Hybrid Genome Assembly and Gene Repertoire of the Root Endophyte Clitopilus hobsonii QYL-10 (Entolomataceae, Agaricales, Basidiomycetes)

Long Peng, Xiaoliang Shan, Yuchen Wang, Francis Martin, Rytas Vilgalys, Zhilin Yuan

MPMI Vol. 34, No. 6, 2021, pp. 711–714, https://doi.org/10.1094/MPMI-11-20-0328-A

Hybrid Genome Assembly and Gene Repertoire of the Root Endophyte *Clitopilus hobsonii* QYL-10 (Entolomataceae, Agaricales, Basidiomycetes)

Long Peng,[1,2] Xiaoliang Shan,[1,2,3] Yuchen Wang,[1,2,3] Francis Martin,[4] Rytas Vilgalys,[5] and Zhilin Yuan[1,2,†]

[1] State Key Laboratory of Tree Genetics and Breeding, Chinese Academy of Forestry, Beijing 100091, China
[2] Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Hangzhou 311400, China
[3] Nanjing Forestry University, Nanjing 210037, China
[4] Université de Lorraine, INRAe, UMR 1136 Interactions Arbres/Microorganismes, INRAe Grand Est-Nancy, 54280 Champenoux, France
[5] Department of Biology, Duke University, Box 90338, Durham, NC 27708, U.S.A.

Funding
This work was funded by Fundamental Research Funds for the Central Non-profit Research of Chinese Academy of Forestry (CAFYBB2020QB002) and National Natural Science Foundation of China (numbers 31901290 and 31722014).

Keywords
fungal endophytes, genomics, long-read assembly, root-fungal symbiosis, saprotrophy-biotrophy transition

Abstract

Clitopilus hobsonii (Entolomataceae, Agaricales, Basidiomycetes) is a common soil saprotroph. There is also evidence that *C. hobsonii* can act as a root endophyte benefitting tree growth. Here, we report the genome assembly of *C. hobsonii* QYL-10, isolated from ectomycorrhizal root tips of *Quercus lyrata*. The genome size is 36.93 Mb, consisting of 13 contigs (N_{50} = 3.3 Mb) with 49.2% GC content. Of them, 10 contigs approached the length of intact chromosomes, and three had telomeres at one end only. BUSCO analysis reported a completeness score of 98.4%, using Basidiomycota_odb10 lineage data. Combining *ab-initio*, RNA-seq data, and homology-based predictions, we identified 12,710 protein-coding genes. Approximately, 1.43 Mb of transposable elements (3.88% of the assembly), 36 secondary metabolite biosynthetic gene clusters, and 361 genes encoding putative carbohydrate-active enzymes were identified. This genomic resource will allow functional studies aimed to characterize the symbiotic interactions between *C. hobsonii* and its host trees and will also provide a valuable foundation for further research on comparative genomics of the Entolomataceae.

Genome Announcement

Clitopilus (Entolomataceae, Agaricales, Basidiomycota) is considered a common saprotrophic genus (Raj and Manimohan 2018). Several field investigations, however, suggested that *Clitopilus orcella* and *C. subvilis* could establish an ectomycorrhizal association with *Quercus petraea* (Southworth 2013) and *Q. robur* (Howe 1964; Keizer and Arnolds 1994) as well as other trees in the *Fagaceae* family (Molina et al. 1992). In our previous study, strain QYL-10 of *C. hobsonii* was isolated from ectomycorrhizal root tips of *Q. lyrata* seedlings grown in the greenhouse (Zhejiang Province, China) (Jin et al. 2019). *C. hobsonii* has been described as a saprotroph inhabiting the leaf litter layer of *Quercus* (Orton 1960; Raj and Manimohan 2018). *C. hobsonii* is clustered into the section *Scyphoides* (Jian et al. 2020). Although *C. hobsonii* failed to form ectomycorrhizal structure (i.e., mantle and Hartig net) in

[†]Corresponding author: Z. Yuan; yuanzl@caf.ac.cn

The author(s) declare no conflict of interest.

Accepted for publication 29 January 2021.

Table 1. Genome characteristics and predicted features of the assembled homokaryotic *Clitopilus hobsonii* QYL-10

Characteristics	Homokaryotic *C. hobsonii* QYL-10
Total genome size (Mb)	36.93
N_{50} contig length (Mb)	3.3
Contig numbers	13
Maximum contig size (bp)	4,299,753
Minimum contig size (bp)	697,320
Average contig size (bp)	2,840,481
Telomere sequences at both ends	10
Telomere sequences at one end	3
Genome coverage	215.47
G+C (%)	49.2
Total transposable elements (Mb)	1.43
The total number of gene	12,710
Average gene length (bp)	1,903.74
The total number of noncoding genes	248
Total number of secondary metabolite biosynthetic gene clusters	36
Total number of genes encoding putative carbohydrate-active enzymes	361
Genome BUSCO (S/D/F/M)[a]	98.4 (97.8:0.6:0.3:1.3)

[a] S = complete and single-copy BUSCOs (benchmarking universal single-copy orthologs). D = complete and duplicated BUSCOs; F = fragmented BUSCOs; M = missing BUSCOs. Values shown are percent.

Quercus roots, we documented its beneficial effects on tree growth and potassium uptake (Jin et al. 2019; Peng et al. 2021). These findings support a dual mode of nutrition, i.e., litter saprotroph or facultative root biotroph. These ecological traits may imprint in its genome. To the best of our knowledge, no genome of *Clitopilus* species or Entolomataceae is currently available. The presence of the genome of *C. hobsonii* will provide an excellent opportunity to better understand the molecular mechanisms underlying this novel facultative symbiosis and, also, will allow studying its genome evolution and adaptation. To this end, we report the high-quality whole-genome sequence data of *C. hobsonii* QYL-10.

C. hobsonii was isolated as a dikaryotic mycelium in pure culture. To facilitate the genome assembly, we generated a homokaryotic strain of *C. hobsonii* QYL-10 through the protoplast regeneration technique (Zhao and Chang 1993). The homokaryotic culture was grown on potato dextrose agar at 22°C for 14 days. High-quality genomic DNA was extracted from freshly harvested mycelia, using the CATB approach (Lee et al. 1988). Sequencing was performed at the Beijing Novogene Bioinformatics Technology Co., Ltd. For the PacBio library, 20-Kb inserts for this strain were constructed using the SMRTbell template prep kit 1.0. One SMRTbell was used to sequence the DNA library with the PacBio Sequel II System. To polish the PacBio long reads–based genome assembly, a paired-end Illumina DNA library with 500 bp insert size was sequenced on the Illumina HiSeq 2500. A total of 11.77 Gb of long reads and 8.21 Gb of short reads, representing approximately 215× theoretical coverage, were generated respectively. Based on the Illumina short-reads data, we analyzed the k-mer distribution of the genome with Jellyfish (Marçais and Kingsford 2011). The result showed that the k-mer curve of *C. hobsonii* QYL-10 is significantly a single peak without a heterozygous peak. Based on the above plotting, we confirmed that the sequenced genome is a haploid.

To capture the maximum number of genes expressed in the *C. hobsonii* for improving structural gene annotations, we performed a comprehensive transcriptome sequencing of this fungus, grown on three media with different nitrogen sources. The basal medium contained (per liter) 1.0 g of glucose, 1.0 g of sucrose, 0.272 g of KH_2PO_4, 74 mg of $MgSO_4{\cdot}H_2O$, 88 mg of $CaCl_2{\cdot}2H_2O$, 12.5 mg of ferric EDTA, and 0.13 mg of thiamine-HCl, 15 g of agar, pH 5.8. The inorganic N sources, ammonium (0.19 g·per liter of NH_4Cl) and nitrate (0.42 g·per liter of $Ca(NO_3)_2{\cdot}4H_2O$), and organic N source, which was composed of a mixture of acidic, neutral, and aromatic amino acids (glutamine, glycine, valine, leucine, and phenylalanine) respectively, were separately added into the basal medium. For all treatments, a final N concentration was adjusted to 50 mg·per liter. Total fungal RNA was extracted with Trizol reagent (Invitrogen). The purified RNA was then reverse-transcribed to create the final cDNA library, following the protocol for the mRNA-Seq sample preparation kit (Illumina). The average insert size for the paired-end libraries was 300 bp (±50 bp). We performed the paired-end sequencing on an Illumina Hiseq 4000 platform (LC Sciences).

Genome assembly and error correction in the long reads was conducted using MECAT2 (Xiao et al. 2017), then, the initial assembly with Illumina short-reads data was polished by Pilon (v. 1.22) (Walker et al. 2014). The nuclear genome size of *C. hobsonii* is 36.93 Mb and comprises 13 contigs with a N_{50} length of 3.3 Mb and overall GC content of 49.2%. Telomeric 5′-TTAGGG-3′ repeats approximately 100 bp in length were manually identified. Of the 13 nuclear contigs, 10 contigs had characteristic telomere sequences at both ends, indicating that the 10 contigs approach the length of intact chromosomes (contig sizes 4,299,753 to 2,178,782 bp), and three contigs contain telomeric repeat sequences on the 5′ or 3′ end (contig sizes 3,337,404 to 697,320 bp) (Aksenova and Mirkin 2019). Genome-assembly quality was evaluated with the BUSCO (benchmarking universal single-copy orthologs) Basidiomycota_odb10 lineage data; results showed 98.4% completeness (a total of 1,736 BUSCO groups were searched), and five and 23 BUSCO orthologs were fragmented (0.3%) and missing (1.3%), respectively. These results suggest good integrity of the assembled genome (Simão et al. 2015) (Table 1).

Transposable elements (TEs) were identified by RepeatMasker (v. 4.07) and Repbase database (v. 23.06) (Bao et al. 2015), were also modeled *ab-initio* using RepeatModeler (v. 1.0.11) and LTR FINDER (Jurka et al. 2005) with default parameters. Approximately 1.43 Mb of TEs (3.88% of the total assembly) were identified. The most abundant of the transposable and repetitive element types present in the genome was the Class II long interspersed nuclear element, with 0.82 Mb (2.24%), and class I long terminal repeats, with 0.48 Mb (1.31%).

For noncoding genes, transfer RNAs (tRNAs) were predicted using tRNAscan-SE (Schattner et al. 2005), ribosomal RNA (rRNA) genes were annotated using the BLASTN tool against Basidiomycetes rRNA sequence; small nuclear and nucleolar RNAs (snRNA) in the genome were annotated by the infernal tool using Rfam database (Griffiths-Jones et al. 2005). A total of 217 tRNAs, 21 rRNAs, and 10 snRNAs were identified.

For accurate annotation of protein-coding genes, a combination of homology-based (Kent 2002; Slater and Birney 2005), *ab-initio* prediction (Stanke et al. 2006; Ter-Hovhannisyan et al. 2008), and transcriptome-based prediction methods (Kim et al. 2019; Pertea et al. 2015, 2016) was used. MAKER v. 2.31.9 annotation pipeline was then used to integrate *ab initio* gene predictions with protein homology evidence (Holt and Yandell 2011). Three reference genomes are used for homology-based inference: *Tricholoma matsutake* (Tricholomataceae), *Lyophyllum atratum* (Lyophyllaceae), and *Laccaria bicolor* (Tricholomataceae). The three species are taxonomically closed to *C. hobsonii* and their genomes and annotations were downloaded from the Joint Genome Institute MycoCosm website (Grigoriev et al. 2014). The transcripts were assembled and aligned to the reference genome using Hisat v2.2.0 (Kim et al. 2019) and Stringtie v2.0 (Pertea et al. 2015, 2016) with default parameters based on RNA-seq data. Overall, 12,710 protein-coding gene models were predicted, the average gene length was 1,903.74 bp (Table 1). Among them, 10,398 genes (81.81%) could be annotated with InterProScan, Gene Ontology, Kyoto Encyclopedia of Genes and Genomes, SwissProt, TrEMBL, and NR databases.

Genes and gene clusters involved in secondary metabolism were predicted using antiSMASH version 4.0.2 (Blin et al. 2017). A total of 36 secondary metabolite biosynthetic gene clusters were found in *C. hobsonii* QYL-10. Of them, 22 belonged to the terpene synthase group, 10 to the nonribosomal peptide synthases (NRPS)/NRPS-like, one to the type I polyketide synthases group, and the remaining clusters are unknown.

To identify the carbohydrate-active enzymes (CAZyme) repertoire of *C. hobsonii* QYL-10, the dbCAN2 web-based meta server was used for CAZyme annotation (Zhang et al. 2018). HMMER searches against the dbCAN hidden Markov model database, DIAMOND searches against the CAZy preannotated CAZyme sequence database, and Hotpep searches against the conserved CAZyme short peptide database (Buchfink et al. 2015; Busk et al. 2017). The genome of *C. hobsonii* QYL-10 contained 361 genes encoding putative CAZymes. This includes 189 glycoside hydrolases, 70 auxiliary activities, 53 glycosyl transferases, 25 carbohydrate esterases, 15 polysaccharide lyases, and nine carbohydrate-binding modules.

In summary, the high-quality genome assembly and the predicted gene repertoire of the homokaryotic *C. hobsonii* QYL-10 provides a valuable genomic resource for future comparative genomic studies of endophytes and for investigating its evolutionary biology.

Data availability.

Raw genomic data and the assembled sequences have been submitted to the National Center for Biotechnology Information (NCBI) database with BioProject accession number PRJNA675211

and BioSample accession number SAMN16692392. Accession numbers for the assembled genomes are JADPMO000000000. Sequence Read Archive (SRA) numbers are SRR13072277 and SRR13072276. Raw reads of transcriptome were submitted to the NCBI SRA database under accession numbers SRR13089187, SRR13089186, SRR13089185, SRR13089184, SRR13089183, SRR13089182, SRR13089181, SRR13089180, SRR13089179.

Author-Recommended Internet Resources

BUSCO lineage index: https://busco-data.ezlab.org/v4/data/lineages
dbCAN 2 server: http://bcb.unl.edu/dbCAN2/index.php
Repbase database: https://www.girinst.org/server/RepBase
RepeatModeler: http://www.repeatmasker.org/RepeatModeler
RepeatMasker: http://repeatmasker.org

Literature Cited

Aksenova, A. Y., and Mirkin, S. M. 2019. At the beginning of the end and in the middle of the beginning: Structure and maintenance of telomeric dna repeats and interstitial telomeric sequences. Genes (Basel) 10:118.

Bao, W., Kojima, K. K., and Kohany, O. 2015. Repbase update, a database of repetitive elements in eukaryotic genomes. Mob. DNA 6:11.

Blin, K., Wolf, T., Chevrette, M. G., Lu, X., Schwalen, C. J., Kautsar, S. A., Suarez Duran, H. G., de Los Santos, E. L. C., Kim, H. U., Nave, M., Dickschat, J. S., Mitchell, D. A., Shelest, E., Breitling, R., Takano, E., Lee, S. Y., Weber, T., and Medema, M. H. 2017. antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. Nucleic Acids Res. 45 (W1):W36-W41.

Buchfink, B., Xie, C., and Huson, D. H. 2015. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12:59-60.

Busk, P. K., Pilgaard, B., Lezyk, M. J., Meyer, A. S., and Lange, L. 2017. Homology to peptide pattern for annotation of carbohydrate-active enzymes and prediction of function. BMC Bioinformatics 18:214.

Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S. R., and Bateman, A. 2005. Rfam: Annotating non-coding RNAs in complete genomes. Nucleic Acids Res. 33:D121-D124.

Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., Smirnova, T., Nordberg, H., Dubchak, I., and Shabalov, I. 2014. MycoCosm portal: Gearing up for 1000 fungal genomes. Nucleic Acids Res. 42 (D1):D699-D704.

Holt, C., and Yandell, M. 2011. MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12:491.

Howe, V. K. 1964. Mycorrhizal fungi of white oak. Ph.D. dissertation. Iowa State University of Science and Technology, Ames, IA, U.S.A.

Jian, S. P., Bau, T., Zhu, X. T., Deng, W. Q., Yang, Z. L., and Zhao, Z. W. 2020. *Clitopilus, Clitocella*, and *Clitopilopsis* in China. Mycologia 112:371-399.

Jin, W., Peng, L., Zhang, X., Sun, H., and Yuan, Z. 2019. Effects of endophytic and ectomycorrhizal *Basidiomycetes* on *Quercus virginiana* seedling growth and nutrient absorption. J. Sustain. For. 38:457-470.

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. 2005. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. 110:462-467.

Keizer, P. J., and Arnolds, E. 1994. Succession of ectomycorrhizal fungi in roadside verges planted with common oak (*Quercus robur* L.) in Drenthe, The Netherlands. Mycorrhiza 4:147-159.

Kent, W. J. 2002. BLAT–the BLAST-like alignment tool. Genome Res. 12:656-664.

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. 37:907-915.

Lee, S. B., Milgroom, M. G., and Taylor, J. W. 1988. A rapid, high yield mini-prep method for isolation of total genomic DNA from fungi. Fungal Genet. Newsl. 35:23-24.

Marçais, G., and Kingsford, C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. Bioinformatics 27:764-770.

Molina, R., Massicotte, H., and Trappe, J. M. 1992. Specificity phenomena in mycorrhizal symbioses: community-ecological consequences and practical implications. Pages 357-423 in: Mycorrhizal Functioning: An Integrative Plant-fungal Process. M. Allen, ed. Chapman and Hall, New York.

Orton, P. D. 1960. New check list of British agarics and boleti: Part III. Notes on genera and species in the list. Trans. Br. Mycol. Soc. 43:159-384.

Peng, L., Shan, X., Yang, Y., Wang, Y., Druzhinina, I. S., Pan, X., Jin, W., He, X., Wang, X., Zhang, X., Martin, F. M., and Yuan, Z. 2021. Facultative symbiosis with a saprotrophic soil fungus promotes potassium uptake in American sweetgum trees. Plant Cell Environ. Published online.

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., and Salzberg, S. L. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat. Protoc. 11:1650-1667.

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. 33:290-295.

Raj, K. A., and Manimohan, P. 2018. A new species and a new record of *Clitopilus* and a description of *C. orientalis* from India based on morphology and molecular phylogeny. Phytotaxa 343:47-59.

Schattner, P., Brooks, A. N., and Lowe, T. M. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. 33:W686-W689.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210-3212.

Slater, G. S., and Birney, E. 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6:31.

Southworth, D. 2013. Oaks and mycorrhizal fungi. Pages 207-218 in: Oak: Ecology, Types and Management, Environmental Science, Engineering and Technology Series. C. A. Chuteira and A. B. Grao, eds. Nova Science Publishers Inc., New York.

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. 2006. AUGUSTUS: *Ab initio* prediction of alternative transcripts. Nucleic Acids Res. 34:W435-W439.

Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y. O., and Borodovsky, M. 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res. 18:1979-1990.

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., and Earl, A. M. 2014. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963.

Xiao, C. L., Chen, Y., Xie, S. Q., Chen, K. N., Wang, Y., Han, Y., Luo, F., and Xie, Z. 2017. MECAT: Fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. Nat. Methods 14:1072-1074.

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P. K., Xu, Y., and Yin, Y. 2018. dbCAN2: A meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 46 (W1):W95-W101.

Zhao, J., and Chang, S. T. 1993. Monokaryotization by protoplasting heterothallic species of edible mushrooms. World J. Microbiol. Biotechnol. 9:538-543.