Data in Brief 57 (2024) 110911

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Whole genome sequence data of ZEN-degrading strain *Levilactobacillus brevis* PYN10_6_2 isolated from *Tenebrio molitor* larval feces

Mengru Zhao[a,b], Baoyuan Guo[a,b], Shuzhen Jiang[c], Yang Wang[b,*]

[a] *School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China*
[b] *Institute of Grain and Oil Quality and Safety, Academy of National Food and Strategic Reserves Administration, Beijing 100834, China*
[c] *College of Animal Science and Technology, Shandong Agricultural University, Tai'an 271018, China*

ARTICLE INFO

Article history:
Received 21 June 2024
Revised 27 August 2024
Accepted 30 August 2024
Available online 4 September 2024

Dataset link: Levilactobacillus brevis strain PYN10-6-2 plasmid pLB1106_4, complete sequence (Original data)
Dataset link: Levilactobacillus brevis strain PYN10-6-2 plasmid pUCCLB521_C, complete sequence (Original data)
Dataset link: Levilactobacillus brevis strain PYN10-6-2 plasmid pLB1174-1, complete sequence (Original data)
Dataset link: Levilactobacillus brevis strain:PYN10_6_2 Raw sequence reads (Original data)
Dataset link: Levilactobacillus brevis strain PYN10-6-2 chromosome, complete genome (Original data)

ABSTRACT

Levilactobacillus brevis PYN10_6_2, a lactic acid bacterial strain previously isolated from *Tenebrio molitor* larval feces, possesses the ability to convert zearalenone (ZEN) to α-/β-Zearalenol (α-/β-ZEL). However, the genes involved in the ZEN reduction reaction and the biosafety of this strain remain unknown. In this study, we sequenced, assembled, and annotated the whole genome of *L. brevis* PYN10_6_2. Genomic sequencing was conducted using short-read sequencing on the Illumina HiSeq X Ten platform and long-read sequencing on the PacBio RS II Single Molecule Real-Time (SMRT) platform. The assembled genome consisted of one circular chromosome, four circular plasmids, with a total size of 2,745,725 bp and a G + C content of 45.52 %. Annotation identified 2,660 coding sequences, 5 rRNAs, 66 tRNAs, and a single CRISPR locus. Average nucleotide identity (ANI) between *L. brevis* PYN10_6_2 and *L. brevis* DSM 20054^T yielded a value of 98.94 %. Further in-depth analysis revealed 182 antibiotic resistance genes, 237 putative virulence genes, 2 prophages, and 10 genomic islands. Additionally, functional annotation through COG and KEGG

* Corresponding author.
E-mail address: wy@ags.ac.cn (Y. Wang).

https://doi.org/10.1016/j.dib.2024.110911
2352-3409/© 2024 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/)

Dataset link: Levilactobacillus brevis strain PYN10-6-2 plasmid p1, complete sequence (Original data)
Dataset link: SRX24542287: Bridge PCR of Levilactobacillus brevis PYN10-6-2:Tenebrio Molitor feces (Original data)
Dataset link: SRX24542286: Bridge PCR of Levilactobacillus brevis PYN10_6_2:Tenebrio Molitor feces (Original data)

Keywords:
Levilactobacillus brevis
Genome sequencing
Zearalenone
Hydroxysteroid dehydrogenase
Biosafety

databases revealed the presence of three genes encoding 3α- and 3β-hydroxysteroid dehydrogenase (3α-/3β-HSD) within the bacterial chromosome. This comprehensive genomic characterization provides valuable insights into the genetic basis of *L. brevis* PYN10_6_2′s ZEN-reducing ability and its biosafety profile.

© 2024 The Author(s). Published by Elsevier Inc.
This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/)

Specifications Table

Subject	Microbiology: Applied Microbiology.
Specific subject area	Molecular microbiology and Genomics
Type of data	Table, Figures
Data collection	Genomic DNA was extracted using the Wizard® Genomic DNA Purification Kit (Promega). Short-read sequencing libraries were prepared with the NEXTflex™ Rapid DNA-Seq Kit and sequenced on an Illumina HiSeq X Ten platform (2 × 150 bp paired-end reads). Long-read sequencing was performed on a PacBio RS II Single Molecule Real Time (SMRT) platform. Using Unicycler v0.4.8 for long-read assembly, followed by sequence polishing with Pilon v1.22. Genome annotation was performed with the Prokaryotic Genome Annotation Pipeline (PGAP) and functionally characterized using Clusters of Orthologous Groups (COG) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. Antibiotic resistance genes were identified using the Comprehensive Antibiotic Resistance Database (CARD) v3.2.6. Virulence factors were predicted based on homology searches against the Virulence Factors Database (VFDB). Genomic islands were identified using IslandViewer v1.2 and IslandPath-DIMOB v1.0.0, while prophages were predicted using Phigaro v2.3.0.
Data source location	Institution: Academy of National Food and Strategic Reserves Administration City/Town/Region: Beijing Country: China
Data accessibility	Repository name: NCBI (National Center for Biotechnology Information) GenBank Nucleotide database Data identification number: BioProject accession number: PRJNA1111238, NCBI SRA Accession Number: SRR29015671 and SRR29015672, GenBank Accession Number: CP156920-CP156924. Direct URL to data: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1111238 https://www.ncbi.nlm.nih.gov/sra/SRR29015671 https://www.ncbi.nlm.nih.gov/sra/SRR29015672 https://www.ncbi.nlm.nih.gov/nuccore/CP156920:CP156924[accn]
Related research article	None

1. Value of the Data

- The genomic data reveals the potential genes responsible for converting ZEN to α-/β-ZEL in *L. brevis* PYN10_6_2.
- The genomic data serves as a valuable resource for comparative genomic studies of *Levilactobacillus* spp. and related taxa, and can be used to assess the evolutionary relevance of short-promoted *Levilactobacillus* species.
- The genomic data can contribute to the evaluation of the biosafety of *L. brevis* PYN10_6_2 by identifying genetic elements associated with pathogenicity, resistance, and other risk factors.

Table 1
General genomic features of *L. brevis* PYN10_6_2.

Feature	Value
Total sequence length (bp)	2,745,725 bp
G+C content (%)	45.52 %
Plasmids	4
CDSs (total)	2,660
CDSs (with protein)	2,623
Genes assigned to COG	2,053
Genes assigned to KEGG	1,856
rRNA	15
tRNA	66
Number of CRISPR locus	1
Number of Prophages	2

2. Background

Zearalenone (ZEN), a prevalent mycotoxin produced by *Fusarium* species, is frequently found contaminating cereals [1]. Microbial degradation as a promising and environmentally friendly approach for ZEN mitigation [2]. One of the known mechanisms of ZEN degradation is the reduction of the ketocarbonyl groups, which reduces ZEN to α-/β-ZEL [3].

In a previous study, we isolated a strain, *L. brevis* 10_6_2, from *T. molitor* larval feces, demonstrating its ability to reduce ZEN to α-/β-ZEL [4]. Studies have shown that one of the metabolic pathways of zearalenone in rat liver is the reduction of the ketocarbonyl groups, which is thought to be catalyzed by 3α-/3β-HSD [5]. However, the genetic basis of this biotransformation process and the potential biosafety of the strain remain unexplored. The present study aims to address these knowledge gaps through a comprehensive genomic analysis of *L. brevis* PYN10_6_2. Specifically, we aim to: (1) identify potential genes responsible for ZEN reduction, elucidating the molecular mechanisms underlying this biotransformation process; (2) confirm the taxonomic classification of the strain through whole-genome sequencing; (3) evaluate its biosafety profile by screening for pathogenicity islands, virulence factors, and antibiotic resistance genes, a preliminary assessment for its potential application in food and feed safety.

3. Data Description

This study presents the whole genome sequencing information of *L. brevis* PYN10_6_2, which is able to convert ZEN to α-/β-ZEL. Genome sequencing was performed using the PacBio RS II and Illumina HiSeq X Ten sequencing platforms. *L. brevis* PYN10_6_2 generated 3,276,526 paired-end reads on the Illumina platform with a genome sequencing depth of 368.47×. The *L. brevis* PYN10_6_2 genome was sequenced to 104.05× depth using the PacBio platform, which generated 32,302 reads. The assembly coverage was calculated to be 99.95× based on the overall bases mapped to the genome. Genome assembly revealed that *L. brevis* PYN10_6_2 possesses a 2,745,725 bp, consisting of one circular chromosome and four circular plasmids, with a G + C content of 45.52 % (Fig. 1). Raw read length data can be found in the Sequence Read Archive (SRA), where the PacBio RS II and Illumina HiSeq X Ten reads are available under the accession numbers SRR29015671 and SRR29015672, respectively. The assembled genome can be found in Whole Gun Shotgun (WGS) under the accession numbers CP156920-CP156924. The four plasmids found in *L. brevis* PYN10_6_2 with the accession numbers CP156921, CP156922, CP156923, and CP156924. The genome was annotated using NCBI PGAP yielded a total of 2,660 CDSs including 2,623 CDSs containing coding proteins, 15 rRNAs, 66 tRNAs and 1 CRISPR (Table 1). The ANI between typical *Levilactobacillus* strains was obtained using OAT based on the complete genome sequence, and the ANI of *L. brevis* PYN10_6_2 with *L. brevis* DSM 20054^{T} was 98.94 % (Fig. 2).

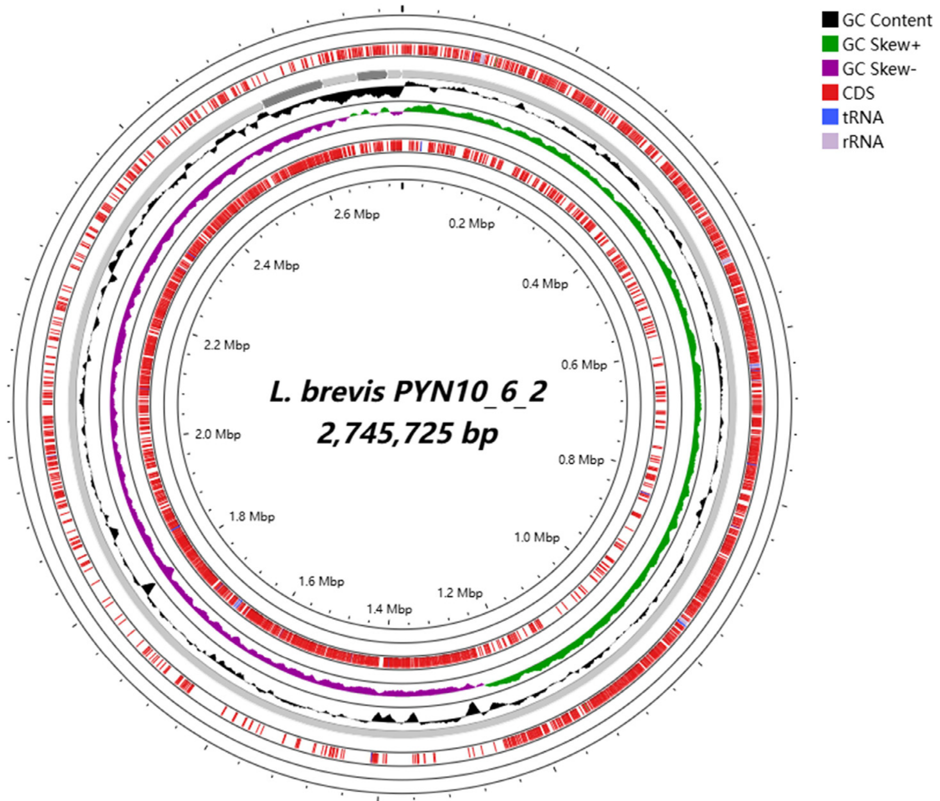

Fig. 1. Construction of a circular genome map of *L. brevis* PYN10_6_2 strain using CGView server (https://proksee.ca/ assessed on 26th May 2024). The first and fourth circles of the circle map from outside to inside are CDS, tRNA and rRNA on the positive and negative strands; the third and second circles are GC skew and GC content on the positive and negative strands, respectively; red color represents CDS; the green peaks represent GC-skew+; purple color represents GC-skew-; and the black peaks are G + C content. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

When combined with the annotation results, Functional annotation of CDS by gene evolutionary genealogy, COG annotation results showed that a total of 2,053 gene annotations were obtained, accounting for 76.52 % of the total number of genes. In contrast, KEGG annotation results showed 1,856 genes assigned, representing 69.18 % of the total number of genes (Table 1). The pathway of Steroid hormone biosynthesis in the Lipid metabolism hierarchy classification of KEGG and the Lipid transport and metabolism type of COG are known to have three coding sequences capable of producing 3α-/3β-HSD (ABE785_00920, ABE785_01750, ABE785_01835) present in the bacterial chromosome (Fig. 3). This genome annotation data is valuable for elucidating the molecular mechanism of the biotransformation process of *L. brevis* PYN10_6_2 to ZEN to α-/β-ZEL.

Information on resistance genes contained in each genome was obtained by CARD comparison prediction, and a total of 182 resistance genes appeared on chromosomes containing resistance genes to tetracyclines, fluoroquinolones, macrolides, and other antibiotics. An annotation map of the virulence factor genes was obtained by comparison with the VFDB, and a total of 237 virulence factors were statistically obtained, of which 8.43 % were present in the plasmids, 17 virulence factors were present in plasmid CP156921, and 3 virulence factors were present in plasmids CP156923 and CP156924, respectively. Ten genomic islands were obtained by ge-

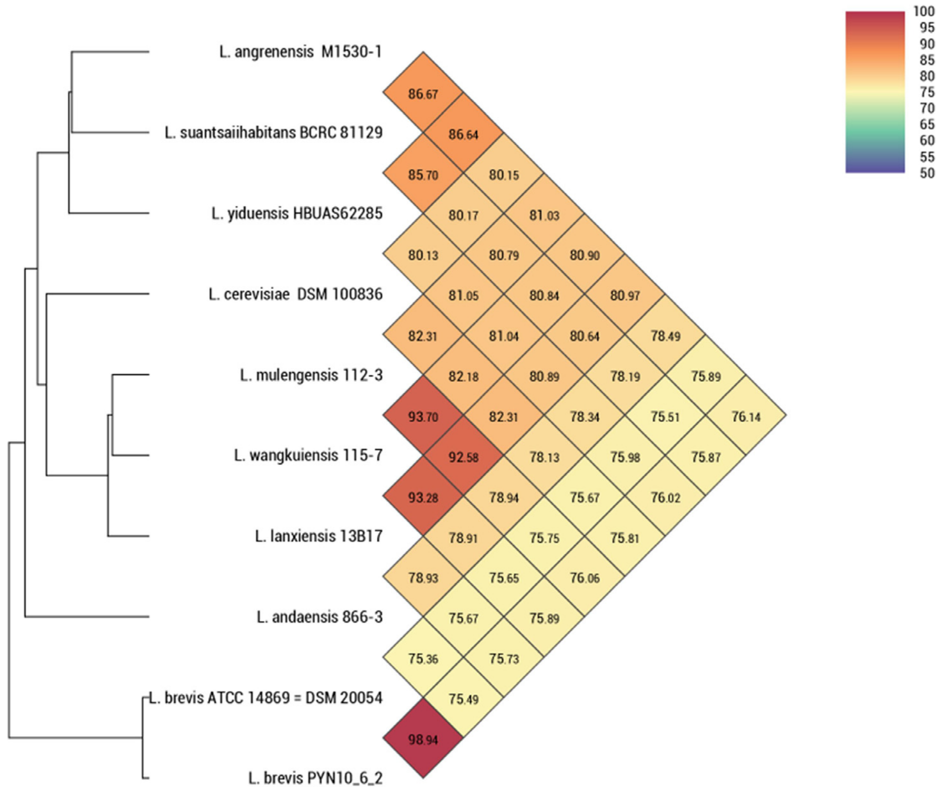

Fig. 2. Heatmap generated by OAT software showing ANI values for *L. brevis* PYN10_6_2 and related *Levilactobacillus* reference strains.

nomic island prediction using IslandViewer and IslandPath-DIMOB, four of which were present on plasmids. Two prophages present on the chromosome were obtained by Phigaro's prediction. This genome annotation data is valuable for the preliminary assessment of the biosafety of *L. brevis* PYN10_6_2 in food and feed applications.

4. Experimental Design, Materials and Methods

4.1. Sample collection

L. brevis PYN10_6_2, previously isolated from *T. molitor* larval feces and shown to convert ZEN into α-/β-ZEL [4], was subjected to whole-genome sequencing. The microbial strain is inoculated into fresh MRS broth and incubated at 37 °C with a shaking speed of 220 revolutions per minute (rpm) for a duration of 9 h to reach the exponential growth phase and collection of bacterial cells. Bacterial cells were harvested and submitted to Shanghai Majorbio Bio-Pharm Technology Co., Ltd. for whole-genome sequencing.

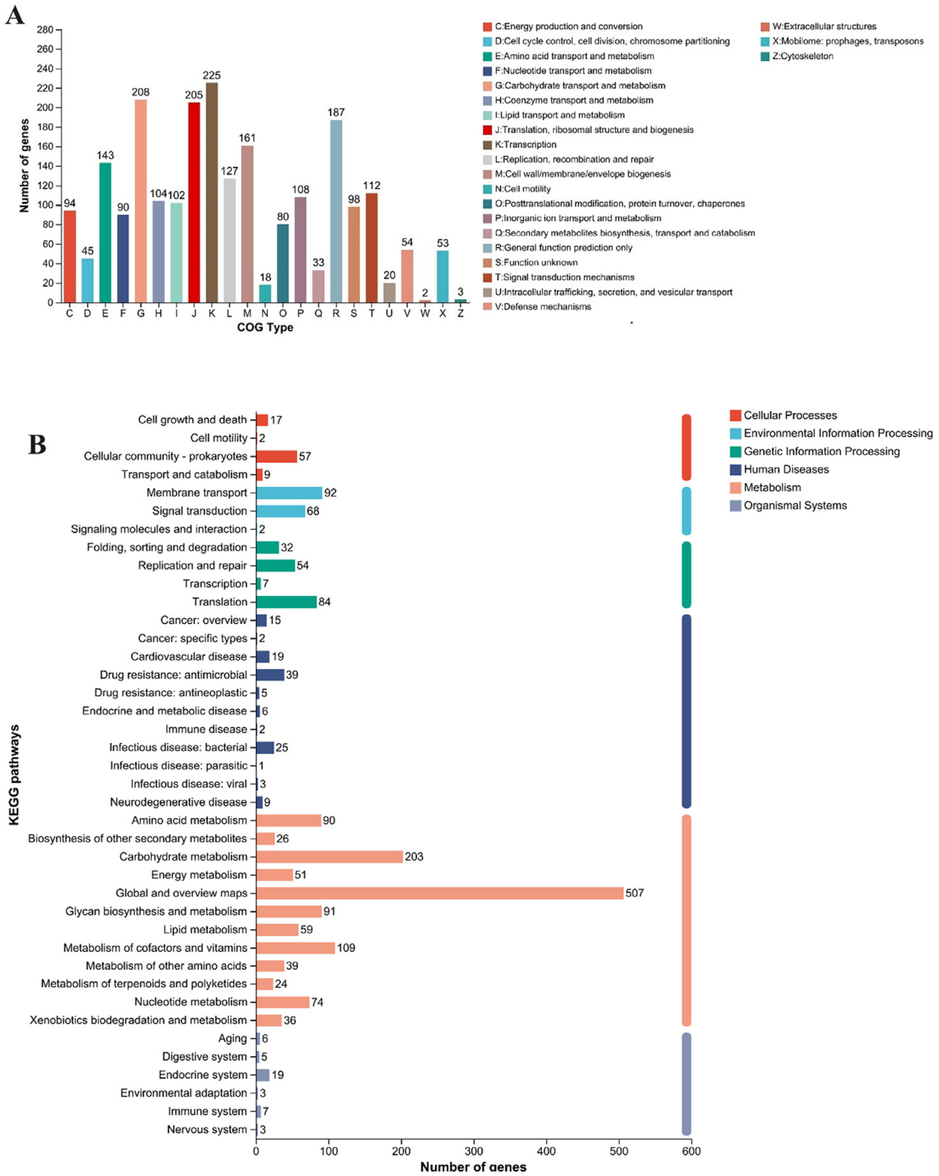

Fig. 3. COG and KEGG annotations for the coding sequence of *L. brevis* PYN10_6_2. A, COG annotation for the coding sequence; B, KEGG annotation for the sequence.

4.2. Genomic DNA preparation

Genomic DNA was extracted using Wizard® Genomic DNA Purification Kit (Promega) according to manufacturer's protocol. The concentration and purity of genomic DNA were assessed by a TBS-380 fluorometer (Turner BioSystems Inc., Sunnyvale, CA). High quality DNA (OD260/280=1.8–2.0, >20 μg) was used for subsequent analyses.

4.3. Whole genome sequencing and assembly

The genome of *L. brevis* PYN10_6_2 was sequenced using a hybrid approach combining short-read Illumina and long-read PacBio SMRT sequencing technologies. For Illumina sequencing, genomic DNA was fragmented to 400–500 bp using a Covaris M220 Focused ultrasonicator and libraries were prepared with the NEXTflex^{TM} Rapid DNA-Seq Kit. Sequencing was performed on an Illumina HiSeq X Ten platform (2 $\times$ 150 bp paired-end reads). For PacBio sequencing, genomic DNA was sheared, purified, end-repaired, and ligated to SMRTbell adapters following manufacturer's recommendations. The resulting sequencing library was purified three times using 0.45$\times$ volumes of Agencourt AMPure XP beads (Beckman Coulter Genomics, MA) according to the manufacturer's instructions. A $\sim$10 kb insert library was prepared and sequenced on a SMRT cell following standard methods. Raw reads from both platforms underwent quality assessment and trimming. Genome assembly was conducted using Unicycler v0.4.8[6], and the resulting assembly was polished with Pilon v1.22[7]. Complete circular chromosomes and plasmids were identified based on overlapping contig ends.

4.4. Whole genome prediction and annotation

Genome annotation was performed using the PGAP v6.7. Functional annotation of CDS was performed through the EggNOG and KEGG databases using the Diamond v0.8.35 Sequence Comparison Tool with an E-value threshold of 10^{-5} [8–10]. The assembled genome was visualized as a circular map using the CGView web service (https://proksee.ca/) [11]. ANI of strain with closely related species was determined using the Homologous Average Nucleotide Identity Tool v0.93.1 (OAT), resulting in the construction of a phylogenetic tree and a heatmap [12].

Genomic island prediction was performed using IslandViewer 4 and IslandPath-DIMOB v1.0.0 [13]. Prophage prediction was performed using Phigaro v2.3.0 [14]. The virulence factors were obtained from the VFDB using the Diamond with an E-value threshold of 10^{-5} [15]. Antibiotic resistance genes were obtained by Diamond with an E-value threshold of 10^{-5} via CARD v3.2.6 [16].

Limitations

None.

Ethics Statement

The authors have read and follow the ethical requirements for publication in Data in Brief and confirming that the current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

CRediT Author Statement

Mengru Zhao: Conceptualization, Data curation, Formal analysis, Methodology, Validation, Writing – original draft; **Baoyuan Guo:** Project administration, Funding acquisition; **Shuzhen Jiang:** Supervision; **Yang Wang:** Conceptualization, Investigation, Supervision, Writing – review and editing, Funding acquisition.

Data Availability

Levilactobacillus brevis strain PYN10-6-2 plasmid pLB1106_4, complete sequence (Original data) (NCBI GenBank Nucleotide database).

Levilactobacillus brevis strain PYN10-6-2 plasmid pUCCLB521_C, complete sequence (Original data) (NCBI GenBank Nucleotide database).

Levilactobacillus brevis strain PYN10-6-2 plasmid pLB1174-1, complete sequence (Original data) (NCBI GenBank Nucleotide database).

Levilactobacillus brevis strain:PYN10_6_2 Raw sequence reads (Original data) (NCBI GenBank Nucleotide database).

Levilactobacillus brevis strain PYN10-6-2 chromosome, complete genome (Original data) (NCBI GenBank Nucleotide database).

Levilactobacillus brevis strain PYN10-6-2 plasmid p1, complete sequence (Original data) (NCBI GenBank Nucleotide database).

SRX24542287: Bridge PCR of Levilactobacillus brevis PYN10-6-2:Tenebrio Molitor feces (Original data) (NCBI GenBank Nucleotide database).

SRX24542286: Bridge PCR of Levilactobacillus brevis PYN10_6_2:Tenebrio Molitor feces (Original data) (NCBI GenBank Nucleotide database).

Acknowledgments

This research was supported by grants from the Key Research and Development Program of Ningxia, China (2022BBF03031) and the National Natural Science Foundation of China (31972605).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] L. Kong, A.-H. Zhao, Q.-W. Wang, Y.-Q. Feng, Z.-H. Yan, M.-H. Li, F.-L. Zhang, H. Wang, K.-Y. Shen, Y. Liu, Y.-J. Sun, W. Shen, L. Li, Maternal Zearalenone exposure impacted ovarian follicle formation and development of suckled offspring, Sci. Total Environ. 788 (2021) 147792, doi:10.1016/j.scitotenv.2021.147792.

[2] B. Murtaza, B. Jin, L. Wang, X. Li, M.K. Saleemi, S. Majeed, A. Khatoon, G. Li, Y. Xu, Mitigation of zearalenone in vitro using probiotic strains, LWT–Food Sci. Technol. 186 (2023) 115265, doi:10.1016/j.lwt.2023.115265.

[3] P. Cai, S. Liu, Y. Tu, T. Shan, Toxicity, biodegradation, and nutritional intervention mechanism of zearalenone, Sci. Total Environ. 911 (2024) 168648, doi:10.1016/j.scitotenv.2023.168648.

[4] P. Yuetong, Degradation and Metabolic Behavior of Zearalenone in Tenebrio Molitor [D], Beijing: Beijing Union University, 2022.

[5] M. Olsen, H. Pettersson, K.H. Kiessling, Reduction of zearalenone to zearalenol in female rat liver by 3α-hydroxysteroid dehydrogenase, Acta Pharmacol. Toxicol. 48 (1981) 157–161, doi:10.1111/j.1600-0773.1981.tb01602.x.

[6] R.R. Wick, L.M. Judd, C.L. Gorrie, K.E. Holt, Unicycler: resolving bacterial genome assemblies from short and long sequencing reads, PLoS Comput. Biol. 13 (2017) e1005595, doi:10.1371/journal.pcbi.1005595.

[7] B.J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C.A. Cuomo, Q. Zeng, J. Wortman, S.K. Young, A.M. Earl, Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement, PLoS ONE 9 (2014) e112963, doi:10.1371/journal.pone.0112963.

[8] L.J. Jensen, P. Julien, M. Kuhn, C. von Mering, J. Muller, T. Doerks, P. Bork, eggNOG: automated construction and annotation of orthologous groups of genes, Nucleic Acids Res. 36 (2008) D250–D254, doi:10.1093/nar/gkm796.

[9] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 28 (2000) 27–30, doi:10.1093/nar/28.1.27.

[10] B. Buchfink, C. Xie, D.H. Huson, Fast and sensitive protein alignment using DIAMOND, Nat. Methods 12 (2015) 59–60, doi:10.1038/nmeth.3176.

[11] J.R. Grant, E. Enns, E. Marinier, A. Mandal, E.K. Herman, C.Y. Chen, M. Graham, G. Van Domselaar, P. Stothard, Proksee: in-depth characterization and visualization of bacterial genomes, Nucleic Acids Res. 51 (2023) W484-w492, doi:10.1093/nar/gkad326.

[12] I. Lee, Y. Ouk Kim, S.-C. Park, J. Chun, OrthoANI: an improved algorithm and software for calculating average nucleotide identity, Int. J. Syst. Evol. Microbiol. 66 (2016) 1100–1103, doi:10.1099/ijsem.0.000760.

[13] C. Bertelli, M.R. Laird, K.P. Williams, B.Y. Lau, G. Hoad, G.L. Winsor, F.S.L. Brinkman, IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets, Nucleic Acids Res. 45 (2017) W30-w35, doi:10.1093/nar/gkx343.

[14] D.E. Fouts, Phage_Finder: automated identification and classification of prophage regions in complete bacterial genome sequences, Nucleic Acids Res. 34 (2006) 5839–5851, doi:10.1093/nar/gkl732.

[15] L. Chen, D. Zheng, B. Liu, J. Yang, Q. Jin, VFDB 2016: hierarchical and refined dataset for big data analysis–10 years on, Nucleic Acids Res. 44 (2016) D694–D697, doi:10.1093/nar/gkv1239.

[16] B. Jia, A.R. Raphenya, B. Alcock, N. Waglechner, P. Guo, K.K. Tsang, B.A. Lago, B.M. Dave, S. Pereira, A.N. Sharma, S. Doshi, M. Courtot, R. Lo, L.E. Williams, J.G. Frye, T. Elsayegh, D. Sardar, E.L. Westman, A.C. Pawlowski, T.A. Johnson, F.S. Brinkman, G.D. Wright, A.G. McArthur, CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database, Nucleic Acids Res. 45 (2017) D566-d573, doi:10.1093/nar/gkw1004.