BMC Genomic Data

DATA NOTE

Open Access

Genome sequencing highlights the fungal sclerotium formation of medicinal fungi *Polyporus umbellatus*

Li Chi[1,2†], Tianrui Liu[3†], Zhongyi Hua[5], Pengjie Han[4*], Honghong Jiao[1] and Yuan Yuan[1,2,3*]

Abstract

Objectives *Polyporus umbellatus* is a well-known medicinal fungus in Asia. Due to its long growth cycle, wild resources of *P. umbellatus* are rapidly declining. Research on *P. umbellatus* growth is scarce, thereby impeding the investigation into the mechanism of sclerotium formation. In this study, the whole genome sequence of *P. umbellatus* was assembled and annotated.

Data description The generated genome was 79.74 Mb with an N50 of 969.73 Kbp, a GC content of 50.63%, encoding 10,864 protein-coding genes. Additionally, 82 ribosomal RNAs, 488 transfer RNAs, 25 small nuclear RNAs (snRNAs) and 741 micro RNA (miRNAs) were identified. This genomic resource will unveil molecular mechanisms of sclerotium formation and provides insight into the underlying molecular processes in medicinal fungi.

Keywords *Polyporus umbellatus*, Whole-genome sequencing, Fungal sclerotium formation, CAZymes

Objective

Polyporus umbellatus is a well-known edible and medicinal mushroom. The life cycle of *P. umbellatus* includes three stages of mycelium, sclerotium, and fruiting bodies with the dominant stage of sclerotium development. *P. umbellatus* sclerotia have been utilized in traditional Chinese medicine for over a millennium [1, 2]. In recent years, the polysaccharide extracted from *P. umbellatus* sclerotia has been demonstrated anti-cancer, immunomodulatory, antioxidant, anti-inflammatory and renoprotective activity [3, 4]. However, wild sclerotia of *P. umbellatus* have been largely depleted due to insufficient protection, over-harvesting and severe habitat loss. At present, the artificial sclerotium cultivation primarily relies on the asexual breeding of *P. umbellatus* sclerotium and its symbiotic fungi, *Armillaria* species. The growth cycle of *P. umbellatus* sclerotia typically spans 3 to 4 years before mature sclerotia can be harvested. Our previous studies have shown that *P. umbellatus* sclerotium

†Li Chi and Tianrui Liu contributed equally to this work.

*Correspondence:
Pengjie Han
hanpengjie2023@126.com
Yuan Yuan
y_yuan0732@163.com
[1]School of Pharmacy, Shaanxi University of Chinese Medicine, Xianyang 712046, China
[2]Experimental Research Center, China Academy of Chinese Medical Sciences, Beijing 100700, China
[3]Jiangxi Province Key Laboratory of Sustainable Utilization of Traditional Chinese Medicine Resources, Institute of Traditional Chinese Medicine Health Industry, China Academy of Chinese Medical Sciences, Nanchang 330115, China
[4]Shandong Engineering Research Center for Innovation and Application of General Technology for Separation of Natural Products, Shandong Analysis and Test Center, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250014, China
[5]School of Pharmacy, Jiangsu University, Zhenjiang 212013, China

Table 1 Overview of data files/data sets

Label	Name of the data files/data sets	File types	Data repository and identifiers
Data file 1	Detailed methodology	Word file (.doc)	Zenodo https://doi.org/10.5281/zenodo.14192471 [13]
Data file 2	Genomic features of the *P. umbellatus*	Word file (.doc)	Zenodo https://doi.org/10.5281/zenodo.14190274 [14]
Data file 3	Functional annotation of the *P. umbellatus* genome.	Spreadsheet (.xlsx)	Zenodo https://doi.org/10.5281/zenodo.14192681 [15]
Data set 1	Genome assembly of *P. umbellatus*	Fasta file (.fasta.gz)	NGDC https://ngdc.cncb.ac.cn/gwh/Assembly/30670/show [16]
Data set 2	Predicted genes of *P. umbellatus*	General feature format (.gff)	NGDC https://ngdc.cncb.ac.cn/gwh/Assembly/30670/show [17]
Data set 3	Pacbio sequencing reads of *P. umbellatus*	Bam file (.bam)	NGDC https://ngdc.cncb.ac.cn/gsa/browse/CRA022677/CRX14 92289 [18]
Data set 4	Illumina sequencing reads of *P. umbellatus*	Fastq file (.fq.gz)	NGDC https://ngdc.cncb.ac.cn/gsa/browse/CRA022677/CRX14 92288 [19]

formation is induced by abiotic stress, such as low temperature, hypoxia and osmotic stress [5–7], as well as the invasion of symbiotic *Armillaria* species and associated bacteria [8, 9].

Genome sequencing of fungi offers an opportunity to decipher the mechanisms underpinning the gene regulation in sclerotium development. The advent of genome sequencing and comparative analysis has largely contributed to identifying factors involved in sclerotium formation in related species, such as *Wolfiporia cocos* [10]. Many studies have explored the secondary metabolites and formation conditions of *P. umbellatus* sclerotium [11], yet a genomic perspective is lacking. Here, we report the complete genome sequence of *P. umbellatus* to facilitate the exploration the molecular mechanism underlying its slow growth, development and sclerotium formation.

Data description

The detail description of the methodology is represented in Table 1 Data file1. The ultimate clean Illumina and Pacbio reads used for genome assembly were 5.73 Gbp and 7.29 Gbp, respectively. The final genome assembly of *P. umbellatus* comprises 318 scaffolds with an estimated size of 79.74 Mb, a GC content of 50.63%, and an N50 of 969.73 Kbp (Table 1 Data file2). The repeats of *P. umbellatus* genome account for 58.62% (46.75 Mp) of the whole genome, among which long terminal repeats (LTR) were the most, accounting for 43.43%. For non-coding RNA (ncRNA), there were 488 tRNAs, 25 snRNAs and 741 miRNAs in *P. umbellatus* genome. The ribosomal RNA genes of *P. umbellatus* were composed of twenty-two 2121 bp 28 S subunit, fifty-one 15,804 bp 18 S subunit, and nine 900 bp 5.8 S subunits.

A total of 10,864 CDSs were predicted for the *P. umbellatus* genome with an average length of 1,339 bp. The cumulative length of encoded genes was 19.38 Mb, which accounted for 24.3% of the whole genome. The average exon numbers per gene was 5.79. Our genome assembly contains 96.2% (920/956) complete BUSCO genes, being 811 complete and single-copy BUSCOs, 109 complete and duplicated BUSCOs, 7 fragmented genes, and 29 missing genes, that indicating our high quality in terms of completeness [12].

Total 94.4% (10,258/10,864) of the protein-coding genes matched to public databases in similarity searches. Among them, 6,062 genes could be annotated by all these databases, and 5,340 predicted proteins have corresponding GO term annotations while the number for KEGG pathway is 6,955 (Table 1 Data file3). *P. umbellatus* has 6,780 genes that were assigned to the NOG categories (Table 1 Data file3), the majority of which belong to the "Posttranslational modification, protein turnover, chaperones" (692, 10.21%), "Carbohydrate transport and metabolism (512, 7.55%)" and "Intracellular trafficking, secretion, and vesicular transport (502, 7.40%)".

A total of 330 candidate CAZymes were identified in the genome of *P. umbellatus*, which included 152 glycoside hydrolases (GHs), 86 auxiliary activities enzymes (AAs), 41 carbohydrate-binding modules (CBMs), 33 glycosyl transferases (GTs), 11 carbohydrate esterases (CEs), and 7 polysaccharide lyases (PLs) (Table 1 Data file3). Among the 152 GHs, there were just 24 enzymes for the digestion of plant cell wall polysaccharides (6 cellulose, 6 hemicelluloses and 12 pectin).

According to antiSMASH investigations, *P. umbellatus* possesses at least 21 gene clusters involved in biosynthesis, including ten terpene gene clusters, four NRPS-like gene clusters, five T1PKS gene clusters, one fungal-RiPP-like gene cluster, and one other gene cluster (Table 1 Data file3).

In conclusion, the complete sequencing, assembly, and annotation of *P. umbellatus* have provided valuable new insights into the genomic structure, genetics, evolution, and phylogeny related to sclerotium development. The data are pivotal for the domestication, cultivation, production, and evaluation of the medicinal properties of this traditional medicine.

Limitations

The primary limitation is that the *P. umbellatus* assembly reported in this study is not completeness, lacking chromosome-level informaiton. This restriction limits

the subsequent biological insights we can derive from the data. To address this limitation, future research should incorporate Nanopore sequencing technology, which has the potential to significantly enhance the completeness and accuracy of the genome assembly by generating longer reads that can span complex genomic regions.

Abbreviations

bp	Base pair
DNA	Deoxyribonucleic acid
BUSCO	Benchmarking Universal SingleCopy Orthologs
KEGG	Kyoto Encyclopedia of Genes and Genomes
GO	Gene Ontology
NOG	Non-supervised Orthologous Groups
CAZymes	Carbohydrate-active enzymes
NGDC	National Genomics Data Center

Authors' contributions

YY designed the experiments; PH, LC performed the experiments and PH wrote the original draft paper; PH, ZH, TL analyzed the data; HJ supervised the study; YY supervised the study and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding

This research was supported by the Key Project at Central Government Level for the Ability Establishment of Sustainable Use for Valuable Chinese Medicine Resources (No.2060302), the National Science Fund for Distinguished Young Scholars of China (No. 82325049), the Scientific and Technological Innovation Project of China, Academy of Chinese Medical Science (CI2023D001/CI2023E002-04), the Fundamental Research Funds for the Central public welfare research institutes (No.ZZ16-ND-12/No.ZZ17-ND-12/No.ZZ18-ND-12), Scientific Research Foundation for the introduction of talent (QD-2024007B).

Data availability

The genomic data of Polyporus umbellatus for this study has been deposited at the National Genomics Data Center (NGDC) with accession number: GWHBQTP00000000(https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA008746).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 22 November 2024 / Accepted: 10 March 2025
Published online: 17 March 2025

References

1. Chinese Pharmacopoeia Commission. Pharmacopoeia of the People's Republic of China, vol. 1. Beijing: China Medical Science Press; 2020. p. 331–2.
2. Xing YM, Yin WQ, Liu MM, Wang CL, Guo SX. Oxalic acid and sclerotial differentiation of *Polyporus umbellatus*. Sci Rep. 2015;5:10759.
3. Li H, Yan Z, Xiong Q, Chen X, Lin Y, Xu Y, et al. Renoprotective effect and mechanism of polysaccharide from *Polyporus umbellatus* sclerotia on renal fibrosis. Carbohydr Polym. 2019;212:1–10.
4. Liu GK, Yang TX, Wang JR. Polysaccharides from *Polyporus umbellatus*: A review on their extraction, modification, structure, and bioactivities. Int J Biol Macromol. 2021;189:124–34.
5. Han P, Hua Z, Zhao Y, Huang L, Yuan Y. PuCRZ1, an C2H2 transcription factor from *Polyporus umbellatus*, positively regulates mycelium response to osmotic stress. Front Microbiol. 2023;14:1131605.
6. Li B, Tian X, Wang C, Zeng X, Xing Y, Ling H, et al. SWATH label-free proteomics analyses revealed the roles of oxidative stress and antioxidant defensing system in sclerotia formation of *Polyporus umbellatus*. Sci Rep. 2017;7:41283.
7. Xing YM, Zhang LC, Liang HQ, Lv J, Song C, Guo SX, et al. Sclerotial formation of *Polyporus umbellatus* by low temperature treatment under artificial conditions. PLoS One. 2013;8(2):e56190.
8. Liu T, Cheng R, Hua Z, Gao H, Wang C, Li H. et al. Identification of Growth-Promoting Bacterial Resources by Investigating the Microbial Community Composition of *Polyporus umbellatus* Sclerotia. J Fungi (Basel). 2024;10(6).
9. Kikuchi G, Yamaji H. Identification of *Armillaria* species associated with *Polyporus umbellatus* using ITS sequences of nuclear ribosomal DNA. Mycoscience. 2010;51:366–72.
10. Cao S, Yang Y, Bi G, Nelson D, Hu S, Makunga NP. et al. Genomic and Transcriptomic Insight of Giant Sclerotium Formation of Wood-Decay Fungi. Front Microbiol. 2021;12:746121.
11. Zhao YY. Traditional uses, phytochemistry, pharmacology, pharmacokinetics and quality control of *Polyporus umbellatus* (Pers.) Fries: a review. J Ethnopharmacol. 2013;149(1):35–48.
12. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, EM Z. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210–3212.
13. Detailed methodology. zenodo dataset. 2024. https://doi.org/10.5281/zenodo.14192471.
14. Genomic features of the *P. umbellatus*. zenodo dataset. 2024. https://doi.org/10.5281/zenodo.14190274.
15. Functional annotation of the *P. umbellatus*. zenodo dataset. 2024. https://doi.org/10.5281/zenodo.14192681.
16. Genome assembly of *P. umbellatus*. NGDC. 2023:GWHBQTP00000000. https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA008746.
17. Predicted genes of *P. umbellatus*. NGDC. 2023:GWHBQTP00000000. https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA008746.
18. Illumina and Pacbio sequencing reads of *P. umbellatus*. NGDC dataset. 2025. https://ngdc.cncb.ac.cn/gsa/browse/CRA022677/CRX1492289.
19. Illumina sequencing reads of *P. umbellatus*. NGDC dataset. 2025. https://ngdc.cncb.ac.cn/gsa/browse/CRA022677/CRX1492288.

Publisher's note