

DATA NOTE

Open Access



# High-quality genome resource of *Lasiodiplodia pseudotheobromae* associated with die-back on *Eucalyptus* trees

LinQin Lu<sup>1</sup>, GuoQing Li<sup>1</sup> and FeiFei Liu<sup>1\*</sup>

## Abstract

**Objectives** *Lasiodiplodia pseudotheobromae* is an important fungal pathogen associated with die-back, canker and shoot blight in many plant hosts with a wide geographic distribution. The aim of our study was to provide high-quality genome assemblies and sequence annotation resources of *L. pseudotheobromae*, to facilitate future studies on the systematics, population genetics and genomics of the fungal pathogen *L. pseudotheobromae*.

**Data description** High-quality genomes of five *L. pseudotheobromae* isolates were sequenced based on Oxford Nanopore technology (ONT) and Illumina HiSeq sequencing platform. The total size of each assembly ranged from 43 Mb to 43.86 Mb and over 11,000 protein-coding genes were predicted from each genome. The proteins of predicted genes were annotated using multiple public databases, among the annotated protein-coding genes, more than 4,300 genes were predicted as potential virulence genes by the Pathogen Host Interactions (PHI) database. Moreover, the genome comparative analysis among *L. pseudotheobromae* and other closely related species revealed that 7,408 gene clusters were shared among them and 152 gene clusters unique to *L. pseudotheobromae*. This genome and associated datasets provided here will serve as a useful resource for further analyses of this fungal pathogen species.

**Keywords** Reference genome, Comparative genome, Function annotation, Fungal disease, Canker

## Objective

Members of *Botryosphaeriaceae* are considered as latent pathogens and can infect numerous hosts almost all woody plants [1]. Diseases associated with them usually occur under environmental stresses such as drought, frost and heat, and typical symptoms include canker, dieback, root rot, fruit rot and twig blight [1, 2]. *Lasiodiplodia pseudotheobromae* (*Botryosphaeriaceae*,

*Botryosphaeriales*) was first described in 2008, which is closely related to *L. theobromae* [3]. The known hosts include nearly 100 species in 40 families, such as forest trees of *Eucalyptus* spp., *Acacia* spp., *Pinus* spp., crop plants of *Gossypium hirsutum*, *Citrus* spp., and ornamental plants of *Bougainvillea spectabilis*, *Magnolia candolei* [4]. The geographic distribution of this pathogen recorded includes China [5–7], Malaysia [8], Brazil [9], Venezuela [10], South Africa [11], Tunisia [12] and Spain [13].

In southern China, studies on *Botryosphaeriaceae* showed that *L. pseudotheobromae* is one of the dominant causal agents of *Eucalyptus* die-back, canker and shoot blight in plantations, especially in [6, 14]. Inoculation

\*Correspondence:

FeiFei Liu  
shelleyfei@163.com

<sup>1</sup>Research Institute of Fast-growing Trees (RIFT), Chinese Academy of Forestry (CAF), Zhanjiang 524022, China



© 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.

trials in the greenhouse and field suggested that this pathogen has a relatively high virulence to different *Eucalyptus* species or hybrids, compared to other species in *Botryosphaeria* and *Neofusicoccum* [7]. For this important pathogen, there are three isolates with publicly available genomic data in the NCBI database, CBS 116459 from *Gemlina arborea* [15], KET9 from *Prunus persica* [16] and BaA from *Morinda officinalis* [17] (DataFile 1; Table 1) [18]. These genome assemblies are fragmented and not suitable as reference genomes. Thus, high-quality genome assemblies based on long-read sequencing technology by Oxford Nanopore Technologies (ONT) were conducted in this study. These new genomic resources can provide more information for future studies aimed at fungal biology and pathogenic mechanism of *L. pseudotheobromae*.

### Data description

Five *L. pseudotheobromae* isolates originated from plantation trees of *Eucalyptus* spp. and *Cunninghamia lanceolata* in southern China were selected for genome sequencing in this study (DataFile 1; Table 1) [18]. Fresh mycelia of the single hyphal tip isolates were harvested from 2% MEA plates (20 g malt extract powder and 20 g agar per litre of water) covered with cellophane for 2 days at 25 °C and immediately frozen in liquid nitrogen, followed by preservation at -80 °C in the laboratory prior to DNA extraction. High-quality genomic DNA was extracted using a modified CTAB (cetyltrimethylammonium bromide) method [26]. The integrity and purity of DNA were detected by 0.8% agarose gel electrophoresis and the precise concentration of which was quantified by a Qubit 2.0 fluorescence detector (Life Technologies). All five isolates were confirmed as *L. pseudotheobromae* by sequencing the elongation factor 1- $\alpha$  (*EF1- $\alpha$* ) gene and phylogenetic analyses.

Whole genome sequencing was conducted using both the short-read platform and the long-read Oxford Nanopore Technologies (ONT) in Zhenyue Biotechnology Co., Ltd (WuHan, China). The Illumina sequencing was performed for all the five isolates (RIFT3495, RIFT 6050, RIFT 15092, RIFT 18431 and RIFT 19273). Paired-end library with 350 bp median insert size was generated and 150 bp paired-end reads were sequenced using the Illumina HiSeq 2500 platform. Poor-quality data and adapters were removed using the program Trimmomatic v. 0.36 [27]. The program SPAdes v. 3.14 [28] was used to assemble the genome *de novo* into contigs. The ONT sequencing was performed for the two isolates RIFT 3495 and RIFT 18431. The library was loaded on a MinION R10.3 flow cell (FLO-MIN111) and the sequencing run was carried out for 48 h. Base calling was conducted using the ONT Guppy base calling software v. 4.0.14 (<https://community.nanoporetech.com>). GenomeScope was used to estimate the size of genomes [29]. The ONT reads were assembled with the program Mecat2 (20,190,226) with default parameters after filtration of the low-quality reads [30]. The assembled genome was then polished with ONT reads and Illumina reads by using Racon v. 1.4.11 [31] and Pilon v. 1.23 [32], respectively.

Genome size of the five strains were generated by GenomeScope, ranging from 42 to 44.61 Mb, and the heterozygosity was estimated to be 0.01 to 0.24%. An average of 2,081,811 ONT reads (up to 332  $\times$  coverage) and 49,479,273 Illumina clean reads (up to 192  $\times$  coverage) were generated in this study (DataFile 1; Table 1) [18]. The assembled draft genomes were about 43 Mb in size and with the highest  $N_{50}$  value (5,817,267 bp) and the minimum contig numbers (8 contigs) among all the published *L. pseudotheobromae* genomes (DataFile 1; Table 1) [15–17, 33]. For each of the five genomes, a perfect spectra graph performed by KAT program [34] was acquired, clearly showing a complete haplotype

**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	Table S1, Genome assembly and annotation features of <i>Lasiodiplodia pseudotheobromae</i> isolates	Document file (.docx)	Figshare, <a href="https://doi.org/10.6084/m9.figshare.24419041.v3">https://doi.org/10.6084/m9.figshare.24419041.v3</a> [18]
Data file 2	Figure S1, Venn diagram	Image file (.jpg)	Figshare, <a href="https://doi.org/10.6084/m9.figshare.24418999.v1">https://doi.org/10.6084/m9.figshare.24418999.v1</a> [19]
Data file 3	Figure S2, Phylogenetics analyses	Image file (.jpg)	Figshare, <a href="https://doi.org/10.6084/m9.figshare.24419029.v1">https://doi.org/10.6084/m9.figshare.24419029.v1</a> [20]
Data set 1	Genome assembly of <i>Lasiodiplodia pseudotheobromae</i> strain RIFT 3495	Fasta file (.fasta)	NCBI Nucleotide, <a href="https://identifiers.org/nucleotide:JAWM-WM000000000">https://identifiers.org/nucleotide:JAWM-WM000000000</a> [21]
Data set 2	Genome assembly of <i>Lasiodiplodia pseudotheobromae</i> strain RIFT 6050	Fasta file (.fasta)	NCBI Nucleotide, <a href="https://identifiers.org/nucleotide:JAWM-WL000000000">https://identifiers.org/nucleotide:JAWM-WL000000000</a> [22]
Data set 3	Genome assembly of <i>Lasiodiplodia pseudotheobromae</i> strain RIFT 15092	Fasta file (.fasta)	NCBI Nucleotide, <a href="https://identifiers.org/nucleotide:JAWM-WK000000000">https://identifiers.org/nucleotide:JAWM-WK000000000</a> [23]
Data set 4	Genome assembly of <i>Lasiodiplodia pseudotheobromae</i> strain RIFT 18431	Fasta file (.fasta)	NCBI Nucleotide, <a href="https://identifiers.org/nucleotide:JAWM-WJ000000000">https://identifiers.org/nucleotide:JAWM-WJ000000000</a> [24]
Data set 5	Genome assembly of <i>Lasiodiplodia pseudotheobromae</i> strain RIFT 19273	Fasta file (.fasta)	NCBI Nucleotide, <a href="https://identifiers.org/nucleotide:JAWM-WI000000000">https://identifiers.org/nucleotide:JAWM-WI000000000</a> [25]

achieved. Benchmarking Universal Single-Copy Orthologs (BUSCO) based on *fungi\_odb 10* [35] was used to evaluate the completeness of the genome assemblies. The results showed a high completeness score of up to 99.2% of all the five assemblies in this study, which indicated that the continuity of these assemblies is comparable with the publicly available genomes but is essentially better than them (DataFile 1; Table 1) [15–17, 33].

Maker2 v. 2.31.9 [36] was used for *de novo* gene prediction. In total, up to 12,237 genes were predicted as protein-coding genes with an average length of 1,937.92 bp for all the five genomes in this study (DataFile 1; Table 1) [18]. In addition, about 245 noncoding RNAs (transfer RNA, ribosomal RNA and small nuclear RNA) were predicted using tRNAscan-SE v. 2.0 [37] and Barmap v. 0.8 (<https://github.com/tseemannbarmap>). Further, repeat family identification and modeling were performed *de novo* using Repeatmasker v. 4.0.7 [38]. An average of 59,444 bp of repeat sequences that accounted for about 0.14% of the assemblies were detected in the assembled genomes (DataFile 1; Table 1) [18].

## Annotation

Functional annotation of the predicted gene sequences was done using BLAST to search against multiple public databases, including the InterProScan database (ave. 8,453 genes, 73.76%), Gene Ontology (GO; ave. 1,858 genes, 16.21%), Kyoto Encyclopedia of Genes and Genomes (KEGG; ave. 10,868 genes, 94.82%), Swiss-Prot database (ave. 7,323 genes, 63.91%), TrEMBL database (ave. 11,410 genes, 99.62%) and NCBI's Nonredundant Protein (Nr; ave. 11,453 genes, 99.91%). Additional annotation was carried out based on the Pathogen Host Interactions (PHI) database [39], and Carbohydrate-Active Enzymes (CAZys) databases [40]. Meanwhile, secretory proteins were analyzed using Signal P v. 4.1 and TMHMM v. 2.0 [33]. A total of average 4,429 (PHI) genes were identified in the five genomes, and nearly 900 genes of each genome were annotated from the CAZys databases, including 405 genes related to glycoside hydrolases (GHs), 185 genes related to glycosyl transferases (GTs), 57 genes related to carbohydrate esterases (CEs), 28 genes related to polysaccharide lyases (PLs), 108 genes predicted to have auxiliary activities (AAs) and 87 genes associated with carbohydrate-binding modules (CBMs). Moreover, a total of average 835 putative secondary proteins were identified in the five genomes.

The comparative genomics of the orthologous gene cluster between *L. pseudotheobromae* RIFT 3495 and three related species (*Lasiodiplodia theobromae*, *Botryosphaeria dothidea*, *Neofusicoccum parvum*) were analyzed using the CD-HIT v. 4.6.1 rapid clustering of similar proteins software with a threshold of 50% pairwise identity and 0.7 length difference cutoff in amino acids, which

revealed 7,408 common gene clusters and 152 gene clusters unique to RIFT 3495. RIFT 3495 shared 786, 93 and 13 gene clusters with *L. theobromae*, *B. dothidea* and *N. parvum*, respectively (DataFile 2; Table 1) [19]. Software RAxML was used to construct the evolutionary tree by the maximum likelihood method [41], phylogenetic analysis of single copy orthologous genes from twelve genomes along with *Aplosporella prunicola* (as outgroup) showed a similar association of *L. pseudotheobromae* with *L. theobromae*, followed by *Diplodia corticola* and *D. seriata* (DataFile 3; Table 1) [20].

This study presents five draft genome sequence resources of *L. pseudotheobromae*, a fungal pathogen causing trunk disease in southern China, which is of great importance for elucidating the biology and pathogenicity of this fungus on woody perennial trees.

## Limitation

The *de novo* assemblies resulted in a number of contigs, the genomic quality of the three *L. pseudotheobromae* isolates which sequenced only based on the Illumina HiSeq platform were still fragmented and not suitable for genome structure analysis. Further high-quality genome assemblies using long-read sequencing technologies for those isolates are still needed.

## Acknowledgements

Not applicable.

## Author contributions

GuoQing Li and Feifei Liu conceived the experiments; LinQin Lu completed experiments and wrote the manuscript. All authors edited and approved the final manuscript.

## Funding

This study was supported by the Natural Science Foundation of Guangdong Province, China (Grant No. 2022A1515010874).

## Data availability

The data described in this Data note were deposited under NCBI BioProject ID PRJNA1030934 [19–23]. Associated Datafiles are available on Figshare: Table S1, Genome assembly and annotation features of *Lasiodiplodia pseudotheobromae* isolates [18], Figure S1, Venn diagram [19], Figure S2, Phylogenetics analyses [20]. Please see Table 1 for details and links to the data.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent to publish

Not applicable.

### Competing interests

The authors declare no competing interests.

Received: 24 October 2023 / Accepted: 15 December 2023

Published online: 02 January 2024

## References

- Slippers B, Wingfield MJ. *Botryosphaeriaceae* as endophytes and latent pathogens of woody plants: diversity, ecology and impact. *Fungal Biol Rev*. 2007;21:90–106. <https://doi.org/10.1016/j.fbr.2007.06.002>.
- Slippers B, Crous PW, Jami F, Groenewald JZ, Wingfield MJ. Diversity in the *Botryosphaeriales*: looking back, looking forward. *Fungal Biol*. 2017;121:307–21. <https://doi.org/10.1016/j.funbio.2017.02.002>.
- Alves A, Crous PW, Correia A, Phillips AJL. Morphological and molecular data reveal cryptic speciation in *Lasiodiplodia theobromae*. *Fungal Divers*. 2008;28:1–13.
- EFSA Panel on Plant Health (PLH), Bragard C, Baptista P, Chatzivassiliou E, di Serio F, Gonthier P, Reignault PL. Pest categorisation of *Lasiodiplodia pseudotheobromae*. *EFSA J*. 2023;21:e07737. <https://doi.org/10.2903/j.efsa.2023>.
- Zhao JP, Lu Q, Liang J, Decock C, Zhang XY. *Lasiodiplodia pseudotheobromae*, a new record of pathogenic fungus from some subtropical and tropical trees in southern China. *Cryptogamie Mycol*. 2010;31:431.
- Li GQ, Liu FF, Li JQ, Liu QL, Chen SF. *Botryosphaeriaceae* from *Eucalyptus* plantations and adjacent plants in China. *Persoonia*. 2018;40:63–95. <https://doi.org/10.3767/persoonia.2018.40.03>.
- Li GQ, Slippers B, Wingfield MJ, Chen SF. Variation in *Botryosphaeriaceae* from *Eucalyptus* plantations in YunNan Province in southwestern China across a climatic gradient. *IMA Fungus*. 2020;11:1–49.
- Munirah MS, Azmi AR, Yong SYC, Nur Ain Izzati MZ. Characterization of *Lasiodiplodia theobromae* and *L. Pseudotheobromae* causing fruit rot on pre-harvest mango in Malaysia. *Plant Pathol Quar*. 2017;7:202–13.
- Júnior AFN, Santos RF, Pagenotto ACV, Spósito MB. First report of *Lasiodiplodia pseudotheobromae* causing fruit rot of persimmon in Brazil. *New Dis Rep*. 2017;36:1.
- Castro-Medina F, Mohali SR, Urbez-Torres JR, Gubler WD. First report of *Lasiodiplodia pseudotheobromae* causing trunk cankers in *Acacia mangium* in Venezuela. *Plant Dis*. 2014;98:686. <https://doi.org/10.1094/PDIS-02-13-0160-PDN>.
- Cruywagen EM, Slippers B, Roux J, Wingfield MJ. Phylogenetic species recognition and hybridisation in *Lasiodiplodia*: a case study on species from baobabs. *Fungal Biol*. 2017;121:420–36. <https://doi.org/10.1016/j.funbio.2016.07.01>.
- Rezgui A, Vallance J, Ben Ghnaya-Chakroun A, Bruez E, Dridi M, Demasse RD, Rey P, Sadfi-Zouaoui N. Study of *Lasiodiplodia pseudotheobromae*, *Neofusicoccum parvum* and *Schizophyllum commune*, three pathogenic fungi associated with the Grapevine Trunk Diseases in the North of Tunisia. *Eur J Plant Pathol*. 2018;152:127–42.
- López-Moral A, del Carmen Raya M, Ruiz-Blancas C, Medialdea I, Lovera M, Arquero O, Agustí-Brisach C. Aetiology of branch dieback, panicle and shoot blight of pistachio associated with fungal trunk pathogens in southern Spain. *Plant Pathol*. 2020;69:1237–69. <https://doi.org/10.1111/ppa.13209>.
- Li GQ, Arnold RJ, Liu FF, Li JQ, Chen SF. Identification and pathogenicity of *Lasiodiplodia* species from *Eucalyptus urophylla* × *grandis*, *Polyscias balfouriana* and *Bougainvillea spectabilis* in southern China. *J Phytopathol*. 2015;163:956–67. <https://doi.org/10.1111/ppa.13209>.
- Nagel JH, Cruywagen EM, Machua J, Wingfield MJ, Slippers B. Highly transferable microsatellite markers for the genera *Lasiodiplodia* and *Neofusicoccum*. *Fungal Ecol*. 2020;44:100903. <https://doi.org/10.1016/j.funeco.2020.100903>.
- Yu CM, Diao YF, Lu Q, Zhao JP, Cui SN, Xiong X, Lu A, Zhang XY, Liu HX. Comparative genomics reveals evolutionary traits, mating strategies, and pathogenicity-related genes variation of *Botryosphaeriaceae*. *Front Microbiol*. 2022;13:800981. <https://doi.org/10.3389/fmicb.2022.800981>.
- Li XY, Luo M, Song HD, Dong ZY. Whole-genome resource of *Lasiodiplodia pseudotheobromae* BaA, the causative agent of black root rot *Morinda Officinalis*. *Plant Dis*. 2023;107:542–5. <https://doi.org/10.1094/PDIS-06-22-1507-A>.
- Lu LQ, Li GQ, Liu FF. Data file 1-Table S1, genome assembly and annotation features of *Lasiodiplodia pseudotheobromae* isolates. Figshare. 2023. <https://doi.org/10.6084/m9.figshare.24419041.v3>.
- Lu LQ, Li GQ, Liu FF. Data file 2-Figure S1, Venn diagram. Figshare. 2023. <https://doi.org/10.6084/m9.figshare.24418999.v1>.
- Lu LQ, Li GQ, Liu FF. Data file 3- figure S2, Phylogenetics analyses. Figshare. 2023. <https://doi.org/10.6084/m9.figshare.24419029.v1>.
- Lu LQ, Li GQ, Liu FF. Dataset 1- Genome assembly of *Lasiodiplodia pseudotheobromae* strain RIFT 3495. NCBI. 2023. <https://identifiers.org/nucleotide:JAWMMW000000000>.
- Lu LQ, Li GQ, Liu FF. Dataset 2- Genome assembly of *Lasiodiplodia pseudotheobromae* strain RIFT 6050. NCBI. 2023. <https://identifiers.org/nucleotide:JAWMML000000000>.
- Lu LQ, Li GQ, Liu FF. Dataset 3- Genome assembly of *Lasiodiplodia pseudotheobromae* strain RIFT 15072. NCBI. 2023. <https://identifiers.org/nucleotide:JAWMMW000000000>.
- Lu LQ, Li GQ, Liu FF. Dataset 4- Genome assembly of *Lasiodiplodia pseudotheobromae* strain RIFT 18431. NCBI. 2023. <https://identifiers.org/nucleotide:JAWMMW000000000>.
- Lu LQ, Li GQ, Liu FF. Dataset 5- Genome assembly of *Lasiodiplodia pseudotheobromae* strain RIFT 19273. NCBI. 2023. <https://identifiers.org/nucleotide:JAWMMW000000000>.
- Möller EM, Bahnweg G, Sandermann H, Geiger HH. A simple and efficient protocol for isolation of high molecular weight DNA from filamentous fungi, fruit bodies, and infected plant tissues. *Nucleic Acids Res*. 1992;20:6115. <https://doi.org/10.1093/nar/20.22.6115>.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20. <https://doi.org/10.1093/bioinformatics/btu170>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Pevzner PA. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19:455–77. <https://doi.org/10.1089/cmb.2012.0021>.
- Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics*. 2017;33:2202–4.
- Xiao CL, Chen Y, Xie SQ, Chen KN, Wang Y, Han Y, Xie Z. MECAT: fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. *Nat Methods*. 2017;14:1072–4.
- Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res*. 2017;27:737–46.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Earl AM. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE*. 2014;9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.
- Petersen TN, Brunak S, Von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*. 2011;8:785–6.
- Mapleson D, Garcia Accinelli G, Kettleborough G, Wright J, Clavijo BJ. KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics*. 2017;33(4):574–6.
- Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31:32103212. <https://doi.org/10.1093/bioinformatics/btv351>.
- Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*. 2011;12:1–14.
- Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. 1997;25:955–64. <https://doi.org/10.1093/nar/25.5.955>.
- Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics*. 2009;25:4.10.1–4.10.14.
- Urban M, Pant R, Raghunath A, Irvine AG, Pedro H, Hammond-Kosack KE. The Pathogen-host interactions database (PHI-base): additions and future developments. *Nucleic Acids Res*. 2015;43:D645–55. <https://doi.org/10.1093/nar/gku1165>.
- Jia F, Zhang L, Pang X, Gu X, Abdelazez A, Liang Y, Meng X. Complete genome sequence of bacteriocin-producing *Lactobacillus plantarum* KLD5.0391, a probiotic strain with gastrointestinal tract resistance and adhesion to the intestinal epithelial cells. *Genomics*. 2017;109:432–7. <https://doi.org/10.1016/j.ygeno.2017.109.432>.
- Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019;20:1–14. <https://doi.org/10.1186/s13059-019-1832-y>.

## Publisher's Note

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