Data in Brief 53 (2024) 110143

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Draft genome sequence data of Antarctic *Penicillium sp.* strain E22, from Deception Island

Teoh Chui Peng[a], Paris Lavin[b,c,*], Rómulo Oses Pedraza[d], Natalia Fierro-Vásquez[b], Cristina Purcarea[e], Sheau Ting Yong[f], Clemente M.V.L. Wong[a]

[a] *Biotechnology Research Institute, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia*
[b] *Departamento de Biotecnología, Facultad de Ciencias del Mar y Recursos Biológicos, Universidad de Antofagasta, Antofagasta 1240300, Chile*
[c] *Centro de Investigación en Inmunología y Biotecnología Biomédica de Antofagasta, (CIIBBA), Universidad de Antofagasta, Antofagasta 1240300, Chile*
[d] *Centro Regional de Investigación y Desarrollo Sustentable de Atacama (CRIDESAT), Universidad de Atacama, Copiapó, Chile*
[e] *Institute of Biology Bucharest of the Romanian Academy, 296 Splaiul Independentei, 060031 Bucharest, Romania*
[f] *INEOS Oxford Institute for Antimicrobial Research, Department of Biology, University of Oxford, Oxford, UK*

ARTICLE INFO

Article history:
Received 21 December 2023
Accepted 26 January 2024
Available online 9 February 2024

Dataset link: Draft genome of Penicillium sp. strain E22 (Original data)

Keywords:
Penicillium
Antarctica
Deception Island, Draft genome sequencing

ABSTRACT

Here, we report the draft genome sequence and assembly of the *Penicillium sp.* strain E22, which was isolated from Antarctic soil of Deception Island, South Shetland Islands close to the Antarctic Peninsula. The genome was sequenced using a 2 # 250 bp paired-end method by Illumina MiSeq 6000. The genome assembly was performed using softwares implemented in the Kbase web service. The phylogenetic tree of strain E22 comparing its internal transcribed spacer (ITS) region with the other Penicillium showed high genetic similarity to Penicillium griseofulvum MN545450 and Penicillium camemberti MT530220. Draf genome of *Penicillium sp.* strain E22 comprises 33,653 coding sequences, with a high G + C content of 48.32% and a total size of 37,484,944 bp. This

* Corresponding author.
E-mail address: paris.lavin@uantof.cl (P. Lavin).

https://doi.org/10.1016/j.dib.2024.110143

draft genome assembly version has been deposited at GenBank under accession JASJUN000000000.

Specifications Table

Subject	Microbiology • Fungal Biology
Specific subject area	The genome sequence was processed in Illumina MiSeq 6000 De novo assembly: SPAdes Genome Assembler software (v3.15.3), Annotation: DRAM (Distilled and Refined Annotation of Metabolism) software (v0.1.2) as implemented in the Kbase web service.
Data format	Raw, Analyzed, Filtered and deposited
Type of data	Table, Figure
Data collection	Purification of genomic DNA from a pure culture of *Penicillium sp*. strain E22 isolated from Antarctic soil, The sequencing library was generated using the Nextera® XT DNA sample preparation kit for Illumina. Illumina MiSeq PE250 was used for whole genome sequencing.
Data source location	The strain E22 was isolated from the soil of Deception Island (S 62° 55′ 58.1′′ W 60° 35′ 26.8′′), Antarctica.
Data accessibility	Data are deposited at the NCBI GenBank https://www.ncbi.nlm.nih.gov/bioproject/PRJNA970415 https://www.ncbi.nlm.nih.gov/sra/SRR24472943

1. Value of the Data

- The availability of the draft genome assembly for *Penicillium sp*. strain E22 provides significant benefits for microbial taxonomy and ecological studies, especially in terms of identifying and mapping species distribution.
- The information presented in this article has the potential to be beneficial for researchers who are engaged in environmental microbiology, environmental biotechnology, extremophiles and genomics.
- The genomic data of *Penicillium sp*. strain E22 contained in this report could be a valuable asset for scientists who wish to conduct comparative genomic analyses across different strains and environment.

2. Background

The genus Penicillium comprises the most extensively distributed fungi, which are present universally in both outdoor and indoor environments, including food, water, plants, and soils. Presently, there are 354 acknowledged species in this genus, and numerous among them have the ability to generate a wide range of natural products and enzymes, including amylases, glucoamylase, cellulase, proteases, and xylanase [1–3].

2.1. Data description

The data presented here represents the genome sequencing, assembly, and annotation of the Antarctic *Penicillium* strain E22, isolated from Deception Island soil. Illumina sequencing yielded

Table 1
QUAST report and genome features for Penicillium sp. strain E22 assembly.

Statistics without reference	*Penicillium sp.* strain E22
# contigs	2,704
# contigs (>= 0 bp)	2,705
# contigs (>= 1000 bp)	1,846
# contigs (>= 10000 bp)	769
# contigs (>= 100000 bp)	70
# contigs (>= 1000000 bp)	0
Largest contig	298,884
Total length	37,484,944
Total length (>= 0 bp)	37,485,436
Total length (>= 1000 bp)	36,903,868
Total length (>= 10000 bp)	33,127,976
Total length (>= 100000 bp)	9,345,970
Total length (>= 1000000 bp)	0
N50	53,495
N75	24,743
L50	203
L75	460
GC (%)	48.32
Mismatches	
# N's	2597
# N's per 100 kbp	6.93
Genome features	
Total coding sequences	33,653
tRNA genes	198
rRNA genes	50

874.33 million paired-end reads. The N50 contig length was 53.9 Kb with an average coverage of 24 × . The resulting draft genome was 37,484,944 bp in size with a G+C content of 48.32 %. Gene prediction analysis using the kb_DRAM web-based app in KBase (v.0.1.2) [4], resulted in 33,653 protein coding genes (Table 1).

Based on the comparison of the internal transcribed spacer (ITS) region of the 18S–5.8S–26S nuclear ribosomal of the isolate to other strains, it was found that it had the closest genetic similarity to *Penicillium griseofulvum* MN545450 and *Penicillium camemberti* MT530220, with a 99.15% identity with both species (Fig. 1). Functional gene annotation of the draft genome predicted about 3253 genes using KEGG. The carbohydrate-active enzyme analysis showed that *Penicillium sp.* strain E22 was dominated by AA1, AA3, GH13, GT2, GH16, GH43 and GH5. Different types of secondary metabolite clusters that may be involved in the formation of secondary metabolites were found: T1PKS, NRPS, NRPS-like, fungal-RiPP-like, NI-siderophore, betalactone, indole, terpene and and several hybrids (NRPS,T1PKS; NRPS,indole; NRPS-like,T1PKS; NRPS,fungal-RiPP-like; NRP-metallophore-NRPS and T1PKS,indole,NRPS-like,terpene). This whole genome project has been deposited at NCBI GenBank under accession number for Bioproject, Biosample and SRA as PRJNA970415, SAMN35003752 and SRR24472943, respectively. The assembly version described in this paper is version JASJUN000000000.

3. Experimental Design, Materials and Methods

3.1. Genome DNA extraction and sequencing

Penicillium sp. strain E22 was isolated from Deception Island (62° 55′ 58.1′'S 60° 35′ 26.8′'W), Antarctica. Strain E22 was routinely cultivated in Yeast Malt Extract Agar medium at 28°C for 3 days. TRIzol™ (Invitrogen™, USA) was used for genomic DNA extraction. The genomic library of strain E22 was generated using Nextera® XT DNA sample preparation kit according to the

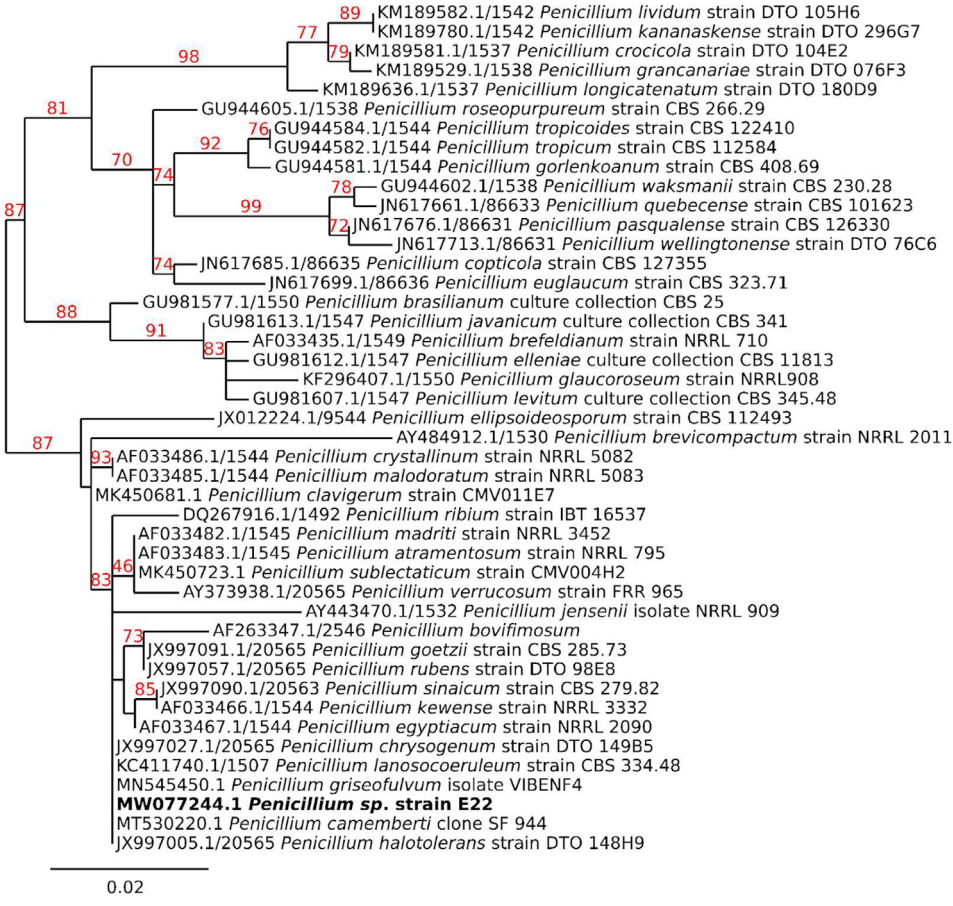

Fig. 1. Phylogenetic tree of ITS region sequences inferred by maximum-likelihood method. Numbers above branches indicate 1000 bootstrap replicates values and 44 sequences of Penicillium species used are presented with GenBank accession numbers followed by the name of strains.

manufacturer's instructions. The whole genome sequencing was then performed by using an Illumina MiSeq PE250 at the Biotechnology Research Institute, Universiti Malaysia Sabah.

3.2. Species identification

The DNA fragment was amplified using universal primer set ITS1 (forward primer) 5-TCCGTAGGTGAACCTGCGG-3 and ITS4 (reverse primer) 5-TCCTCCGCTTATTGATATGC-3. The PCR product was sequenced using bi-directional sequencing. The sequence was analyzed by BLAST and then compared to the NCBI database. The phylogenetic tree was constructed using the Maximum-likelihood phylogenetic tree based on ITS rRNA gene sequences (879 base pair alignment positions including gaps; Substitution model: HKY85; Gamma shape parameter: 0.467; Transition/transversion ratio: 3.761; Number of categories: 4 and Proportion of invariant: 0.740) showing the relationship between strain E22 and the 43 most closely related reference species. The alignment, substitution model and construction of the phylogenetic tree were performed using the online Phylogeny.fr tool [4].

3.3. Reads pre-processing, genome assembly, quality assessment, and annotation

The raw reads were pre-processed using the Trimmomatic (v1.2.14) tool to trim low-quality bases and short reads (minimum length=36), then assembled using SPAdes Genome Assembler software (v3.15.3), Quast Report (QUality ASsessment Tool, v4.4) and the Annotation performed with DRAM (Distilled and Refined Annotation of Metabolism) software (v0.1.2). All software used was implemented in the Kbase web service [5].

Limitations

'Not applicable'.

Ethics Statement

This work neither involves human subjects nor animal subjects. The authors declare that this manuscript is original work and has not been published elsewhere.

Data Availability

Draft genome of Penicillium sp. strain E22 (Original data) (OSF)

CRediT Author Statement

Teoh Chui Peng: Formal analysis, Writing – original draft; **Paris Lavin:** Conceptualization, Supervision, Methodology, Writing – review & editing; **Rómulo Oses Pedraza:** Writing – review & editing; **Natalia Fierro-Vásquez:** Formal analysis, Writing – original draft; **Cristina Purcarea:** Writing – review & editing; **Sheau Ting Yong:** Data curation; **Clemente M.V.L. Wong:** Writing – review & editing, Conceptualization, Supervision, Methodology.

Acknowledgements

This work was supported by the INACH RT_20-19 Project (Instituto Antartico Chileno), Fondecyt Iniciación grant No 11190754 (National Agency of Research and Development, ANID), Convenio Mineduc-UA ANT20992, Funding granted by the Vice-Rector's Office for Research, Innovation and Postgraduate Studies at the University of Antofagasta and Romanian Academy project RO1567-IBB05/2022.

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] C.M. Visagie, J. Houbraken, J.C. Frisvad, S.-B. Hong, C.H.W. Klaassen, G. Perrone, K.A. Seifert, J. Varga, T. Yaguchi, R.A. Samson, Identification and nomenclature of the genus Penicillium, Stud. Mycol. 78 (2014) 343–371, doi:10.1016/j.simyco.2014.09.001.

[2] D.M. Kita, P. Giovanella, T.T. Yoshinaga, E.P. Pellizzer, L.D. Sette, Antarctic fungi applied to textile dye bioremediation, Anais da Academia Brasileira de Ciências 94 (2022), doi:10.1590/0001-3765202220210234.

[3] N. Vaishnav, A. Singh, M. Adsul, P. Dixit, S.K. Sandhu, A. Mathur, S.K. Puri, R.R. Singhania, Penicillium: the next emerging champion for cellulase production, Bioresour. Technol. Rep. 2 (2018) 131–140, doi:10.1016/j.biteb.2018.04.003.
[4] A. Dereeper, V. Guignon, G. Blanc, S. Audic, S. Buffet, F. Chevenet, J.F. Dufayard, S. Guindon, V. Lefort, M. Lescot, J.M. Claverie, O. Gascuel, Phylogeny.fr: robust phylogenetic analysis for the non-specialist, Nucleic Acids Res. 1 (Web Server issue) (2008) W465–W469 36, doi:10.1093/nar/gkn180.
[5] AP Arkin, RW Cottingham, CS Henry, NL Harris, RL Stevens, S Maslov, et al., KBase: the United States department of energy systems biology knowledgebase, Nat. Biotechnol. 36 (2018) 566, doi:10.1038/nbt.4163.