AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

EUKARYOTES

Draft Genome Sequences of *Trichophyton rubrum* CMCC(F)T_{1i} and *Trichophyton violaceum* CMCC(F)T_{3l} by Illumina 2000 and Pacific Biosciences

Ping Zhan,[a,b] Sybren de Hoog,[c] Weida Liu[a]

Department of Mycology, Institute of Dermatology, Chinese Academy of Medical Science & Peking Union Medical College, Nanjing, China[a]; Dermatology Hospital of Jiangxi Province, Dermatology Institute of Jiangxi Province, Nanjing, China[b]; Westerdijk Fungal Biodiversity Institute, Utrecht, The Netherlands[c]

ABSTRACT One strain of *Trichophyton rubrum* CMCC(F)T_{1i} (=CBS 139224) isolated from onychomycosis and one strain of *Trichophyton violaceum* CMCC(F)T_{3l} (=CBS 141829) isolated from tinea capitis in China were whole-genome sequenced by Illumina/Solexa, while the former was also sequenced by Pacific Biosciences sequencing in parallel.

Received 8 August 2017 **Accepted** 10 August 2017 **Published** 28 September 2017

Citation Zhan P, de Hoog S, Liu W. 2017. Draft genome sequences of *Trichophyton rubrum* CMCC(F)T_{1i} and *Trichophyton violaceum* CMCC(F)T_{3l} by Illumina 2000 and Pacific Biosciences. Genome Announc 5:e00920-17. https://doi.org/10.1128/genomeA.00920-17.

Address correspondence to Weida Liu, liumyco@hotmail.com.

Trichophyton rubrum and *Trichophyton violaceum* are the most common siblings in the *T. rubrum* complex (1). They are highly conserved at the sequencing level and have quite different phenotypes (1, 2). To explore the different pathogenicity of these two dermatophytes, we whole-genome sequenced one strain of *T. rubrum* CMCC(F)T_{1i} and one strain of *T. violaceum* CMCC(F)T_{3l} in South China by Illumina/Solexa, while the former was sequenced by PacBio RS in parallel.

The strains were cultured in solid Sabouraud glucose agar and incubated at 27°C for 7 to 20 days. Genomic DNA was extracted using the EZNA fungal DNA kit (Omega, USA).

For Illumina sequencing, 5-μg genomic DNA was used in the sequencing library construction. Paired-end libraries with insert sizes of ~300 bp were constructed using the AIR paired-end DNA sequencing kit (Bioscientific). Subsequently, 100 bp at each end was sequenced using Illumina Hiseq2000. The *T. violaceum* genome was assembled using Velvet assembler (v1.2.09), and contigs with a length less than 200 bp were discarded to get reliable assembled results.

For Pacific Biosciences sequencing, 8- to 10-kb insert whole-genome shotgun libraries were generated and sequenced on a Pacific Biosciences RS instrument using standard methods. The complete genome sequence of strain *T. rubrum* was assembled using both the PacBio reads and Illumina reads. The assembly was produced first using a hybrid *de novo* assembly solution modified by Koren et al., in which a de Bruijn-based assembly algorithm and a CLR reads correction algorithm were integrated in the PacBioToCA pipeline in the Celera assembler (3, 4). The last circular step was checked and finished manually. The final assembly generated a circular genome sequence with no existing gap.

Identification of predicted coding sequences (CDSs) was performed using AUGUSTUS v3.0.1 (http://bioinf.uni-greifswald.de/augustus/). Open reading frames (ORFs) with less than 300 bp were discarded. Then, the remaining ORFs were queried against the nonredundant (NR) NCBI, SwissProt (http://uniprot.org), KEGG (http://www.genome.jp/kegg/), and COG (http://www.ncbi.nlm.nih.gov/COG) databases to do functional annotation.

The third-generation sequencing technique can supply long, perfect sequences, as well as accurate mitochondrial structure. *T. rubrum* obtained 19 superscaffolds (scaf-

fold$_{50}$, 2,198,313 bp) with a whole-genome length of 22,289,584 bp, while *T. violaceum* obtained 278 scaffolds (scaffold$_{50}$, 1335,347 bp) with 23,310,379 bp as the whole length. The overall G+C content of the entire genome is 48.34%, and the *N* rates are as low as 0.055% for *T. rubrum* and 47.22% and 0.76% for *T. violaceum*, respectively. A total of 7,170 genes were predicted for *T. rubrum* and 7,415 for *T. violaceum*. The two genomes show a quite high genome similarity and synteny, with 99.38% identity at the nucleic acid level. As for genes, there were only 18 different CDSs altogether.

This is the first report of the whole-genome sequence of *T. violaceum* and the first application of PacBio RS for dermatophyte genome sequencing. These high-quality sequences and subsequent comparative genomic analysis might provide a better understanding of the pathogenicity and diversity of *T. rubrum* and *T. violaceum* at the genome level.

Accession number(s). This whole-genome shotgun project has been deposited in DDBJ/EMBL/GenBank under the accession numbers LHPM00000000 for *T. rubrum* CMCC(F)T$_{1i}$ and LHPN00000000 for *T. violaceum* CMCC(F)T$_{3l}$.

ACKNOWLEDGMENTS

This research was supported by the National Natural Science Foundation of China (13008316), National Key Basic Research Program of China (2013CB531605), and the Ministry of Science and Technology Foundation (2013FY113700).

REFERENCES

1. Gräser Y, Kuijpers AFA, Presber W, de Hoog GS. 2000. Molecular taxonomy of the *Trichophyton rubrum* complex, J Clin Microbiol 38:3329–3336.
2. Ohst T, de Hoog S, Presber W, Stavrakieva V, Gräser Y. 2004. Origins of microsatellite diversity in the *Trichophyton rubrum-T. violaceum* clade (Dermatophytes). J Clin Microbiol 42:4444–4448. https://doi.org/10.1128/JCM.42.10.4444-4448.2004.
3. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10:563–569. https://doi.org/10.1038/nmeth.2474.
4. Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, Wang Z, Rasko DA, McCombie WR, Jarvis ED, Adam M Phillippy AM. 2012. Hybrid error correction and *de novo* assembly of single-molecule sequencing reads. Nat Biotechnol 30:693–700. https://doi.org/10.1038/nbt.2280.