Data in brief 25 (2019) 104295

Contents lists available at ScienceDirect

Data in brief

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

Data Article

Data on the genome analysis of the wild edible mushroom, *Russula griseocarnosa*

Fei Yu, Junfeng Liang*

Key Laboratory of State Forestry Administration on Tropical Forestry Research, Research Institute of Tropical Forestry, Chinese Academy of Forestry, Guangzhou 510520, China

ARTICLE INFO

Article history:
Received 18 April 2019
Received in revised form 21 May 2019
Accepted 11 July 2019
Available online 19 July 2019

Keywords:
Russula griseocarnosa
Ectomycorrhizal fungus
Whole genome
Genome annotation

ABSTRACT

In the present article, we report data on the whole genome sequence of a wild edible and medicinal ectomycorrhizal fungus *Russula griseocarnosa*. The *R. griseocarnosa* genome consists of 64.81 Mb with a GC-pair content of 49.41%. The genome assembly consists of 471 scaffolds and 16128 coding protein genes. The coding protein genes was annotated in different databases (GO, KEGG and CAZys), respectively. The whole genome sequence and functional annotation provide important information for ectomycorrhizal fungus, which can be used as a basis for cultivation and breeding of *R. griseocarnosa*. The Whole Genome project of *Russula griseocarnosa* has been deposited at DDBJ/ENA/GenBank under the accession RMVF00000000. The version described is RMVF01000000. To further interpretation of the data provided in this article, please refer to the research article 'Whole genome sequencing and genome annotation of the wild edible mushroom, *Russula griseocarnosa*' [1].

DOI of original article: https://doi.org/10.1016/j.ygeno.2019.04.012.
* Corresponding author.
E-mail address: jfliang2000@163.com (J. Liang).

https://doi.org/10.1016/j.dib.2019.104295

Specifications Table

Subject area	Biology
More specific subject area	Microbiology, Genomics
Type of data	Table, figures
How data was acquired	PacBio RS and Illumina Hiseq X-Ten sequencing
Data format	Annotated and comparative analyzed
Experimental factors	The fruiting body samples were obtained and quickly frozen in liquid nitrogen before stored in a −80 °C freezer. Total DNA of fruiting body was extracted immediately.
Experimental features	DNA Sequencing was performed by using PacBio RS and Illumina Hiseq X-Ten, genome assembly, annotation and analysis were carried out.
Data source location	The fruiting bodies of *Russula griseocarnosa* were collected from Linjing Town, Teng County, Guangxi Province, China (2 Jun. 2017) (23.15 N, 110.66 E)
Data accessibility	The whole genome sequence of *Russula griseocarnosa* has been deposited at DDBJ/ENA/GenBank under the accession RMVF00000000. The version described is RMVF01000000. The BioSample, BioProject and SRA accession number are SAMN09602224, PRJNA479704 and SRP153002, respectively.
Related research article	F. Yu, J. Song, J.F. Liang, S.K. Wang, J.K. Lu, Whole genome sequencing and genome annotation of the wild edible mushroom, *Russula griseocarnosa*. Genomics. (2019) in press [1] https://doi:10.1016/j.ygeno.2019.04.012.

Value of the data

- The first genome under the genus *Russula* to be reported.
- The data provide valuable information of the potential function and gene expression mechanisms about ectomycorrhizal fungus *Russula griseocarnosa.*
- The CAZymes of *Russula griseocarnosa* confirms the adaptation to symbiosis, and reveals the strategy for host interaction.

1. Data

Russula griseocarnosa (Fig. 1) is a wild edible and medicinal ectomycorrhizal fungus that is native to southern China. The resulting draft genome of *R. griseocarnosa* present the 64.81 Mb in size with a G+C content of 49.41%. The genome sequence was assembly with 471 scaffolds and 16128 coding protein genes [1]. The data illustrated in Fig. 2 show the Gene Ontology (GO) distribution of the protein coding genes and Fig. 3 gives a complete overview of the KEGG pathway. According comparative analysis, The GO annotations of *Russula griseocarnosa* genes were similar with *Agaricus bisporus* [2] in “Localization”, “Biological regulation”, and “Regulation of biological process”, and fewer numbers than that of *Laccaria bicolor* [1,3]. Compared with KEGG metabolic annotations, the most genes of *Russula griseocarnosa* pathways was not significantly in *Laccaria bicolor* and *Agaricus bisporus*, but *R. griseocarnosa* had less genes in "Tryptophan metabolism" and "Starch and sucrose metabolism" pathways [1].

The CAZymes coding genes of *R. griseocarnosa* encode enzymes involved in the degradation of plant cell wall polysaccharides, non-plant polysaccharides (for example, animal and bacterial polysaccharides) and fungal cell wall (Fig. 4). The CAZymes coding genes of *R. griseocarnosa* was similar to the symbiotic fungal species *Scleroderma citrinum* [4] in non-plant polysaccharides degradation and fungal cell wall degradation, and higer number of plant cell wall polysaccharides degradation. The plant cell wall polysaccharides degradation associated with cellulose degrading enzymes (GH6, GH7, GH44 and GH45), hemicellulose-degrading enzymes (GH10, GH11 and GH115) and pectin-degrading enzymes (GH43, GH51, GH78, GH93, PL1, PL3, and PL4) were absent in *Russula griseocarnosa*, *Laccaria bicolor*, and *Scleroderma citrinum* genomes [1].

2. Experimental design, materials and methods

2.1. Fungal material

Fruiting bodies of *R. griseocarnosa* were collected from Linjing Town, Teng County, Guangxi Province, China in 2017. The fruiting body samples was frozen in liquid nitrogen and stored at −80 °C freezer until DNA extract.

Fig. 1. Fruiting bodies of *Russula griseocarnosa*.

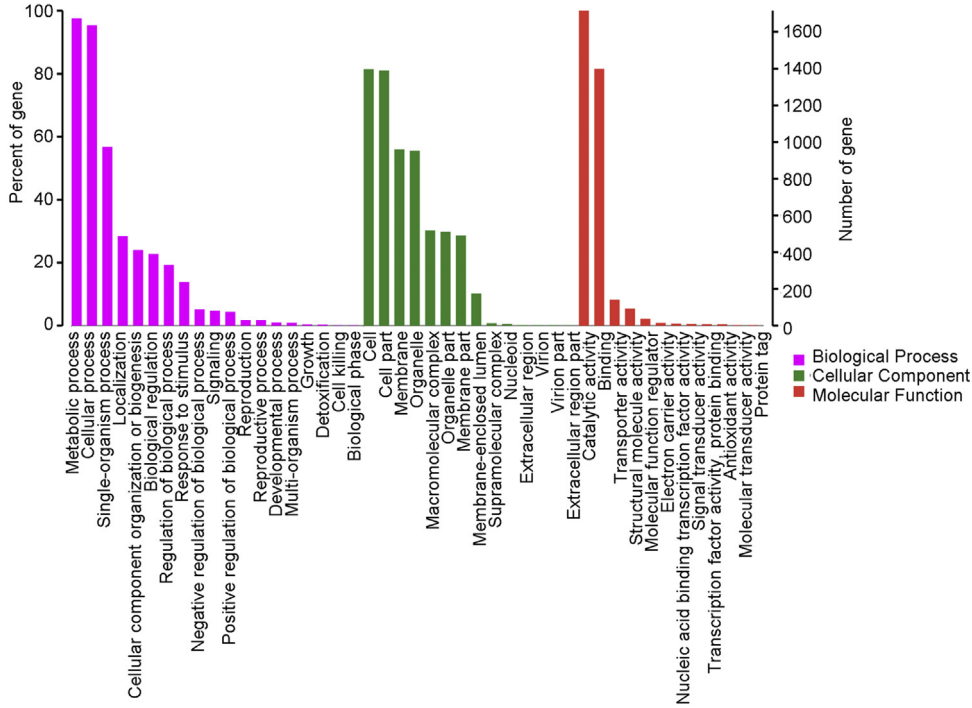

Fig. 2. The Gene Ontology (GO) function annotation of *Russula griseocarnosa*.

2.2. DNA extraction and sequencing

Genomic DNA was extracted using the Omega Fungal DNA Kit D3390-02. Quality of DNA was determined using TBS-380 fluorometer (Turner BioSystems Inc., Sunnyvale, CA). The concentration of at least 20 mg/L (OD260/280 = 1.8–2.0).

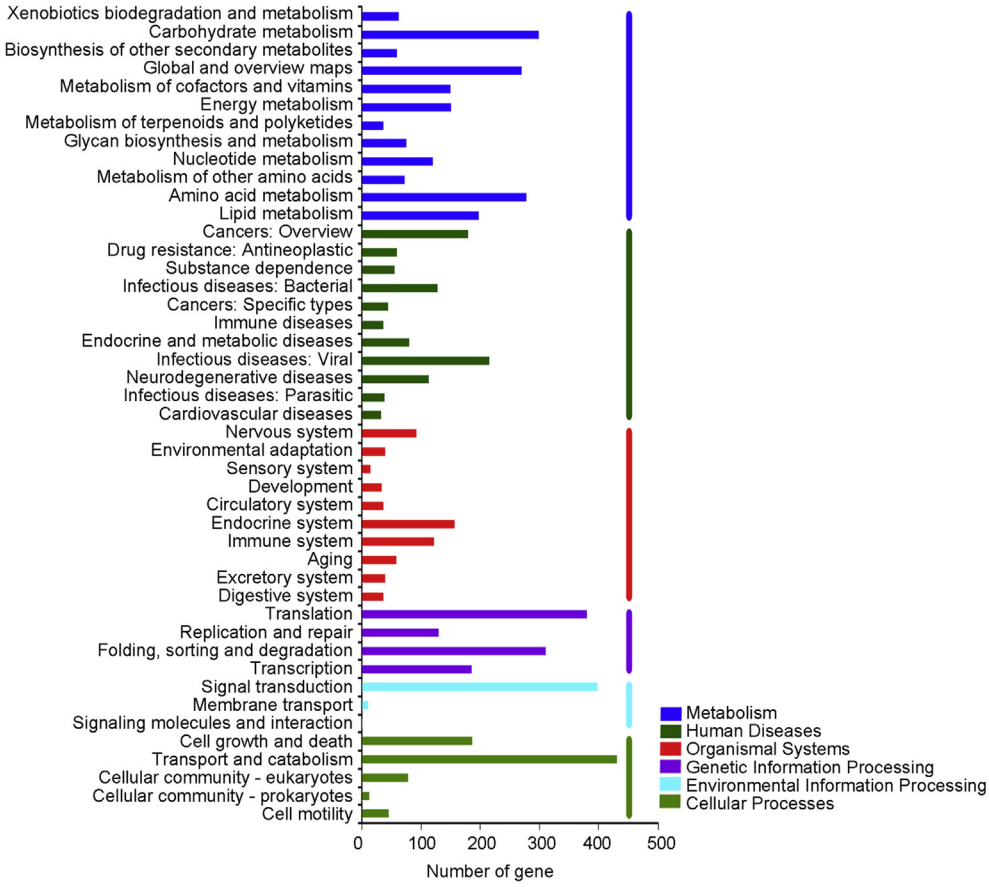

Fig. 3. The KEGG function annotation of *Russula griseocarnosa*.

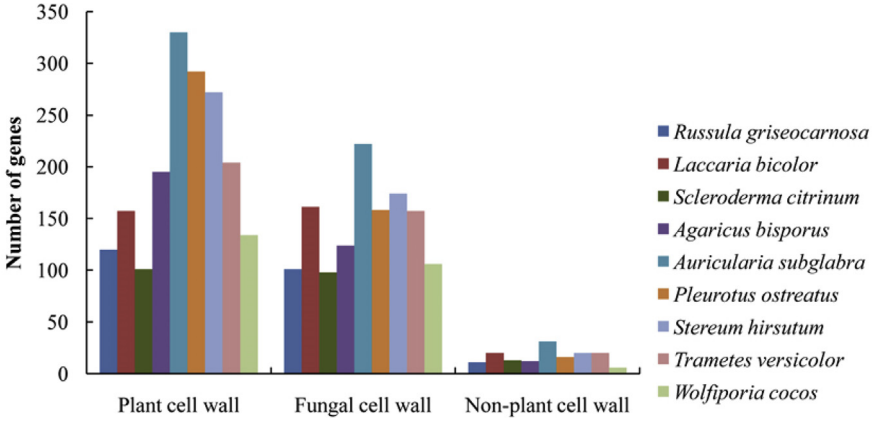

Fig. 4. Comparison of CAZys associated with cell wall degradation.

R. griseocarnosa genome was sequenced using Illumina HiSeq X-ten sequencing and PacBio RS sequencing at Shanghai Majorbio Bio-pharm Biotechnology Co., Ltd, China. Paired-end libraries with 300 bp inserts were constructed in Illumina HiSeq X-ten sequencing. 8-10k insert shotgun libraries were generated in Pacific Biosciences RS sequencing.

2.3. Genome assembly and annotation

The genome sequence was assembled as follows: (1) PacBio long reads were corrected and assembled by Canu (v1.7) [5]; (2) Illumina reads corrected and used for scaffolding by SOAPdenovo (v2.04). Fill the gaps using GapCloser (v1.12) package; and (3) PacBio reads were modified based on Illumina reads. The final assembly produced a circular genome sequence without gaps.

Protein coding sequences were predicted using the automated pipeline MAKER2 (v2.31.9) [6]. It combining data for mRNAs, proteins, the ab initio predictions of SNAP [7] and GeneMark-ES (v2.3a) [8].

The predicted protein coding sequences was annotated in Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) database using Blastp (v2.3.0). The Carbohydrate-active enzymes (CAZymes) were performed using blastp (cut off e-value$\leq$1e−5) at http://www.cazy.org/ [9].

Acknowledgements

This study was supported by the Science and Technology Project of Guangdong Province (2017B020205002), and the National Natural Science Foundation of China (No. 31770657 and 31570544). The authors thank Dr Hong Luo from the Kunming Institute of Botany, Chinese Academy of Sciences for research guidance.

Conflict of 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] F. Yu, J. Song, J.F. Liang, S.K. Wang, J.K. Lu, Whole genome sequencing and genome annotation of the wild edible mushroom, *Russula griseocarnosa*, Genomics (2019). https://doi:10.1016/j.ygeno.2019.04.012.

[2] E. Morin, A. Kohler, A.R. Baker, M. Foulongne-Oriol, V. Lombard, L.G. Nagy, R.A. Ohm, A. Patyshakuliyeva, A. Brun, A.L. Aerts, A.M. Bailey, C. Billette, P.M. Coutinho, G. Deakin, H. Doddapaneni, D. Floudas, J. Grimwood, K. Hildén, U. Kües, K.M. LaButti, A. Lapidus, E.A. Lindquist, S.M. Lucas, C. Murat, R.W. Riley, A.A. Salamov, J. Schmutz, V. Subramanian, H.A.B. Wösten, J. Xu, D. C. Eastwood, G.D. Foster, A.S.M. Sonnenberg, D. Cullen, R.P. De Vries, T. Lundell, D.S. Hibbett, B. Henrissat, K.S. Burton, R.W. Kerrigan, M.P. Challen, I.V. Grigoriev, F. Martin, Genome sequence of the button mushroom *Agaricus bisporus* reveals mechanisms governing adaptation to a humic-rich ecological niche, Proc. Natl. Acad. Sci. U. S. A 109 (2012) 17501−17506. https://doi.org/10.1073/pnas.1206847109.

[3] J. Labbé, X. Zhang, T. Yin, J. Schmutz, J. Grimwood, F. Martin, G.A. Tuskan, F.L. Tacon, A genetic linkage map for the ectomycorrhizal fungus *Laccaria bicolor* and its alignment to the whole-genome sequence assemblies, New Phytol. 180 (2008) 316−328. https://doi.org/10.1111/j.1469-8137.2008.02614.x.

[4] A. Kohler, A. Kuo, L.G. Nagy, E. Morin, K.W. Barry, F. Buscot, B. Canbäck, C. Choi, N. Cichocki, A. Clum, J. Colpaert, A. Copeland, M.D. Costa, J. Doré, D. Floudas, G. Gay, M. Girlanda, B. Henrissat, S. Herrmann, J. Hess, N. Högberg, T. Johansson, H. Khouja, K. LaButti, U. Lahrmann, A. Levasseur, E.A. Lindquist, A. Lipzen, R. Marmeisse, E. Martino, C. Murat, C.Y. Ngan, U. Nehls, J.M. Plett, A. Pringle, R.A. Ohm, S. Perotto, M. Peter, R. Riley, F. Rineau, J. Ruytinx, A. Salamov, F. Shah, H. Sun, M. Tarkka, A. Tritt, C. Veneault-Fourrey, A. Zuccaro, M.G.I. Consortium, A. Tunlid, I.V. Grigoriev, D.S. Hibbett, F. Martin, Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists, Nat. Genet. 47 (2015) 410−415. https://doi.org/10.1038/ng.3223.

[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 adaptive k-mer weighting and repeat separation, Genome Res. 27 (2017) 722. https://doi.org/10.1101/gr.215087.116.

[6] C. Holt, M. Yandell, MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects, BMC Bioinf. 12 (2011) 491. https://doi.org/10.1186/1471-2105-12-491.

[7] I. Korf, Gene finding in novel genomes, BMC Bioinf. 5 (2004) 59. https://doi.org/10.1186/1471-2105-5-59.

[8] M. Borodovsky, A. Lomsadze, Eukaryotic gene prediction using genemark.hmm-E and genemark-ES, Curr. Protoc. Bioinformatics 25 (2011) 1−10. https://doi.org/10.1002/0471250953.bi0406s35.

[9] B.L. Cantarel, P.M. Coutinho, C. Rancurel, The carbohydrate-active EnZymes database (CAZy): an expert resource for glycogenomics, Nucleic Acids Res. 37 (2009) D233−D238. https://doi.org/10.1093/nar/gkn663.