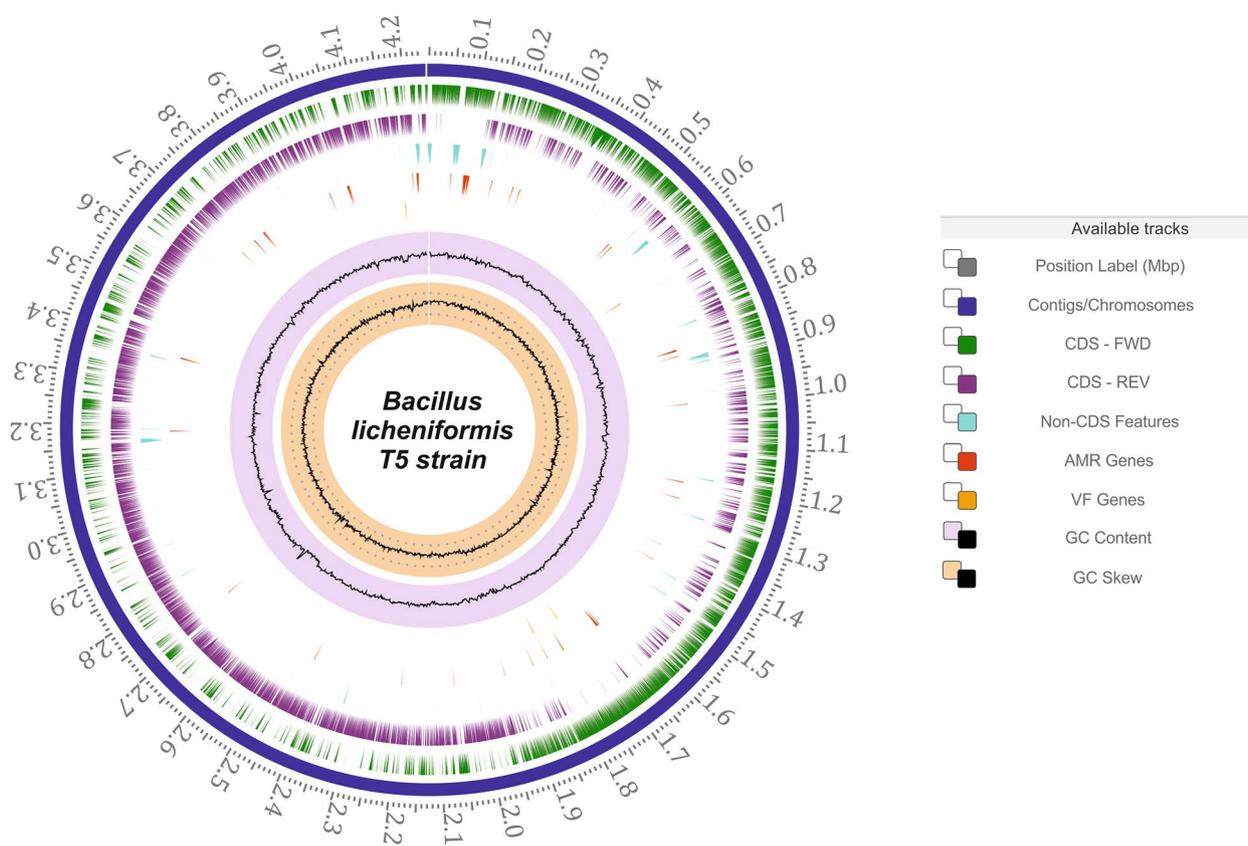
through the BV-BRC [12] platforms (<https://www.bv-brc.org/>). The BUSCO v5.4.7 [13] tool was used to assess the final genome assembly. Genome assembly annotation revealed 5391 protein-coding sequences, 81 tRNAs, 51 repeat regions, 24 rRNAs, 3 virulence factors and 53 antibiotic resistance genes (according to PATRIC DB) [14].

The resulting genome assembly of *Bacillus licheniformis* T5 has been deposited in NCBI GenBank with the accession number CP124852 under the Bioproject number PRJNA967102, as shown in Table 1 [15].

In summary, the data collected in this study involve the genome sequence assembly and annotation [16] of *Bacillus licheniformis* T5 (Fig. 2). This assembly offers a comprehensive understanding of the genetic information within this strain, including the presence of specific genetic elements responsible for observed phenotypic traits. The data were processed through various steps, such as bacterial cultivation and growth observation under different conditions, DNA isolation, library preparation, Oxford Nanopore sequencing, base calling, genome de novo assembly and genome annotation. The resulting assembly has been deposited in NCBI GenBank and is freely available for further comparative studies and explorations.

**Table 1** Overview of data files/data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	Genome assembly of <i>Bacillus licheniformis</i> strain T5	Fasta file (.fasta)	NCBI GenBank BioProject: <a href="https://identifiers.org/ncbi/bioproject:PRJNA967102">https://identifiers.org/ncbi/bioproject:PRJNA967102</a> [15]
Data file 2	ONT sequencing read quality control, genome coverage and circular genome representation of <i>Bacillus licheniformis</i> strain T5	Word file (.docx)	Figshare <a href="https://doi.org/10.6084/m9.figshare.24018087">https://doi.org/10.6084/m9.figshare.24018087</a> [8]
Data file 3	Read and assembly statistics, Genome statistics, features and specialty genes of <i>Bacillus licheniformis</i> strain T5	Word file (.docx)	Figshare <a href="https://doi.org/10.6084/m9.figshare.24025071">https://doi.org/10.6084/m9.figshare.24025071</a> [14]
Data file 4	Genome assembly annotation results	Excel file (.xlsx)	Figshare <a href="https://doi.org/10.6084/m9.figshare.24025113">https://doi.org/10.6084/m9.figshare.24025113</a> [16]



**Fig. 2** Circular genome representation of *Bacillus licheniformis* T5 strain

**Limitations**

Oxford Nanopore technology has been used to generate the genome sequences of *B. licheniformis* T5 strain in order to ensure complete assembly. Nanopore sequencing differs from these earlier methods in that it directly detects nucleotides without active DNA synthesis, as a long stretch of single-stranded DNA passes through a protein nanopore stabilized by an electrically refractory polymer membrane [17, 18]. The nanopore

sequencing does not require imaging apparatus to detect nucleotides, making the system portable and significantly reduces the initial cost of the full-genome sequencing [19]. The important elements of nanopore sequencing are a membrane with nanometer-sized pores and a chamber filled with an electrolytic solution. The principle of operation is that when nucleotides pass through the pore, the cross-section available for ions decreases and the current strength, which is measured,

falls accordingly [20]. The accuracy of the method is determined by the number of times the DNA chain passes through the pore [21]. Genome assembly, analysis and further annotation were performed using novel, robust and validated bioinformatics methods and tools. Whole-genome was sequenced with high coverage and the final genome assembly finalized as the one circular contig [14]. Therefore, the authors are not aware of any limitations in the data.

#### Abbreviations

ONT	Oxford Nanopore Technology
DNA	Deoxyribonucleic acid
bp	Base pair
NCBI	National Center for Biotechnology Information

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12863-023-01177-8>.

**Additional file 1: Table S1.** Genome statistics of *Bacillus licheniformis* T5 strain. **Table S2.** Genomic features of *Bacillus licheniformis* T5 strain. **Table S3.** Specialty genes of *Bacillus licheniformis* T5 strain. **Table S4.** Read and assembly statistics of *Bacillus licheniformis* T5 strain.

**Additional file 2.**

#### Acknowledgements

The authors would like to thank Dr. A. Shevtsov and A. Amirgazin (National Center for Biotechnology) for providing access to the ONT MinION sequencing platform and software.

#### Authors' contributions

BK, AM and UK wrote the manuscript. AM, AK and BK contributed to the experimental work. UK and AD contributed to the bioinformatics analysis and implementation of software/pipelines/codes. BK and UK contributed to the interpretation and critical revision of the manuscript. BK and AB conceptualized and supervised the research and were involved in funding acquisition. All authors contributed to the article and approved the final manuscript.

#### Funding

The research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP14869708) and the Ministry of Agriculture of the Republic of Kazakhstan (Grant No. BR10764944). UK and AD have been supported by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. BR18574184).

#### Availability of data and materials

The data described in this manuscript are available and openly accessed on NCBI GenBank under Bioproject no. PRJNA967102. Please see Table 1 and references [8, 14–16] for details and links to the data.

#### Declarations

##### Ethics approval and consent to participate

Not applicable.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare no competing interests.

#### Author details

<sup>1</sup>National Center for Biotechnology, 13/5 Korgalzhyn Road, Astana 010000, Kazakhstan. <sup>2</sup>Laboratory of Bioinformatics and Systems Biology, Center for Life Sciences, National Laboratory Astana, Nazarbayev University, 53 Kabanbay Batyr Avenue, Astana 010000, Kazakhstan. <sup>3</sup>S. Seifullin Kazakh Agrotechnical Research University, 62 Zhenis Avenue, Astana 010001, Kazakhstan. <sup>4</sup>Faculty of Natural Sciences, L.N. Gumilyev, Eurasian National University, 2 Kanysh Satpayev Street, Astana 010008, Kazakhstan.

Received: 25 August 2023 Accepted: 23 November 2023

Published online: 02 January 2024

#### References

- Muras A, Romero M, Mayer C, Otero A. Biotechnological applications of *Bacillus licheniformis*. Crit Rev Biotechnol. 2021;41(4):609–27. <https://doi.org/10.1080/07388551.2021.1873239>.
- Aktayeva S, Baltin K, Kiribayeva A, Akishev Z, Silayev D, Ramankulov Y, Khassenov B. Isolation of *Bacillus* sp. A5.3 strain with keratinolytic activity. Biology (Basel). 2022;11(2):244. <https://doi.org/10.3390/biology11020244>.
- Kiribayeva A, Mukanov B, Silayev D, Akishev Z, Ramankulov Y, Khassenov B. Cloning, expression, and characterization of a recombinant xylanase from *Bacillus sonorensis* T6. PLoS One. 2022;17(3).
- Fincan SA, Özdemir S, Karakaya A, Enez B, Mustafaov SD, Ulutaş MS, Şen F. Purification and characterization of thermostable  $\alpha$ -amylase produced from *Bacillus licheniformis* So-B3 and its potential in hydrolyzing raw starch. Life Sci. 2021;1(264): 118639. <https://doi.org/10.1016/j.lfs.2020.118639>.
- Rana N, Walia A, Gaur A.  $\alpha$ -Amylases from microbial sources and its potential applications in various industries. Natl Acad Sci Lett. 2013;36(1):9–17. <https://doi.org/10.1007/s40009-012-0104-0>.
- Aktayeva S, Kiribayeva A, Makasheva D, Astrakhanov M, Tursunbekova A, Baltin K, Khassenov B. Isolation, identification and usage of *Bacillus* strains in microbial inhibition test in milk. Eur J Appl Biotechnol. 2022 (4) P49–57. <https://doi.org/10.11134/btp.4.2022.6>.
- Ligation Sequencing Kit (SQK-LSK109) [Sequencing kit]. Oxford Nanopore Technologies. Retrieved from <https://pr.vwr.com/store/product/36311524/ligation-sequencing-kits-oxford-nanopore-technologies>. Accessed 23 Aug 2023.
- Data file 2: ONT sequencing read quality control, genome coverage and circular genome representation of *Bacillus licheniformis* strain T5. Figshare. (2023). <https://doi.org/10.6084/m9.figshare.24018087>.
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37(5):540–6. <https://doi.org/10.1038/s41587-019-0072-8>.
- Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068–9. <https://doi.org/10.1093/bioinformatics/btu153>.
- Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM, Disz T, Gabbard JL, Gerdes S, Henry CS, Kenyon RW, Machi D, Mao C, Nordberg EK, Olsen GJ, Murphy-Olson DE, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Vonstein V, Warren A, Xia F, Yoo H, Stevens RL. Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. Nucleic Acids Res. 2017;45(D1):D535–42. <https://doi.org/10.1093/nar/gkw1017>.
- Olson RD, Assaf R, Brettin T, Conrad N, Cucinell C, Davis JJ, Dempsey DM, Dickerman A, Dietrich EM, Kenyon RW, Kuscuoglu M, Lefkowitz EJ, Lu J, Machi D, Macken C, Mao C, Niewiadomska A, Nguyen M, Olsen GJ, Overbeek JC, Parrello B, Parrello V, Porter JS, Pusch GD, Shukla M, Singh I, Stewart L, Tan G, Thomas C, VanOeffelen M, Vonstein V, Wallace ZS, Warren AS, Wattam AR, Xia F, Yoo H, Zhang Y, Zmasek CM, Scheuermann RH, Stevens RL. Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR Nucleic Acids Res. 2023;51(D1):D678–89. <https://doi.org/10.1093/nar/gkac1003>.
- Manni M, Berkeley MR, Seppely M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol. 2021;38(10):4647–54. <https://doi.org/10.1093/molbev/msab199>.
- Data file 3: Genome statistics, features and specialty genes of *Bacillus licheniformis* strain T5. Figshare. (2023). <https://doi.org/10.6084/m9.figshare.24025071>.

15. Data file 1: NCBI GenBank Bioproject. 2023. <https://identifiers.org/ncbi/bioproject:PRJNA967102>. Accessed 23 Aug 2023.
16. Data file 4: Genome assembly annotation results. Figshare. 2023. <https://doi.org/10.6084/m9.figshare.24025113>.
17. Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, et al. The potential and challenges of nanopore sequencing. *Nat Biotechnol*. 2008;26(10):1146–53. <https://doi.org/10.1038/nbt.1495>.
18. Deamer D, Akeson M, Branton D. Three decades of nanopore sequencing. *Nat Biotechnol*. 2016;34(5):518–24. <https://doi.org/10.1038/nbt.3423>.
19. Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol*. 2016;17(1):239. Epub 2016/11/27. <https://doi.org/10.1186/s13059-016-1103-0>.
20. Mikheyev AS, Tin MMY. A first look at the Oxford Nanopore MinION sequencer. *Mol Ecol Resour*. 2014;14(6):1097–102. <https://doi.org/10.1111/1755-0998.12324>.
21. Feng Y, Zhang Y, Ying C, Wang D, Du C. Nanopore-based fourth-generation DNA sequencing technology. *Genomics, proteomics & bioinformatics*. 2015;13(1):4–16. Epub 2015/03/07. <https://doi.org/10.1016/j.gpb.2015.01.009>.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

