Check for updates

Resequencing the Genome of *Malassezia restricta* Strain KCTC 27527

Yong-Joon Cho,ᵃ Minji Park,ᵇ Won Hee Jungᵇ

ᵃKorea Polar Research Institute, Incheon, South Korea
ᵇDepartment of Systems Biotechnology, Chung-Ang University, Anseong, South Korea

ABSTRACT The draft genome sequence of *Malassezia restricta* KCTC 27527, a clinical isolate from a patient with dandruff, was previously reported. Using the PacBio Sequel platform, we completed and reannotated the genome of *M. restricta* KCTC 27527 for a better understanding of the genome of this fungus.

Malassezia species are recognized to be involved in skin diseases, including dandruff, seborrheic dermatitis, and atopic dermatitis. Among the 17 identified *Malassezia* species, *M. restricta* is the predominant species on human skin and is particularly associated with dandruff, as suggested by recent microbiome analyses (1–4).

We previously sequenced and analyzed the genome of *M. restricta* KCTC 27527, a clinical isolate from a patient with dandruff, in South Korea (5). The previous assembly generated 51 contigs that were assembled into 18 scaffolds containing 3,580 coding sequences (CDSs), representing an overall completeness of 89.7% with Core Eukaryotic Genes Mapping Approach (CEGMA) analysis (5, 6). To address the incompleteness of the previous genome sequencing, we resequenced and completed the genome of *M. restricta* KCTC 27527 in the current study.

M. restricta KCTC 27527 cells were grown in Leeming and Notman agar (LNA) medium (0.5% glucose, 1% peptone, 0.01% yeast extract, 0.8% bile salt, 0.1% glycerol, 0.05% glycerol monostearate, 0.05% Tween 60, 1.2% agar, 0.5% whole-fat cow milk, and 170 μg/ml chloramphenicol) at 34°C for 3 days (7), and genomic DNA was extracted and a SMRTbell library was prepared according to the manufacturer's instructions (8). Genome sequencing was performed using P6-C4 chemistry on one cell of a PacBio Sequel platform (Pacific Biosciences). Raw reads were *de novo* assembled using the Canu v. 1.7 assembler with the parameter "genomeSize = 7.3m," and the assembled contigs were polished with the Arrow consensus caller in PacBio SMRT Link v. 5.0.1 (9). Telomeric motifs in chromosomal ends and mitochondrial contigs were manually curated. Discrepancies between contigs from the previously reported assembly and assembly of PacBio Sequel reads were corrected through the analysis using the CodonCode Aligner software package (CodonCode Corporation). The first round of gene prediction was performed with BRAKER v. 2.1.0 with the parameter "minimum intron length = 20" (10). In this process, *de novo* assembly of existing transcriptome sequencing (RNA-Seq) data on the Gene Expression Omnibus (GEO) database (accession number GSE112036) was used to reflect the exon-intron structure. RNA-Seq raw reads were cleaned by Trimmomatic v. 0.36 and mapped using Hierarchical Indexing for Spliced Alignment of Transcripts (HISAT) v. 2.1.0 (11, 12). After the first round, the BRAKER config file was modified by setting the normal penalty parameter for introns to 0 and the bonus for RNA-Seq data to 1e + 100 to reflect only introns derived from actual data. The final gene prediction was performed using AUGUSTUS v. 3.2.3 with the parameter "min_intron_len = 15" and

Address correspondence to Won Hee Jung, whjung@cau.ac.kr.

Y.-J.C. and M.P. made equal contributions and should be considered co-first authors.

TABLE 1 Summary of genome statistics for *M. restricta* KCTC 27527

Genome statistic	Value
Genome size (bp)	7,330,907
No. of chromosomes	9
Chromosome size range (Mbp)	0.16–1.42
No. of genes (protein coding genes)	4,390
Mean gene length (bp)	1,492
Gene density (genes per Mb)	598.91
Genome GC content (%)	55.79
Mitochondrial genome size (bp)	38,720
Mitochondrial GC content (%)	31.4

the hints and config files generated in the processes discussed above (13). We corrected the gene structure where the splicing site differed from the actual transcript with the RNA-Seq mapping results in the Integrative Genomics Viewer (IGV) genome browser. Genome annotation was carried out with the NCBI RefSeq database release 88, eggNOG v. 4.5, and the KEGG database (14–16). Mitochondrial gene prediction and annotation were performed using MITOS2 (17).

As a result of the resequencing, all gaps were filled and the assembly was completed. A total of 9 chromosomes and a mitochondrion of 7,330,907 bp (GC content, 55.79%) and 38,720 bp (GC content, 31.4%), respectively, were estimated with a read coverage of 38.8×. Further, 4,390 CDSs, 29 rRNAs, 74 tRNAs, and 9 noncoding RNAs (ncRNAs) were identified in the annotated assemblies of the chromosomes. The annotated mitochondrial genome contained 16 CDSs, 2 rRNAs, and 24 tRNAs. The genome statistics are summarized in Table 1. The new assembly contained an additional 810 and 3 predicted CDSs in the chromosomes and mitochondrion, respectively, compared to our previous assembly.

Data availability. This whole-genome sequence has been deposited in GenBank under the accession numbers CP030251 to CP030260 from BioProject PRJNA477735.

ACKNOWLEDGMENT

REFERENCES

1. Findley K, Oh J, Yang J, Conlan S, Deming C, Meyer JA, Schoenfeld D, Nomicos E, Park M, NIH Intramural Sequencing Center Comparative Sequencing Program, Kong HH, Segre JA. 2013. Topographic diversity of fungal and bacterial communities in human skin. Nature 498:367–370. https://doi.org/10.1038/nature12171.
2. Clavaud C, Jourdain R, Bar-Hen A, Tichit M, Bouchier C, Pouradier F, El Rawadi C, Guillot J, Ménard-Szczebara F, Breton L, Latgé J-P, Mouyna I. 2013. Dandruff is associated with disequilibrium in the proportion of the major bacterial and fungal populations colonizing the scalp. PLoS One 8:e58203. https://doi.org/10.1371/journal.pone.0058203.
3. Xu Z, Wang Z, Yuan C, Liu X, Yang F, Wang T, Wang J, Manabe K, Qin O, Wang X, Zhang Y, Zhang M. 2016. Dandruff is associated with the conjoined interactions between host and microorganisms. Sci Rep 6:24877. https://doi.org/10.1038/srep24877.
4. Park T, Kim HJ, Myeong NR, Lee HG, Kwack I, Lee J, Kim BJ, Sul WJ, An S. 2017. Collapse of human scalp microbiome network in dandruff and seborrhoeic dermatitis. Exp Dermatol 26:835–838. https://doi.org/10.1111/exd.13293.
5. Park M, Cho YJ, Lee YW, Jung WH. 2017. Whole genome sequencing analysis of the cutaneous pathogenic yeast Malassezia restricta and identification of the major lipase expressed on the scalp of patients with dandruff. Mycoses 60:188–197. https://doi.org/10.1111/myc.12586.
6. Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23:1061–1067. https://doi.org/10.1093/bioinformatics/btm071.
7. Leeming JP, Notman FH. 1987. Improved methods for isolation and enumeration of Malassezia furfur from human skin. J Clin Microbiol 25:2017–2019.
8. Martin F. 2017. High quality genomic DNA extraction using CTAB and Qiagen genomic-tip. http://1000.fungalgenomes.org/home/protocols/high-quality-genomic-dna-extraction/.
9. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.
10. Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics 32:767–769. https://doi.org/10.1093/bioinformatics/btv661.
11. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
12. Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12:357–360. https://doi.org/10.1038/nmeth.3317.
13. Stanke M, Tzvetkova A, Morgenstern B. 2006. AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. Genome Biol 7:s1–s11. https://doi.org/10.1186/gb-2006-7-s1-s11.
14. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome

sequences. J Mol Biol 428:726–731. https://doi.org/10.1016/j.jmb.2015.11.006.

15. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44:D733–D745. https://doi.org/10.1093/nar/gkv1189.

16. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, Jensen LJ, von Mering C, Bork P. 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res 44:D286–D293. https://doi.org/10.1093/nar/gkv1248.

17. Bernt M, Donath A, Juhling F, Externbrink F, Florentz C, Fritzsch G, Putz J, Middendorf M, Stadler PF. 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol 69:313–319. https://doi.org/10.1016/j.ympev.2012.08.023.