Draft genome sequence data of Micrococcus yunnanesis strain ORF15-23 from rice rhizosphere soil in Thailand

Kawiporn Chinachanta, Fapailin Chaiwan, Doan Trung Luu, Wasu Pathom-Aree

Data in Brief 54 (2024) 110466

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Draft genome sequence data of *Micrococcus yunnanesis* strain ORF15-23 from rice rhizosphere soil in Thailand

Kawiporn Chinachanta[a,b,c], Fapailin Chaiwan[a], Doan Trung Luu[c], Wasu Pathom-aree[b,d,*]

[a] *Department of Plant and Soil Sciences, Faculty of Agriculture, Chiang Mai University, Chiang Mai 50200, Thailand*
[b] *Center of Excellence in Microbial Diversity and Sustainable Utilization, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand*
[c] *IPSiM, Univ Montpellier, CNRS, INRAE, Institut Agro, 34060 Montpellier, France*
[d] *Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand*

ARTICLE INFO

Article history:
Received 13 March 2024
Revised 16 April 2024
Accepted 18 April 2024
Available online 24 April 2024

Dataset link: Micrococcus yunnanesis strain ORF15-23 (Original data)

Keywords:
Micrococcus
Genome sequence
Plant growth promoting actinobacteria
Indole-3-acetic acid
Potassium solubilization
2-acetyl-1-pyrroline

ABSTRACT

A Gram-positive bacterium designated as strain ORF15-23 was isolated from a soil sample collected from rainfed organic paddy fields in Roi Et province, Thailand. This strain is previously reported to produce indole-3-acetic acid and 2-acetyl-1-pyrroline (2AP) compound, solubilize potassium feldspar and promote growth of rice seedlings. The genome sequencing was carried out using Illumina MiSeq platform. The draft genome of strain ORF15-23 was 2,562,005 bp in length with 1677 protein coding sequences and an average G + C content of 72.97 mol.%. Phylogenomic tree supports the assignment of strain ORF15-23 as member of the genus *Micrococcus*. A comparison of average nucleotide identity (ANIb) values revealed that strain ORF15-23 shared 96.95 % identity with the genome of *M. yunnanensis* DSM 21948^T. The draft genome sequence of *M. yunnanesis* ORF15-23 has been deposited in the DDBJ/EMBL/GenBank databases under the accession number JAZDRZ000000000. This genome sequence data provides insightful information for the taxonomic

* Corresponding author at: Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand.

E-mail addresses: kawiporn.ch@cmu.ac.th (K. Chinachanta), fapailin.c@cmu.ac.th (F. Chaiwan), doantrung.luu@cnrs.fr (D.T. Luu), wasu.p@cmu.ac.th (W. Pathom-aree).

https://doi.org/10.1016/j.dib.2024.110466

characterization and further biotechnological exploitation of *M. yunnanesis* ORF15-23.

Specifications Table

Subject	Biology
Specific subject area	Microbiology, Genomics
Type of data	Table Figure Draft genome sequence
Data collection	A pure culture of *Micrococcus yunnanesis* ORF15-23 was routinely cultured on tryptic soy agar (TSA) at 37 °C. Genomic DNA was extracted from a 24 h culture on TSA and used as template for sequencing reaction
Data source location	• *District: Kasetwisai* • *City: Roi Et* • *Country: Thailand* • *Latitude and longitude*: 15.64 N, 103.65 E
Data accessibility	Repository name: DDBJ/GenBank/EMBL Data identification number: JAZDRZ000000000 Direct URL to data: https://www.ncbi.nlm.nih.gov/nuccore/JAZDRZ000000000

1. Value of the Data

- The draft genome data of *Micrococcus yunnanesis* ORF15-23 can provide insights for the understanding of several of its properties, such as indole-3-acetic acid production, potassium solubilization from *K*-feldspar rocks, improvement of 2-acetyl-1-pyrroline content in KDML105 rice seedlings and promotion of their growth.
- These data are valuable resources for researchers working in the field of microbiology, agronomy, genomics, and molecular biology.
- This genome data can be used in comparative genomics of members of the genus *Micrococcus* for biotechnological and taxonomic purposes and allow in-depth analysis of *Micrococcus yunnanesis* ORF15-23 via genome mining.

2. Background

Micrococcus yunnanesis ORF15-23 was isolated from soil collected from rainfed paddy fields, in Tung Kula Rong Hai (TKR) areas, Kasetwisai district, Roi Et province, Thailand (15.64 N, 103.65 E) during the dry season before the rice harvest (November 2018) [1]. Rhizosphere soil samples (0–15 cm depth) were randomly collected using sterile spoon from 10 spots per composite soil sample. The strain was isolated by dilution spread plate on nutrient agar. In a series of experiments, we previously found that the strain ORF15-23 was able to produce indole-3-acetic acid [2], solubilize potassium from *K*-feldspar rocks [3], produce 2-acetyl-1-pyrroline (2AP) in KDML105 rice seedlings [4]. *M. yunnanesis* ORF15-23 is also a salt tolerant strain which exhibited ability to promote the growth of KDML105 rice seedlings under salinity stress [5]. Here we report the genome sequence of strain ORF15-23 to allow its identification at genus and species levels, and also to facilitate further molecular studies.

3. Data Description

3.1. Genome assembly and annotation

The annotated genome of *Micrococcus yunnanesis* ORF15-23 was analyzed using the PATRIC genome analysis server (https://www.patricbrc.org/) [6]. Table 1 summarized the genome characteristics of *M. yunnanesis* ORF15-23. The draft genome contains 64 contigs, with genome length of 2,562,005 bp, N50 and L50 values of 136,364 and 6, respectively. The genome contains 2428 protein coding sequence (CDS), 48 transfer RNA (tRNA) genes, 2 ribosomal RNA (rRNA) genes with 72.97 G + C content (%). The sequence was deposited in DDBJ/GenBank/EMBL databases under accession number JAZDRZ000000000 and can be accessed at https://www.ncbi.nlm.nih.gov/nuccore/JAZDRZ000000000.

M. yunnanensis ORF15-23 genome was also annotated using RAST tool kit (RASTtk) [7]. This genome is in the superkingdom Bacteria and was annotated using genetic code 11. The taxonomy of this genome is: cellular organisms > Bacteria > Terrabacteria group > Actinomycetota > Actinomycetes > Micrococcales > Micrococcaceae > *Micrococcus* > *Micrococcus yunnanensis*.

The annotation included 751 hypothetical proteins and 1677 proteins with functional assignments (Table 2). The proteins with functional assignments included 666 proteins with Enzyme Commission (EC) numbers [8], 564 with Gene Ontology (GO) assignments [9], and 512 proteins that were mapped to KEGG pathways [10]. PATRIC annotation includes two types of protein families [11], and this genome has 2297 proteins that belong to the genus-specific protein families (PLFams), and 2340 proteins that belong to the cross-genus protein families (PGFams). A circular map of *Micrococcus yunnanesis* ORF15-23 genome presents the distribution of genome annotation is showed in Fig. 1.

Table 1
Genome characteristics of *Micrococcus yunnanesis* ORF15-23.

Features	Value
Number of contigs	64
Genome length	2,562,005 bp
Genome coverage	150 X
Largest contig	298,861 bp
GC Content (%)	72.97
Plasmid	0
Contig N50	136,364
Contig L50	6
Protein coding sequence (CDS)	2428
CDS	2428
tRNA	48
rRNA	2
Repeat regions	13

Table 2
Protein features of *Micrococcus yunnanesis* ORF15-23.

Features	Value
Hypothetical proteins	751
Proteins with functional assignments	1677
Proteins with EC number assignments	666
Proteins with GO assignments	564
Proteins with Pathways assignments	512
Proteins with PATRIC genus-specific family (PLfams) assignments	2297
Proteins with PATRIC cross-genus family (PGfams) assignments	2340

Fig. 1. The distribution of annotated genomic features. This includes, from outer to inner rings, the contigs, CDS on the forward strand, CDS on the reverse strand, RNA genes, CDS with homology to know antimicrobial resistance genes, CDS with homology to known virulence factors, GC content and GC skew.

3.2. Subsystem analysis

An overview of the subsystems for this genome is provided in Fig. 2. A subsystem is a set of proteins that together implement a specific biological process or structural complex [12] and PATRIC annotation includes an analysis of the subsystems unique to each genome. The number of genes assigned to each biological processes is as followed: metabolism (490), cellular processes (66), protein processing (206), energy (149), stress response, defense and virulence (80), DNA processing (67), membrane transport (48), RNA processing (39), cell envelop (7), miscellaneous (5) and regulation and cell signaling (13). Many of the annotated genes showed homology to known transporters [13], virulence factors [14,15], drug targets [16,17], and antibiotic resistance genes [18]. The number of genes and the specific source database where homology was found is provided in Table 3.

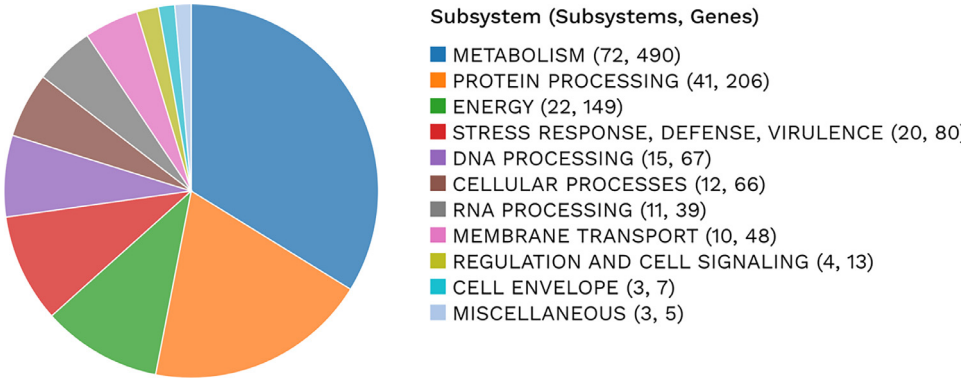

Fig. 2. PATRIC annotation using RAST tool kit (RASTtk) of *Micrococcus yunnanesis* ORF15-23's genome.

Table 3
Specialty genes of *Micrococcus yunnanesis* ORF15-23.

Specialty genes	Source	Genes
Antibiotic resistance	CARD	1
Antibiotic resistance	PATRIC	27
Drug targets	DrugBank	5
Drug targets	TTD	1
Transporters	TCDB	5
Virulence factors	PATRIC_VF	3
Virulence factors	VFDB	1
Virulence factors	Victors	1

Table 4
Antimicrobial resistance (AMR) genes of *Micrococcus yunnanesis* ORF15-23.

AMR mechanism	Genes
Antibiotic target in susceptible species	Alr, Ddl, dxr, EF-G, EF-Tu, folA, Dfr, folP, gyrA, gyrB, Iso-tRNA, kasA, MurA, rho, rpoB, rpoC, S10p, S12p
Antibiotic target replacement protein	FabG, FabL-like, HtdX
Gene conferring resistance via absence	gidB
Protein altering cell wall charge conferring antibiotic resistance	GdpD, PgsA
Regulator modulating expression of antibiotic resistance gene	LpqB, Mtra, MtrB

3.3. Antimicrobial resistance genes

The annotation of *Micrococcus yunnanesis* ORF15-23 genome in PATRIC uses k-mer-based antimicrobial resistance (AMR) genes detection method which utilizes PATRIC's curated collection of representative AMR gene sequence variants [6]. Each AMR gene was assigned functional annotation, broad mechanism of antibiotic resistance, drug class and, in some cases, specific antibiotic it confers resistance to. A summary of the AMR genes annotated in this genome and corresponding AMR mechanism are shown in Table 4.

The phylogenomic tree showed that *M. yunnanesis* ORF15-23 formed clade with several genera within of the family Micrococcaceae including the type of strain of *M. luteus* NCTC 2665 465515.4 and *Citriciccus* sp. CH26A 1045009.3 (Fig. 3). Additionally, the average nucleotide identity (ANIb) values between *M. yunnanesis* ORF15-23 and type strains of *M. yunnanensis* DSM 21948^T and *M. aloeverae* DSM 27472^T were 96.95 and 96.83 %, respectively (Table 5). These data strongly supported the assignment of strain ORF15-23 as *M. yunnanesis* ORF15-23.

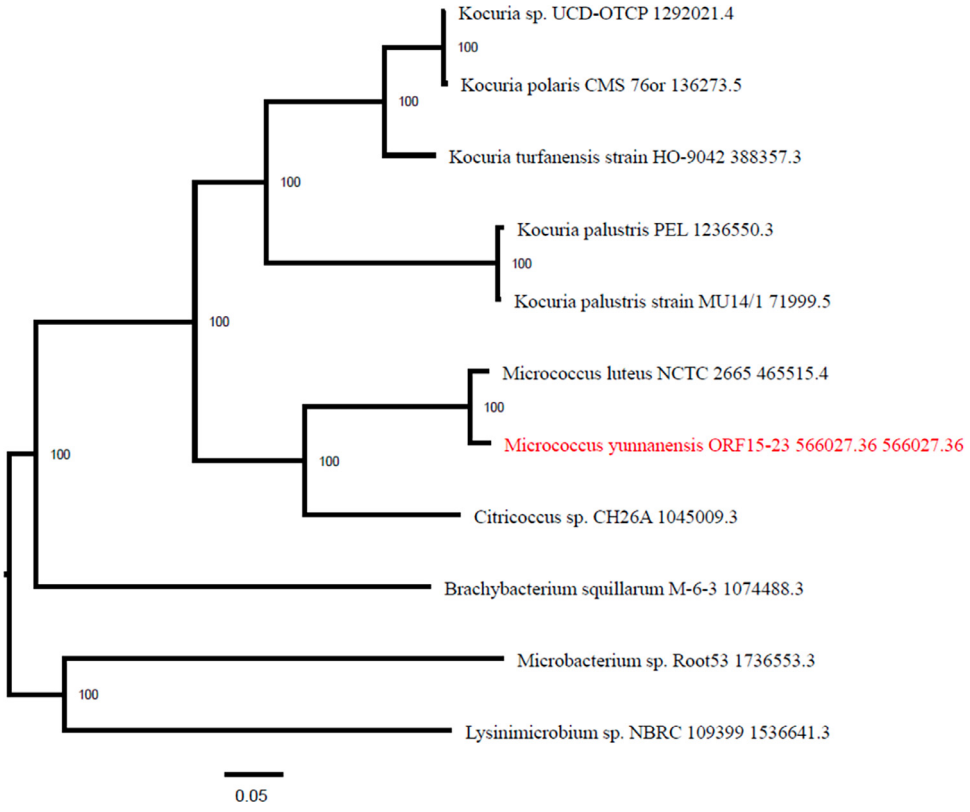

Fig. 3. Phylogenomic tree of *Micrococcus yunnanesis* ORF15-23 and its closely related *Micrococcus* spp. genome generated using the Type (Strain) Genome Server (TYGS).

Table 5
ANIb values of *M. yunnanesis* ORF15-23 with its closely related *Micrococcus* spp.

Genome	ANIb (%)
Micrococcus yunnanensis DSM 21948 [T]	96.95
Micrococcus aloeverae DSM 27472 [T]	96.82
Micrococcus yunnanensis DSM 24531 [T]	96.36
Micrococcus luteus NCTC 2665 NCTC 2665 [T]	96.35
Micrococcus luteus NCTC 2665 NCTC 46698 [T]	96.33
Micrococcus luteus ATCC 4698 [T]	96.31

4. Experimental Design, Materials and Methods

4.1. Sample collection and bacterial isolation

Rhizosphere soil sample was randomly collected (0–15 cm depth) from organic rainfed paddy fields of KDML105 rice variety, in Tung Kula Rong Hai (TKR) areas, Kasetwisai district, Roi Et province, Thailand (15.64 N, 103.65 E) during dry season before the rice harvesting in November 2018. The soil was collected using sterile spoon. The strain was isolated by dilution spread plate on nutrient agar.

4.2. Bacterial cultivation and genomic DNA extraction

M. yunnanesis ORF15-23 was grown on tryptic soy agar (TSA) at 37 °C for 24 h. Genomic DNA extraction was performed by the following procedures [19]. Bacterial cells were lysed in an extraction buffer. The contaminated protein was removed by phenol extraction and centrifuged at 15,000 rpm for 5 min at 4 °C. The supernatant was collected, and phenol extraction step was repeated. Sodium acetate, isopropanol and absolute ethanol were added to precipitate DNA, and incubated at −20 °C for 15 min. After incubation, precipitated DNA was harvested by centrifugation at 15,000 rpm for 5 min at 4 °C. The DNA was washed by 70 % (v/v) ethanol and centrifuged at 15,000 rpm for 5 min at 4 °C. DNA was dried for 30 min and dissolved with sterile ultrapure water.

4.3. Whole genome sequencing, assembly annotation and analysis

The sequencing of *Micrococcus yunnanesis* ORF15-23 genomic DNA was carried out using service of Omics Science and Bioinformatics Center, Faculty of Science, Chulalongkorn University, Bangkok, Thailand. The genomic DNA library was prepared using QIASEQ FX DNA library preparation kit (Qiagen, USA). The libraries were sequenced on Illumina MiSeq sequencer in 2 × 250 bp paired end. FASTQC software version 0.11.9 [20] was used to check raw reads quality. Adaptors and poor-quality reads were removed using Fastp version 0.23.2 [21], and the filtered reads were used as an input for Unicycler, genome assembly program [22]. The assembled genome was annotated using the PATRIC RASTtk-enabled Genome Annotation Service [7]. ANIb value was calculated using JSpeciesWS version: 3.9.7, web server tool [23]. In addition, the phylogenomic tree was constructed using the Type (Strain) Genome Server (TYGS) [24] (https://tygs.dsmz.de/). All software were run using the default parameters.

Limitations

Not applicable.

Ethics Statement

This study did not involve any human subjects and animal experiments. No ethical approval was required.

Data Availability

Micrococcus yunnanesis strain ORF15-23 (Original data) (GenBank).

CRediT Author Statement

Kawiporn Chinachanta: Conceptualization, Methodology, Investigation, Data curation, Writing – original draft, Writing – review & editing; **Fapailin Chaiwan:** Writing – review & editing; **Doan Trung Luu:** Supervision, Writing – review & editing; **Wasu Pathom-aree:** Conceptualization, Data curation, Supervision, Writing – review & editing.

Acknowledgments

This research study was funded by the National Research Council of Thailand. K.C. is supported by a New Researcher Career Path for the year 2023 from National Research Council

of Thailand (NRCT), and Chiang Mai University (Project number: N42A660394). D.T.L. and K.C. were funded through LabEx AGRO ANR-10-LABX-0001-01 (under I-Site Université de Montpellier framework). W.P. and K.C. -a. acknowledge the partial support from Chiang Mai University through the Center of Excellence in Microbial Diversity and Sustainable Utilization.

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] K. Chinachanta, L. Herrmann, D. Lesueur, S. Jongkaewwattana, C. Santasup, A. Shutsrirung, Influences of farming practices on soil properties and the 2-acetyl-1-pyrroline content of Khao Dawk Mali 105 rice grains, Appl. Environ. Soil Sci. 2020 (2020) 8818922, doi:10.1155/2020/8818922.
[2] K. Chinachanta, A. Shutsrirung, L. Herrmann, D. Lesueur, Isolation and characterization of KDML105 aromatic rice rhizobacteria producing indole-3-acetic acid: impact of organic and conventional paddy rice practices, Lett. Appl. Microbiol. 74 (3) (2022) 354–366, doi:10.1111/lam.13602.
[3] K. Chinachanta, A. Shutsrirung, Screening for P-and K-solubilizing, and siderophore producing capacity of rhizobacteria from Khao Dawk Mali 105 aromatic rice, IOP Conf. Ser.: Earth Environ. Sci. 858 (1) (2021) 012004, doi:10.1088/1755-1315/858/1/012004.
[4] K. Chinachanta, A. Shutsrirung, L. Herrmann, D. Lesueur, W. Pathom-aree, Enhancement of the aroma compound 2-acetyl-1-pyrroline in Thai Jasmine rice (*Oryza sativa*) by Rhizobacteria under salt stress, Biology 10 (2021) 1065, doi:10.3390/biology10101065.
[5] K. Chinachanta, A. Shutsrirung, C. Santasup, W. Pathom-Aree, D.T. Luu, L. Herrmann, D. Lesueur, C. Prom-u-thai, Rhizoactinobacteria enhance growth and antioxidant activity in Thai Jasmine rice (*Oryza sativa*) KDML105 seedlings under salt stress, Plants 12 (2023) 3441, doi:10.3390/plants12193441.
[6] A.R. Wattam, J.J. Davis, R. Assaf, S. Boisvert, T. Brettin, C. Bun, N. Conrad, E.M. Dietrich, T. Disz, J.L. Gabbard, S. Gerdes, C.S. Henry, R.W. Kenyo, D. Machi, C. Ma, E.K. Nordberg, Gary J Olsen, D.E. Murphy-Olson, R. Olson, R. Overbeek, B. Parrello, G.D. Pusch, M. Shukla, V. Vonstein, A. Warren, F. Xia, H. Yoo, R.L. Stevens, Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center, Nucleic Acids Res. 45 (2017) D535–D542, doi:10.1093/nar/gkw1017.
[7] T. Brettin, J.J. Davis, T. Disz, R.A. Edwards, S. Gerdes, G.J. Olsen, R. Olson, R. Overbeek, B. Parrello, G.D. Pusch, M. Shukla, RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes, Sci. Rep. 5 (8365) (2015) 1–6, doi:10.1038/srep08365.
[8] I. Schomburg, A. Chang, C. Ebeling, M. Gremse, C. Heldt, G. Huhn, D. Schomburg, BRENDA, the enzyme database: updates and major new developments, Nucleic Acids Res. 32 (2004) D431–D433, doi:10.1093/nar/gkh081.
[9] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig. M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, G. Sherlock, Gene ontology: tool for the unification of biology, Nat. Genet. 25 (2000) 25–29, doi:10.1038/75556.
[10] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, M. Tanabe, KEGG as a reference resource for gene and protein annotation, Nucleic Acids Res. 44 (2016) D457–D462, doi:10.1093/nar/gkv1070.
[11] J.J. Davis, S. Gerdes, G.J. Olsen, R. Olson, G.D. Pusch, M. Shukla, V. Vonstein, A.R. Wattam, H. Yoo, PATtyFams: protein families for the microbial genomes in the PATRIC database, Front. Microbiol. 7 (118) (2016), doi:10.3389/fmicb.2016.00118.
[12] R. Overbeek, T. Begley, R.M. Butler, J.V. Choudhuri, H.-Y. Chuang, M. Cohoon, V. de Crécy-Lagard, N. Diaz, T. Disz, R. Edwards, The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes, Nucleic Acids Res. 33 (2005) 5691–5702, doi:10.1093/nar/gki866.
[13] M.H. Saier Jr, V.S. Reddy, B.V. Tsu, M.S. Ahmed, C. Li, G. Moreno-Hagelsieb, The transporter classification database (TCDB): recent advances, Nucleic Acids Res. 44 (2015) D372–D379, doi:10.1093/nar/gkv1103.
[14] C. Mao, D. Abraham, A.R. Wattam, M.J. Wilson, M. Shukla, H.S. Yoo, B.W. Sobral, Curation, integration and visualization of bacterial virulence factors in PATRIC, Bioinformatics 31 (2015) 252–258, doi:10.1093/bioinformatics/btu631.
[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] F. Zhu, B. Han, P. Kumar, X. Liu, X. Ma, X. Wei, L. Huang, Y. Guo, L. Han, C. Zheng, Update of TTD: therapeutic target database, Nucleic Acids Res. 38 (2009) D787–D791, doi:10.1093/nar/gkp1014.
[17] V. Law, C. Knox, Y. Djoumbou, T. Jewison, A.C. Guo, Y. Liu, A. Maciejewski, D. Arndt, M. Wilson, V. Neveu, A. Tang, G. Gabriel, C. Ly, S. Adamjee, Z.T. Dame, B. Han, Y. Zhou, D.S. Wishart, DrugBank 4.0: shedding new light on drug metabolism, Nucleic Acids Res. 42 (2014) D1091–D1097, doi:10.1093/nar/gkt1068.
[18] A.G. McArthur, N. Waglechner, F. Nizam, A. Yan, M.A. Azad, A.J. Baylay, K. Bhullar, M.J. Canova, G. De Pascale, L. Ejim, The comprehensive antibiotic resistance database, Antimicrob. Agents Chemother. 57 (2013) 3348–3357, doi:10.1128/AAC.00419-13.
[19] P. Rangseekaew, A. Barros-Rodriguez, W. Pathom-aree, M. Manzanera, Plant beneficial deep-sea actinobacterium, *Dermacoccus abyssi* MT1.1^{T} promote growth of tomato (*Solanum lycopersicum*) under salinity stress, Biology 11 (2) (2022) 191, doi:10.3390/biology11020191.

[20] S. Andrews (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc. Accessed May 18, 2022.
[21] S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor, Bioinformatics 34 (1) (2018) i884–i890, doi:10.1093/bioinformatics/bty560.
[22] 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 (6) (2017) e1005595, doi:10.1371/journal.pcbi.1005595.
[23] M. Richter, R. Rosselló-Móra, F.O. Glöckner, J. Peplies, JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison, Bioinformatics 32 (6) (2015) 929–931, doi:10.1093/bioinformatics/btv681.
[24] J.P. Meier-Kolthoff, M. Goker, TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy, Nat. Commun. 10 (2182) (2019) 1–10, doi:10.1038/s41467-019-10210-3.