BMC Genomic Data

DATA NOTE

Open Access

Genome sequence resources for three strains of the genus *Clonostachys*

Zongping Sun[1†], Fan Zhang[2,3†], Nana Zhong[1], Kang Zhou[1*] and Jun Tang[1*]

Abstract

Objective *Clonostachys*, a genus with rich morphological and ecological diversity in Bionectriaceae, has a wide distribution among diverse habitats. Several studies have reported *Clonostachys* fungi as effective biological agents against multiple fungal plant pathogens. To clarify the diversity and biocontrol mechanisms of the *Clonostachys* fungi, this study was undertaken to sequence and assemble the genomes of two *C. chloroleuca* and one *C. rhizophaga*.

Data description Here, we performed genomic sequencing of three strains of genus *Clonostachys* collected from the China General Microbiological Culture Collection Center (CGMCC) using Illumina HiSeq 2500 sequencing technology. Whole genome analysis indicated that their genomes consist of 58,484,224 bp with a GC content of 48.58%, 58,114,960 bp with a GC content of 47.74% and 58,450,453 bp with a GC content of 48.58%, respectively. BUSCO analysis of the genome assembly indicated that the completeness of these genomes was at least 98%. In summary, these datasets provide a valuable resource for ongoing studies that include further exploration of biological function, marker development, enhanced biological control ability of *Clonostachys* fungi, and population diversity.

Keywords Fungal genome, *Clonostachys*, Biocontrol

Objective

Most studies related to the genus *Clonostachys* have been focused on its powerful capacity as a biocontrol agent, especially of fungal pathogens. For example, *Clonostachys rosea* and *C. chloroleuca* are well-known destructive mycoparasites and can effectively control various plant diseases, caused by *Fusarium* species, *Sclerotinia sclerotiorum* and *Botrytis cinerea* [1–3]. Additionally, *Clonostachys* fungi are well known to produce a variety of secondary metabolites with various biological activities to show their pharmaceutical and agrochemical applications. Many of these compounds exhibit biological activities, such as cytotoxic, antimicrobial, antileishmanial, and antimalarial activities [4–6]. In the past decades, a zearalenone hydrolase encoded by *C. rosea* was discovered, and the zearalenone detoxification ability was proved to be important for the biocontrol of *Fusarium graminearum* [7–10].

It is remarkable as well that *C. rosea* is an important endophyte organism that, besides providing benefits to a wide range of host plants, can successfully mimic their chemical behavior [11, 12]. Some studies have shown the potential of *C. rosea* strains to promote the growth and health of diverse crops, such as tomato (*Lycopersicon esculentum* L.) [13], cucumber (*Cucumis sativus* L.) [14], wheat (*Triticum durum* Desf.) [15], pine (*Pinus radiata* D. Don) [16], and oil palm (*Elaeis guineensis* Jacq.) [17]. *C. rosea* was also classified as 'plant-growth-promoting

[†]Zongping Sun and Fan Zhang contributed equally to this work.

*Correspondence:
Kang Zhou
zhoukang19950601@163.com
Jun Tang
tj751@163.com
[1] Anhui Province Key Laboratory of Pollution Damage and Biological Control for Huaihe River Basin, Fuyang Normal University, Fuyang 236037, China
[2] State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, Wuhan 430070, China
[3] Hubei Key Laboratory of Plant Pathology, Huazhong Agricultural University, Wuhan 430070, China

fungi' (PGPF) [18]. Therefore, in-depth understanding of the molecular mechanisms underlying biocontrol of *Clonostachys* fungi will be conducive to developing sustainable strategies for management of fungal disease. In this regard, high-quality complete genome resources for *Clonostachys* fungi will be helpful.

Data description

In this study, we sequenced the genomes of three strains belonging to the genus *Clonostachys*, which were collected from the China General Microbiological Culture Collection Center (CGMCC). These strains were cultured on potato dextrose agar (PDA) for five days at 22 °C in the dark. Mycelia growing on the cellophane were harvested by scraping the plates with a flame-sterilized metal spatula, frozen in liquid nitrogen for 20 s, and stored at −80 °C until use. The genomic DNA of strains was extracted using the cetyltrimethylammonium bromide (CTAB) method [19]. Quality control for genomic DNA (gDNA) was performed by measuring the absorbance at ratios 260/280 and 260/230 using a NanoDrop ND-2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

DNA samples were sent to Shanghai Biotechnology Corporation (Shanghai, China) for microbial whole genome sequencing. Three genome paired-end libraries were prepared for Illumina HiSeq 2500 System sequencing, yielding 27,300,124 short read pairs for strain CGMCC3.3655, 30,040,177 for strain CGMCC3.3657, and 33,506,775 for strain CGMCC3.4252 (Table 1 Data file 1–3). Fastp v0.20.0 was used to trim and filter the Illumina reads, then fastqc v0.11.9 and multiqc v1.8 were used to evaluate the reads' quality [20–22]. After quality control, the short reads were used for *de novo* assembly by SPAdes v3.6.1 [23]. Potential mitochondrial genes and other possible contaminants from the genomes were filtered out using the Barrnap v0.9 [24]. The remaining contigs were filtered at 1,000 bp, and the quality of the assembly was visualized with QUAST v5.0.2 [25]. The resultant draft genomes were all >58 MB in size. GC content was similar across all three genomes at 48%. However, the largest contig sizes varied with strain CGMCC3.3657 having the largest contig at 3,565,857 bp. Additionally, strain CGMCC3.3655 had the highest overall genome length at 58.5 Mb (Table 1 Data file 4–7). Transposable elements (TE) were identified by Tandem Repeat Finder (v4.04) [26] and The Extensive *de novo* TE Annotator (EDTA) [27]. The analysis revealed that 4.25% of the CGMCC3.3655 genome (LTR/Gypsy type: 1.5 Mb), 8.34% of the CGMCC3657 genome (LTR/Gypsy type: 1.9 Mb) and 3.99% of the CGMCC3.4252 (LTR/Gypsy type: 1.3 Mb) were composed of repetitive DNA (Table 1 Data file 7).

Genome annotation was performed using the BRAKER v2.1.5 pipeline [28] based on GeneMark-ES version 4.68 [29] and Augustus v3.3.3 [30]. 17,770 protein-coding genes for strain CGMCC3.3655, 17,186 for strain CGMCC3.3657, and 17,779 for strain CGMCC3.4252

Table 1 Overview of all 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	Sequencing read dataset of CGMCC3.3655	Fastq file (fastq)	NCBI Sequence Read Archive (https://identifiers.org/ncbi/insdc.sra: SRR30906288) [37]
Data file 2	Sequencing read dataset of CGMCC3.3657	Fastq file (fastq)	NCBI Sequence Read Archive (https://identifiers.org/ncbi/insdc.sra: SRR30906289) [38]
Data file 3	Sequencing read dataset of CGMCC3.4252	Fastq file (fastq)	NCBI Sequence Read Archive (https://identifiers.org/ncbi/insdc.sra: SRR30906290) [39]
Data file 4	Genome assembly of CGMCC3.3655	genbank format (.gbk)	NCBI GenBank (https://identifiers.org/ncbi/nucleotide: JBICSD000000000.1) [40]
Data file 5	Genome assembly of CGMCC3.3657	genbank format (.gbk)	NCBI GenBank (https://identifiers.org/ncbi/nucleotide: JBICSC000000000.1) [41]
Data file 6	Genome assembly of CGMCC3.4252	genbank format (.gbk)	NCBI GenBank (https://identifiers.org/ncbi/nucleotide: JBICSB000000000.1) [42]
Data file 7	Genome statistics and phylogeny of three strains of the genus *Clonostachys*	MS docx file (docx)/ Portable Data Format file (pdf)	Figshare (https://doi.org/10.6084/m9.figshare.27229602.v1) [43]
Data set 8	CGMCC3.3655 genomic transcript, protein sequences and functional annotations	FASTA/Text	Figshare (https://doi.org/10.6084/m9.figshare.27229635.v1) [44]
Data set 9	CGMCC3.3657 genomic transcript, protein sequences and functional annotations	FASTA/Text	Figshare (https://doi.org/10.6084/m9.figshare.27237435.v1) [45]
Data set 10	CGMCC3.4252 genomic transcript, protein sequences and functional annotations	FASTA/Text	Figshare (https://doi.org/10.6084/m9.figshare.27237438.v1) [46]

were predicted, respectively. To gain a functional gene annotation, we annotated whole-genome protein-encoding genes. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotation of protein-coding genes was mainly performed by KAAS-KEGG Automatic Annotation Server (KEGG's KAAS) (http://www.genome.jp/tools/kaas/) [31]. The Gene Ontology (GO) annotation of protein-coding genes was done using InterPro (v5.27) software [32]. The carbohydrate active enzymes (CAZymes) analysis was compared and annotated with the dbCAN database (https://bcb.unl.edu/dbCAN2/) by HMMER software (v3.3.2) [33]. The databases were pfam and Clusters Orthologous Groups (COG), which were compared using InterPro (version 5.27) [32] and diamond (v0.8.22) software [34], respectively. We identified a total of 11,835 proteins with Pfam domains, 3,950 genes with GO items, 5,051 genes involved in different KEGG pathways, 12,007 COG genes, and 237 CAZymes in the genome of strain CGMCC3.3655. There was also a total of 11,412 proteins with Pfam domains, 3,880 genes with GO items, 4,933 genes involved in different KEGG pathways, 11,600 COG genes, and 224 CAZymes in the genome of strain CGMCC3.3657. For strain CGMCC3.4252, we identified 11,851 proteins with Pfam domains, 3,960 genes with GO items, 5,057 genes involved in different KEGG pathways, 12,042 COG genes, and 236 CAZymes (Table 1 Data file 8–10). Their genome annotation completeness's were estimated using benchmarking universal single-copy orthologs (BUSCO v4.1.4) with the fungi dataset [35], identifying 98.3 to 98.5% of the fungal orthologs (Table 1). With 1,000 bootstrap replicates, the phylogenetic tree was built in Molecular Evolutionary Genetics Analysis (MEGA) X software (v10.1.7) utilizing the neighbor-joining (NJ) method [36]. The phylogenetic analysis result based on the ATP citrate lyase (*acl1*) and the largest subunit of RNA polymerase II (*rbp1*) sequences showed that CGMCC3.3655, CGMCC3.4252 and other *C. chloroleuca* strains were grouped into one large branch with good support. But CGMCC3.3657 was clustered with members of *C. rhizophaga* (Table 1 Data file 7).

These genome sequences could contribute to the understanding of genetic and genomic diversities of *Clonostachys* fungi. It could also provide opportunities to analyze the molecular basis of their biocontrol activities.

Limitation

This data note was limited to the description of genomes of three *Clonostachys* strains. A larger collection is needed to help us better understand their genetic and biological characteristics.

Acknowledgements
Not applicable.

Authors' contributions
Z. S. and J. T. conceived and designed the research. K. Z., N., Z., and F. Z. analyzed the data. K. Z., Z. S. and J. T. wrote or revised the paper. All authors read and approved the final manuscript.

Funding
This research was supported by the Key Natural Science Research Projects in Anhui Universities (No. 2022AH051346; No. KJ2021A0676); Science and Technology Innovation Capability Enhancement Projects in Fuyang National Agricultural Sci-Tech Park (No. KJTS2022002; No. KJTS2022003). Natural Science Foundation of Universities of Anhui Province for Distinguished Young Project (No. 2022AH020081); The Fuyang Normal University Research Project (No. 2020KYQD0023). Biological and Medical Sciences of Applied Summit Nurturing Disciplines in Anhui Province, Anhui Education Secretary Department [2023]13.

Data availability
The complete whole-genome sequences of three Clonostachys strains have been deposited at the National Center for Biotechnology Information (NCBI) GenBank with the accession number JBICSC000000000.1 (CGMCC3.3657), JBICSB000000000.1 (CGMCC3.4252) and JBICSD000000000.1 (CGMCC3.3655) (BioProject: PRJNA1170043; BioSample: SAMN44090158, SAMN44090159, and SAMN44090160), respectively.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 12 August 2024 Accepted: 10 January 2025
Published online: 24 January 2025

References

1. Cota LV, Maffia LA, Mizubuti ESG, Macedo PEF. Biological control by *Clonostachys Rosea* as a key component in the integrated management of strawberry gray mold. Biol Control. 2009;50(3):222–30.
2. Sun ZB, Sun MH, Zhou M, Li SD. Transformation of the endochitinase gene Chi67-1 in *Clonostachys Rosea* 67 – 1 increases its biocontrol activity against *Sclerotinia Sclerotiorum*. AMB Express. 2017;7(1):1.
3. de Souza DS, Barth AI, Berté ALW, Bizarro GL, Heidrich D, da Silva GL, Johann L, Maciel MJ. Evaluation of the activity of filamentous fungi isolated from soils of the Pampa biome applied in the biological control of *Tetranychus urticae* (Acari: *Tetranychidae*) and *Polyphagotarsonemus Latus* (Acari: *Tarsonemidae*). Exp Appl Acarol. 2021;85(1):19–30.
4. Han P, Zhang X, Xu D, Zhang B, Lai D, Zhou L. Metabolites from *Clonostachys* Fungi and their biological activities. J Fungi. 2020;6(4):229.
5. Sun ZB, Li SD, Ren Q, Xu JL, Lu X, Sun MH. Biology and applications of *Clonostachys Rosea*. J Appl Microbiol. 2020;129(3):486–95.
6. Zhai MM, Qi FM, Li J, Jiang CX, Hou Y, Shi YP, Di DL, Zhang JW, Wu QX. Isolation of secondary metabolites from the soil-derived Fungus *Clonostachys Rosea* YRS-06, a Biological Control Agent, and evaluation of antibacterial activity. J Agric Food Chem. 2016;64(11):2298–306.
7. El-Sharkawy S, Abul-Hajj YJ. Microbial cleavage of zearalenone. Xenobiotica. 1988;18(4):365–71.
8. Kakeya H, Takahashi-Ando N, Kimura M, Onose R, Yamaguchi I, Osada H. Biotransformation of the mycotoxin, zearalenone, to a non-estrogenic compound by a fungal strain of *Clonostachys* Sp. Biosci Biotechnol Biochem. 2002;66(12):2723–6.
9. Kosawang C, Karlsson M, Vélëz H, Rasmussen PH, Collinge DB, Jensen B, Jensen DF. Zearalenone detoxification by zearalenone hydrolase is

important for the antagonistic ability of *Clonostachys rosea* against mycotoxigenic *Fusarium graminearum*. Fungal Biol. 2014;118(4):364–73.
10. Sun Z, Fang Y, Zhu Y, Tian W, Yu J, Tang J. Biotransformation of zearalenone to non-estrogenic compounds with two novel recombinant lactonases from *Gliocladium*. BMC Microbiol. 2024;24(1):75.
11. Chatterton S, Punja ZK. Factors influencing colonization of cucumber roots by *Clonostachys Rosea* f. *catenulata*, a biological disease control agent. Biocontrol Sci Technol. 2010;20:37–55.
12. Chatterton S, Punja ZK. Colonization of geranium foliage by *Clonostachys Rosea* f. catenulata, a biological control agent of Botrytis grey mould. Botany-botanique. 2012;90:1–10.
13. Ravnskov S, Jensen B, Knudsen IMB, Bodker L, Jensen DF, Karliński L, Larsen L. Soil inoculation with the biocontrol agent *Clonostachys Rosea* and the mycorrhizal fungus *Glomus intraradices* results in mutual inhibition, plant growth promotion and alteration of soil microbial communities. Soil Biol Biochem. 2006;38:3453–62.
14. Sutton JC, Liu W, Ma J, Brown GW, Stewart FJ. Evaluation of the fungal endophyte *Clonostachys Rosea* as an inoculant to enhance growth, fitness and productivity of crop plants. Acta Hortic. 2008;782:279–86.
15. Roberti R, Veronesi AR, Cesari A, Cascone A, Berardino ID, Bertini L, Caruso C. Induction of PR proteins and resistance by the biocontrol agent *Clonostachys Rosea* in wheat plants infected with *Fusarium Culmorum*. Plant Sci. 2008;175:339–47.
16. Moraga-Suazo P, Sanfuentes E. Growth promotion of *Pinus radiata* seedlings by soil inoculation and seed pretreatment with the biological control agent *Clonostachys Rosea*. Gayana Bot. 2017;74:140–6.
17. Goh YK, Marzuki NF, Tuan Pa TNF, Goh TK, Kee ZS, Goh YK, Yusof MT, Vujanovic V, Goh KJ. Biocontrol and plant-growth-promoting traits of *Talaromyces Apiculatus* and *Clonostachys Rosea* consortium against ganoderma basal stem rot disease of oil palm. Microorganisms. 2020;8:1138.
18. Adedayo AA, Babalola OO. Fungi that promote plant growth in the rhizosphere boost crop growth. J Fungi. 2023;9(2):239.
19. 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;20:6115–6.
20. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:i884–90.
21. Andrews S. FastQC: A quality control tool for high throughput sequence data. 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
22. Ewels P, Magnusson M, Lundin S, Kaller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32:3047–8.
23. Nurk S, Meleshko D, Korobeynikov A, Pevzner A. meta-SPAdes: a new versatile metagenomic assembler. Genome Res. 2017;27:824–34.
24. Seemann T. Barrnap 0.9: Rapid ribosomal RNA prediction. 2018. https://github.com/tseemann/barrnap.
25. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29:1072–5.
26. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27:573–80.
27. Su W, Ou S, Hufford MB, Peterson T. A tutorial of EDTA: extensive de novo TE annotator. Methods Mol Biol. 2021;2250:55–67.
28. Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. 2019. Whole-genome annotation with BRAKER. Methods Mol. Biol. 1962:65–95.
29. Borodovsky M, Lomsadze A. Eukaryotic gene prediction using GeneMark. hmm-E and GeneMark-ES. Curr Protoc Bioinf. 2011;4:461–4610.
30. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. Bioinformatics. 2008;24:637–44.
31. Moriya Y, Itoh M, Okuda S, Yoshizawa A, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 2007;35:182–5.
32. McDowall J, Hunter S. InterPro protein classification. Methods Mol Biol. 2011;694:37–47.
33. Zhang Z, Wood WI. A profile hidden Markov model for signal peptides generated by HMMER. Bioinformatics. 2003;19(2):307–8.
34. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12:59–60.
35. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol. 2018;35(3):543–8.
36. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35(6):1547–9.
37. Sun ZP, et al. Sequencing read dataset of CGMCC3.3655, NCBI Sequence Read Archive, Data file 1; 2024.
38. Sun ZP, et al. J. Sequencing read dataset of CGMCC3.3657, NCBI Sequence Read Archive, Data file 2; 2024.
39. Sun ZP, et al. J. Sequencing read dataset of CGMCC3.4252, NCBI Sequence Read Archive, Data file 3; 2024.
40. Sun ZP, et al. Genome assembly of CGMCC3.3655, NCBI GenBank, Data file 4; 2024.
41. Sun ZP, et al. Genome assembly of CGMCC3.3657, NCBI GenBank, Data file 5; 2024.
42. Sun ZP, et al. Genome assembly of CGMCC3.4252, NCBI GenBank, Data file 6; 2024.
43. Sun ZP, et al. Genome statistics and phylogeny of three strains of the genus Clonostachys, Figshare, Data file 7; 2024.
44. Sun ZP, et al. CGMCC3.3655 genomic transcript, protein sequences and functional annotations, Figshare, Data set 8; 2024.
45. Sun ZP, et al. CGMCC3.3657 genomic transcript, protein sequences and functional annotations, Figshare, Data set 9; 2024.
46. Sun ZP, et al. CGMCC3.4252 genomic transcript, protein sequences and functional annotations, Figshare, Data set 10; 2024.

Publisher's Note