



Data Article

Whole genome sequence data of *Comamonas sediminis* FS4_11, a fumonisin B1-transforming bacterium, using hybrid nanopore-illumina sequencing

Yang Wang^a, Mengru Zhao^{a,b}, Zhe Wang^a, Xiaohong Luo^a,
Chengfei Wang^c, Baoyuan Guo^{a,*}

^a Institute of Grain and Oil Quality and Safety, Academy of National Food and Strategic Reserves Administration, Beijing 100037, China

^b School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

^c Jiangsu Aomai Bio-Technology Co., Ltd., Nanjing 211226, China

ARTICLE INFO

Article history:

Received 21 May 2025

Revised 16 June 2025

Accepted 19 June 2025

Available online 24 June 2025

Dataset link: [BioProject \(Original data\)](#)

Dataset link: [SRA data_ONT \(Original data\)](#)

Dataset link: [SRA data_Illumina \(Original data\)](#)

Dataset link: [genome \(Original data\)](#)

Dataset link: [BioSample \(Original data\)](#)

Keywords:

Comamonas sediminis

Mycotoxin

Hybrid sequencing

Genome assembly

Functional annotation

Biosafety

ABSTRACT

The genome of *Comamonas sediminis* FS4_11, a bacterial strain with mycotoxin fumonisin B1 (FB1) transformation capability, was sequenced using Oxford Nanopore Technologies (ONT) and Illumina platforms. The final assembly generated a circular chromosome of 5,148,490 bp with a mean G+C content of 63.74%, representing a contiguous genomic structure. Genome annotation predicted 4565 protein-coding sequences (CDSs), 82 transfer RNAs (tRNAs), 18 ribosomal RNAs (rRNAs; 6 each of 5S, 16S, and 23S rRNA), 1 transfer-messenger RNA (tmRNA), and 8 pseudogenes and other non-coding RNAs. Functional annotation identified 939 potential virulence factors, two putative AdeF-related antibiotic resistance genes, 1486 potential pathogen-host interaction proteins, and a candidate carboxylesterase for FB1 transformation. This dataset primarily aids in identifying potential FB1 detoxification enzyme genes and assessing strain biosafety. It also offers sig-

* Corresponding author.

E-mail address: gby@ags.ac.cn (B. Guo).

nificant reuse potential for comparative genomics and understanding bacterial evolution.

© 2025 The Authors. Published by Elsevier Inc.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Specifications Table

Subject	Microbiology: Applied Microbiology
Specific subject area	Microbiology, Genomics, Biotechnology
Type of data	Tables, Figures
Data collection	High-quality genomic DNA was extracted using the Wizard® genomic DNA purification kit (Promega). Sequencing utilized Oxford Nanopore PromethION 48 for long reads and Illumina NovaSeq 6000 for short reads. Nanopore reads were filtered with Filtlong v0.2.0. Hybrid assembly combined Canu v1.5 (ONT draft), Circlator v1.5.5 (circularization), Racon v1.4.3 (polishing), and Pilon v1.22 (Illumina error correction). Gene prediction used Bakta (software v1.11 and database v6.0.0). Predicted proteins were further analyzed against specialized databases for antibiotic resistance (CARD v3.2.7), pathogen-host interactions (PHI-base v4.17), and virulence factors (VFDB 2022).
Data source location	Institution: Academy of National Food and Strategic Reserves Administration City: Beijing Country: China
Data accessibility	Repository name: National Centre for Biotechnology Information (NCBI) Data identification number: BioProject Accession Number: PRJNA1258737, BioSample: SAMN48330266, NCBI SRA Accession Number: SRR33422104 and SRR33422105, NCBI GenBank Accession Number: CP191350 https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1258737 https://www.ncbi.nlm.nih.gov/biosample/?term=SAMN48330266 https://www.ncbi.nlm.nih.gov/sra/SRX28663058 https://www.ncbi.nlm.nih.gov/sra/SRX28663057 https://www.ncbi.nlm.nih.gov/nucleotide/CP191350
Related research article	N/A

1. Value of the Data

- This genome of *C. sediminis* FS4_11 provides the essential foundation for identifying and characterizing genes potentially responsible for fumonisin B1 transformation. Researchers can use this sequence data as a reference to search for candidate genes based on homology to known FB1-degrading enzymes.
- The genome serves as a valuable reference for comparative genomics studies, contributing to our understanding of bacterial evolution and diversity within this species group.
- The comprehensive annotation of virulence factors, antibiotic resistance genes, and pathogenicity-related elements offers essential data for biosafety assessment, supporting researchers in evaluating the strain's safety for industrial applications.

2. Background

Fumonisin B1 (FB1) is a major mycotoxin that poses significant risks to food safety and human health [1]. While microbial degradation offers a promising approach for FB1 detoxification, the molecular mechanisms remain poorly understood [2]. *C. sediminis* FS4_11, isolated from a maize frass composite from *Sitophilus zeamais* infestation, exhibited FB1 biotransformation ca-

pability. Similar activity has been reported in the phylogenetically closely related strain NCB 1492 (*Delftia/Comamonas*-like), though the enzymes responsible remain unidentified in the *Comamonas* genus [3]. To investigate the genetic basis of this biotransformation and evaluate strain biosafety, we performed whole-genome sequencing using a hybrid approach combining Oxford Nanopore long-read and Illumina short-read technologies. The genome data provides a bioinformatic foundation for identifying potential FB1-degrading enzyme genes and enables biosafety assessment through detection of virulence factors, antibiotic resistance genes, and other safety-related genetic elements.

3. Data Description

The genomic statistics for strain FS4_11 are summarized in Table 1. The genome consists of a single contig representing a chromosome (5,148,490 bp; N50 5,148,490 bp) with a GC content of 63.74% and an average coverage depth of 399.5x. Annotation predicted 4,565 protein-coding sequences (CDS), 82 tRNA genes, 1 transfer-messenger RNA (tmRNA) gene, 8 pseudogenes, and 18 rRNA genes (6 copies each of 5S, 16S, and 23S). Additionally, 8 non-coding RNA (ncRNA) genes and 18 ncRNA regions were identified. No clustered regularly interspaced short palindromic repeats (CRISPR) was found. The overall genome structure and distribution of these features are visualized in Fig. 1.

Comparison of predicted protein sequences against the Virulence Factor Database (VFDB) via BLASTp (E-value < 1e⁻⁵) yielded 939 potential homologs to known virulence factors. These contained various virulence-associated functions, including adhesion, toxin production (e.g., Colibactin, Cytolysin, and Hemolysin), secretion systems (e.g., T3SS, T4SS, and T7SS), immune modulation, surface structures (e.g., O antigen and Vi antigen), and multiple iron uptake pathways (e.g., Pyoverdine, Pyochelin, Aerobactin, Yersiniabactin, FeoAB, and Legiobactin).

Comparison of the predicted protein sequences against the PHI-base database identified 1486 potential pathogen-host interaction (PHI) related proteins. Functional analysis of these candidates revealed homologs associated with a diverse range of processes potentially involved in host interactions, including factors related to iron acquisition (e.g., homologs of ferrichrome receptor, proteins related to Fe3+ transport, Fec operon regulator, and enterobactin synthetase component A), potential plant hormone manipulation (proteins possibly associated with auxin biosynthetic pathways), host tissue degradation (e.g., putative collagenase), metabolism and stress response (e.g., cystathionine beta-lyase and terminal oxidases like cytochrome O ubiquinol

Table 1
Genome statistics and features of strain FS4 11.

Feature	Value
Genome Length (bp)	5,148,490
GC content (%)	63.74
Number of contigs	1
tRNA	82
tmRNA (Transfer-messenger RNA)	1
rRNA (5S, 16S, 23S)	6, 6, 6
ncRNAs	8
ncRNA region	16
CDSs	4,565
Pseudogenes	8
CRISPR	0
N50 value	5,148,490
Genome Coverage (x)	399.5
Completeness	91.73%
Contamination value	5.28%

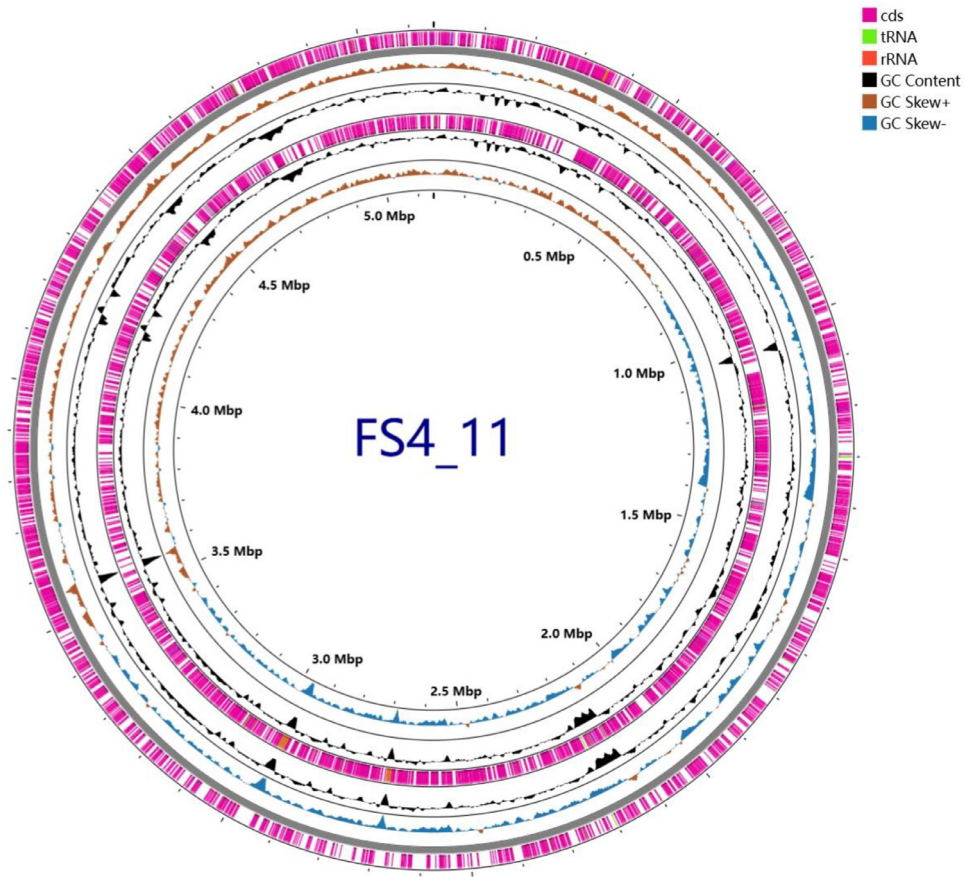


Fig. 1. Circular genome map of *C. sediminis* FS4_11 generated using Proksee. The map displays genomic features arranged in concentric circles. From the outermost circle inwards: (1 & 2) Predicted coding sequences (CDS) on the forward (sense) and reverse (antisense) strands (magenta). (3) Transfer RNA (tRNA, green) and ribosomal RNA (rRNA, orange) genes. (4) GC content variation plot (black, deviation from the mean of 63.74%). (5) GC skew $((G-C)/(G+C))$, with positive values shown in orange/brown and negative values in blue. The scale markers indicate the genome size in megabase pairs (Mbp), corresponding to the total genome size of 5,148,490 bp.

oxidase subunit CyoA), transport and efflux systems (e.g., ABC-type efflux pumps), and regulatory functions (e.g., ECF sigma factors).

To identify the gene responsible for FB1 transformation in strain FS4_11, a homology search was initially performed using the characterized FB1 hydrolase from *Sphingopyxis macrogoltabida* (UniProtKB D2D3B6.1) [4] as a query. However, this BLASTp analysis (E-value cutoff $< 1e^{-5}$) against the predicted proteome of FS4_11 did not yield any significant homologs. Consequently, we turned to functional annotations to identify potential candidates. Examination of the high-quality, manually curated Swiss-Prot database annotations for FS4_11 revealed a single gene annotated as a carboxylesterase. As this enzyme belongs to the type-B carboxylesterase/lipase family, consistent with the family of the known FB1 hydrolase used as the initial probe, this gene was selected as the primary candidate for encoding the FB1-hydrolyzing activity in strain FS4_11.

The pairwise Average Nucleotide Identity based on BLAST (ANIb) values between FS4_11 and related type strains are visualized in the heatmap presented in Fig. 2. The colour gradient, ranging from blue (lower identity) to red (higher identity), clearly delineates genomic relatedness, which is further supported by the hierarchical clustering shown in the accompanying dendro-

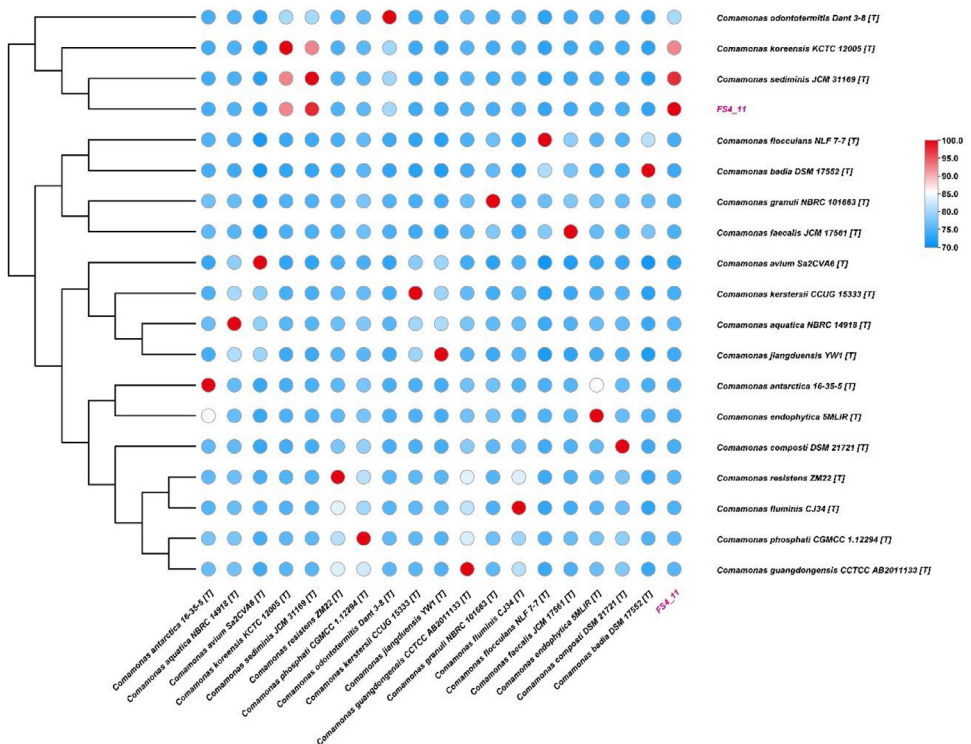


Fig. 2. Heatmap of pairwise ANiB values comparing strain FS4_11 to *Comamonas* type strains.

gram. Consistent with the quantitative results, the heatmap visually confirms the highest genomic similarity between strain FS4_11 and *C. sediminis* JCM 31169^T (indicated by the deep red colour corresponding to 97.65% ANiB). Furthermore, FS4_11 clusters closely with *C. koreensis* KCTC 12005^T and *C. odontotermitis* Dant 3-8^T, showing relatively high ANiB values with these strains and forming a distinct clade. Overall, the heatmap analysis corroborates the phylogenetic placement of FS4_11 within the *Comamonas* genus, highlighting its specific affiliation with *C. sediminis* and closely related species.

Screening against the CARD database identified two potential antibiotic resistance genes (ARGs). Both were identified via the protein homolog model as encoding homologs related to AdeF, a component of resistance-nodulation-cell division (RND) family antibiotic efflux pumps. These two AdeF homologs exhibited sequence identities of 71.36% and 58.54%, respectively, to known resistance determinants in CARD. Based on their homology, these efflux pumps are predicted to contribute to resistance primarily against fluoroquinolone and tetracycline antibiotics.

4. Experimental Design, Materials and Methods

4.1. Strain Source and Culture Conditions

Strain FS4_11, known for its FB1 degradation activity, was isolated from maize frass associated with *Sitophilus zeamais* infestation as part of a previous screening effort by our lab. The detailed methodology for isolating strain FS4_11 and confirming its FB1 degradation potential will be reported in a future publication. For DNA extraction, a fresh single colony grown on Nu-

trient Agar (NA) was picked and inoculated into 5 ml of Nutrient Broth (NB). The culture was incubated at 30°C with shaking at 180 rpm for 18 h. Bacterial cells were subsequently harvested by centrifugation at 12,000 x g for 10 min at 4°C.

4.2. DNA Extraction and Quality Assessment

Genomic DNA was extracted using the Wizard® Genomic DNA Purification Kit (Promega) following the manufacturer's protocol. The genomic DNA quality was assessed using a NanoDrop 2000 (Thermo Fisher Scientific) for purity, a Qubit 4 Fluorometer (Thermo Fisher Scientific) for concentration, and agarose gel electrophoresis for integrity.

4.3. Genome Sequencing, Assembly, Quality Assessment and Annotation

Genome sequencing was outsourced to Biomarker Technologies (Beijing, China) using a hybrid approach combining long-read (Oxford Nanopore Technologies, ONT) and short-read (Illumina) sequencing. For long-read sequencing, genomic DNA library was prepared using the SQK-LSK109 ligation sequencing kit and processed on a PromethION 48 instrument (ONT, Oxford, UK) using a PromethION R9.4.1 flow cell. Raw Nanopore reads were filtered using Filtlong v0.2.0 (<https://github.com/rwwick/Filtlong>) to remove low-quality reads (mean quality score < 6) and/or short reads (length < 2,000 bp). The filtered reads were assembled using Canu v1.5 [5], and the resulting assembly was then circularized using Circlator v1.5.5 [6]. Subsequently, this circularized assembly was polished using the filtered Nanopore reads with Racon v1.4.3 [7]. For short-read sequencing, a paired-end library with an approximate insert size of 350 bp was constructed using the Hieff NGS® OnePot Pro DNA Library Prep Kit for Illumina (Yeasen Biotechnology Co., Ltd., Shanghai, China). This library was sequenced on an Illumina NovaSeq 6000 platform (Illumina, USA) to generate 2 × 150 bp paired-end reads. The quality of the raw Illumina reads was assessed using FastQC v0.11.5 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). These reads were then mapped to the Racon-polished assembly using Burrows-Wheeler Aligner (BWA-MEM) v0.7.19 (<https://github.com/lh3/bwa>). Finally, the mapped Illumina reads were used to perform further sequence and assembly error correction using Pilon v1.23 (<https://github.com/broadinstitute/pilon/releases/>) to obtain the final polished genome sequence. The assembled genome was visualized as a circular map using the CGView web service (<https://proksee.ca/>) [8]. Genome completeness and contamination were assessed using CheckM v1.2.3 (<https://github.com/ECogenomics/CheckM?tab=readme-ov-file>). All software programs were executed using their default parameters unless otherwise specified.

Structural and functional annotation of the final polished genome was performed. Coding sequences (CDS), transfer RNA (tRNA) genes, and ribosomal RNA (rRNA) were predicted using Bakta web (<https://bakta.computational.bio/>, software v1.11.0 and database v6.0.0) [9]. Predicted proteins were functionally annotated by BLASTp comparison against the Swiss-Prot database [10] (E-value < 1e⁻⁵). Further functional insights were gained by searching protein sequences against specialized databases relevant to antibiotic resistance (CARD; Comprehensive Antibiotic Resistance Database v3.2.7 [11]), pathogen-host interactions (PHI; Pathogen-Host Interaction database v4.17 [12]), and virulence factors (VFDB; Virulence Factor Database 2022 [13]).

4.4. Average Nucleotide Identity (ANI) Analysis

To assess the genomic relatedness of *C. sediminis* FS4_11 to other species within the genus, Average Nucleotide Identity based on BLAST (ANiB) was calculated using the JSpeciesWS web service [14]. The genome sequence of FS4_11 was compared against the genome sequences of selected *Comamonas* type strains. Default parameters were used for the ANiB calculation. The resulting pairwise ANiB values were visualized as a heatmap using TBtools v1.098769 [15].

Limitations

Not applicable.

Ethics Statement

The current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

Data Availability

[BioProject \(Original data\)](#) (ncbi)
[SRA data_ONT \(Original data\)](#) (NCBI)
[SRA data_Illumina \(Original data\)](#) (NCBI)
[genome \(Original data\)](#) (NCBI)
[BioSample \(Original data\)](#) (ncbi)

CRediT Author Statement

Yang Wang: Conceptualization, Methodology, Investigation, Formal analysis, Writing – original draft, Funding acquisition; **Mengru Zhao:** Resources, Investigation, Writing – review & editing; **Zhe Wang:** Validation, Investigation, Data curation, Visualization; **Xiaohong Luo:** Investigation, Writing – review & editing; **Chengfei Wang:** Writing – review & editing; **Baoyuan Guo:** Conceptualization, Supervision, Project administration.

Acknowledgements

This research was funded by the Fundamental Research Funds for the Central Public Welfare Research Institutes (Grant No. [JY2415](#)) and the National Natural Science Foundation of China (Grant No. [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] B.A. Obafemi, I.A. Adedara, C.P. Delgado, O.T. Obafemi, M. Aschner, J.B.T. Rocha, Fumonisin B1 neurotoxicity: pre-clinical evidence, biochemical mechanisms and therapeutic strategies, *Toxicol. Rep.* 14 (2025) 101931, doi:[10.1016/j.toxrep.2025.101931](#).
- [2] L. Qu, L. Wang, H. Ji, Y. Fang, P. Lei, X. Zhang, L. Jin, D. Sun, H. Dong, Toxic mechanism and biological detoxification of fumonisins, *Toxins* 14 (2022) 182 (Basel), doi:[10.3390/toxins14030182](#).
- [3] R. Benedetti, F. Nazzi, R. Locci, G. Firrao, Degradation of Fumonisin B1 by a bacterial strain isolated from soil, *Biodegradation* 17 (2006) 31–38, doi:[10.1007/s10532-005-2797-y](#).
- [4] S. Heintl, D. Hartinger, M. Thamhesl, E. Vekiru, R. Krska, G. Schatzmayr, W.D. Moll, R. Grabherr, Degradation of fumonisin B1 by the consecutive action of two bacterial enzymes, *J. Biotechnol.* 145 (2010) 120–129, doi:[10.1016/j.jbiotec.2009.11.004](#).
- [5] S. Koren, B.P. Walenz, K. Berlin, J.R. Miller, N.H. Bergman, A.M. Phillippy, Canu: scalable and accurate long-read assembly via adaptivek-mer weighting and repeat separation, *Genome Res.* 27 (2017) 722–736, doi:[10.1101/gr.215087.116](#).

- [6] M. Hunt, N.D. Silva, T.D. Otto, J. Parkhill, J.A. Keane, S.R. Harris, Circlator: automated circularization of genome assemblies using long sequencing reads, *Genome Biol.* (2015) 16, doi:[10.1186/s13059-015-0849-0](https://doi.org/10.1186/s13059-015-0849-0).
- [7] R. Vaser, I. Sović, N. Nagarajan, M. Šikić, Fast and accurate de novo genome assembly from long uncorrected reads, *Genome Res.* 27 (2017) 737–746, doi:[10.1101/gr.214270.116](https://doi.org/10.1101/gr.214270.116).
- [8] J.R. Grant, E. Enns, E. Marinier, A. Mandal, E.K. Herman, C.Y. Chen, M. Graham, G.V. Domselaar, P. Stothard, Proksee: in-depth characterization and visualization of bacterial genomes, *Nucleic Acids Res.* 51 (2023) W484–W492, doi:[10.1093/nar/gkad326](https://doi.org/10.1093/nar/gkad326).
- [9] S. Beyvers, L. Jelonek, A. Goesmann, O. Schwengers, Bakta Web – rapid and standardized genome annotation on scalable infrastructures, *Nucleic Acids Res.* (2025), doi:[10.1093/nar/gkaf335](https://doi.org/10.1093/nar/gkaf335).
- [10] A. Bairoch, The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000, *Nucleic Acids Res.* 28 (2000) 45–48, doi:[10.1093/nar/28.1.45](https://doi.org/10.1093/nar/28.1.45).
- [11] B. Jia, A.R. Raphenya, B. Alcock, N. Wagglechner, 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, 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](https://doi.org/10.1093/nar/gkw1004).
- [12] M. Urban, A. Cuzick, J. Seager, N. Nonavinakere, J. Sahoo, P. Sahu, V.L. Iyer, L. Khamari, M. Martinez, K.E. Hammond-Kosack, PHI-base – the multi-species pathogen–host interaction database in 2025, *Nucleic Acids Res.* 53 (2024) D826–D838, doi:[10.1093/nar/gkae1084](https://doi.org/10.1093/nar/gkae1084).
- [13] B. Liu, D. Zheng, S. Zhou, L. Chen, J. Yang, VFDB 2022: a general classification scheme for bacterial virulence factors, *Nucleic Acids Res.* 50 (2021) D912–D917, doi:[10.1093/nar/gkab1107](https://doi.org/10.1093/nar/gkab1107).
- [14] M. Richter, R. Rosselló-Móra, F. Oliver Glöckner, J. Peplies, JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison, *Bioinformatics* 32 (2015) 929–931, doi:[10.1093/bioinformatics/btv681](https://doi.org/10.1093/bioinformatics/btv681).
- [15] C. Chen, H. Chen, Y. Zhang, H.R. Thomas, M.H. Frank, Y. He, R. Xia, TBtools: an integrative toolkit developed for interactive analyses of big biological data, *Mol. Plant* 13 (2020) 1194–1202, doi:[10.1016/j.molp.2020.06.009](https://doi.org/10.1016/j.molp.2020.06.009).