

Supplementary File 1: Supplementary materials to this study.

Endogenous accessory chromosome mediates evolutionary trade-off between vegetative fitness and progeny advantage in a non-pathogenic basidiomycete

Miao Wang^{1,2}, Qi Wang³, Jian-Ping Xu⁴, Zheng Wang⁵ and Cai-Hong Dong^{1*}

¹ State Key Laboratory of Microbial Diversity and Innovative Utilization, Institute of Microbiology, Chinese Academy of Sciences, Beijing, 100101, China

² University of Chinese Academy of Sciences, Beijing, 100049, China

³ Hubei Provincial Hospital of Traditional Chinese Medicine, Wuhan 430071, China

⁴ Department of Biology, McMaster University, Hamilton, Ontario, L8S 4K1, Canada

⁵ Yale University, New Haven, Connecticut, 06520, USA

* Corresponding author, E-mail: dongch@im.ac.cn

Materials and methods

Detection of horizontal gene transfer

Potential horizontal gene transfer (HGT) candidates were screened using a BLAST-based taxonomic filtering strategy [1]. All predicted proteins were first searched against the JGI fungal genome database using BLASTP with an E-value cutoff of 10^{-5} , and proteins with no more than 20 hits were retained. The retained proteins were then searched against the NCBI NR database restricted to Fungi (taxid: 4751), using the same E-value cutoff, and proteins with no more than 20 fungal hits were kept. The remaining candidates were further searched against the NCBI NR database excluding Fungi (taxid: 4751), and proteins with no more than 5 non-fungal hits were retained for phylogenetic inspection. Candidate proteins were further evaluated using phylogenetic trees constructed from their homologs (Fig. S2). Proteins showing clear non-fungal affinity in the phylogeny were considered putative HGT candidates.

Results

Genome reassembly of strain SS20

To improve the completeness and accuracy of the genome, a reassembly was performed employing a combination of strategies, including Hifiasm, Canu, and Falcon, along with meticulous manual refinement. The new draft assemblies were then aligned to the previous genome to resolve existing breaks (Fig. S4). Among the seven identified breaks, Breaks 1 and 6 were classified as misassemblies. Chr01 from the previous assembly was reassembled using Falcon from contigs 000002F and 000014F, which contained two and one telomeres, respectively, thereby confirming the presence of a misassembly with three telomeres. Additionally, contig 000014F, with rDNA sequences at one end, could overlap and fuse with contig 000026F, which featured a telomere at one end and rDNA at the other. These results revealed that Chr01 in the previous assembly was actually comprised of two misassembled chromosomes, corresponding to Chr11 and Chr04 in the new assembly. For Break 6, both the Canu and Hifiasm draft assemblies generated contigs with telomeres at both ends, one of which was located near Break 6, supporting its classification as a misassembly. The remaining Breaks (2, 3, 4, 5, and 7) were solved through the Hifiasm assembly. Heatmap generated using Hi-C data showed the gap-filling regions with no noticeable noise, confirming the accuracy of reassembly. The rDNA copy number was assessed through BLAST analysis with a 22.9 kb query sequence that included one rDNA unit and its intergenic spacer region (ISR). The total cumulative length of rDNA-matching sequences amounted to 186,341,015 bp, corresponding to a chromosome coverage of 278×, and yielding an estimated 29.25 rDNA copies. These matching reads were subsequently assembled using Canu v2.2, and manually connected based on variations in the ISR, resulting in a final assembly of 0.67 Mb in length, comprising 29 rDNA units located on Chr11.

After manually resolving all breaks and incomplete rDNA sequences in the previous assembly, anomalous regions with irregular local coverage were also carefully polished, resulting in a complete, gap-free genome. Furthermore, a circular mitochondrial genome of 136,067 bp was assembled, which was identical to the mitochondrial genome of strain PC1609 (GenBank: MH661258.1).

Genome assembly and phasing of heterokaryotic strain 1695

Strain 1695 is a wild isolate collected from Anhui Province, China, exhibiting a slightly yellowish mycelial phenotype compared with strain 776. PacBio circular consensus sequencing of 1695 generated 168,713 HiFi reads, totaling 3.45 Gb of data, corresponding to approximately 60× genomic coverage. Initial assembly with Hifiasm generated two phased contig sets, which were provisionally designated hapA and hapB after telomere completion. SNPs called from NGS data of a protoplast-derived monokaryotic strain of 1695 were used to detect and remove chimeric assemblies and to refine phase assignments across chromosomes. The final haplotype-resolved assemblies, hapA and hapB, were 56.18 Mb and 56.73 Mb in size, respectively, each comprising 14 chromosomes with 28 telomeres and N50 values of 4.34 Mb and 4.17 Mb. Repetitive elements accounted for 41.41% and 41.79% of the two genomes, with 13,602 and 13,766 protein-coding genes predicted, respectively.

Genome assembly of the heterozygous, monosporic strain 7S14

We generated 467.2× raw ultralong ONT data, 121.4× PacBio HiFi data, and 273.6× Hi-C data for the assembly of the 7S14 genome. A hybrid assembly was performed using Hifiasm v0.19.8-r603 with PacBio HiFi, ultralong ONT and Hi-C data, which produced two phased assembly sets. For contigs in either set lacking telomeric sequences at both ends, we extended them until overlaps with other contigs were identified. The accuracy of these extensions was further evaluated by collinearity analysis against T2T_SS20. We then mapped the short-read data separately to set1 and set2, calculated SNP allele frequencies, and used this information to identify and resolve chimeric assemblies.

References

- [1] Zhang Q, Chen X, Xu C, Zhao H, Zhang X, et al. 2019. Horizontal gene transfer allowed the emergence of broad host range entomopathogens. *Proceedings of the National Academy of Sciences of the United States of America* 116:7982–7989 <http://doi.org/10.1073/pnas.1816430116>