Complete Genome Sequence of an Australian Strain of the Lichen-Forming Fungus *Endocarpon pusillum* (Hedwig)

Oliver L. Mead,[a] Cécile Gueidan[a]

[a]Australian National Herbarium, National Research Collections of Australia, Commonwealth Scientific and Industrial Research Organization, Canberra, Australia

ABSTRACT The cosmopolitan lichen-forming fungus *Endocarpon pusillum* (Hedwig) has previously been used as a model for the study of symbiosis and drought resistance. Here, we present the annotated genome of the Australian strain *Endocarpon pusillum* EPUS1.4. This genome sequence provides additional information on the ability of this species to produce secondary metabolites.

Lichens are natural reservoirs of novel compounds, with over 800 lichen compounds described to date (1) and many being used in industry (2). As such, they represent fertile opportunities for bioprospecting. However, lichen-forming fungi are notoriously slow growing and difficult to cultivate under laboratory conditions, and often the symbionts are recalcitrant to isolation (3). To overcome these problems, several lichen metagenomes and genomes from axenic cultures have been sequenced (4, 5). These projects aim to identify the biosynthetic gene clusters responsible for producing valuable molecules. Similarly, to discover novel molecules through genome mining, we sequenced the genome of the model soil crust lichen *Endocarpon pusillum* (Hedwig). Two other *Endocarpon* genome sequences have recently been reported (5, 6). Park et al. (6) employed a genomics approach to investigate desiccation resistance in a Korean isolate of *E. pusillum*. Wang et al. (5) used genomics and transcriptomics to identify the molecular mechanisms underlying lichen symbiosis. To strengthen the emerging community of lichen genomics, we contribute the annotated genome sequence of an Australian isolate of *E. pusillum*, strain EPUS1.4.

The specimen of *E. pusillum* (C. Gueidan 2364) was collected from the CSIRO Black Mountain site, north of Christian Road in Canberra, Australia, in 2016 and deposited in the CANB collection (accession number CANB 913709). Ascospores were shot onto peptone-dextrose agar (PDA) plates, and single ascospores were isolated and grown on liquid potato-dextrose broth (PD) medium in an incubator with 20°C/18°C 12-h day/night cycles. Before extraction, the stock culture was ground and an inoculate transferred to stationary YSSG medium, consisting of yeast extract (5 g/liter), sucrose (10 g/liter), sorbitol (10 g/liter), and γ-aminobutyric acid (GABA) medium (1 g/liter).

Genomic DNA and total RNA were extracted separately from ca. 20 mg of *E. pusillum* EPUS1.4 using a phenol-chloroform and sodium dodecyl sulfate (2% wt/vol)-beta-mercaptoethanol (1% vol/vol) emulsion at room temperature (7). The John Curtin School of Medical Research generated 25 million paired-end 2×300-bp reads from 100 ng genomic DNA on the Illumina MiSeq platform and 150 million paired-end 2×75-bp reads from 100 ng RNA on the Illumina NextSeq 500 platform. The Illumina Nextera XT v3 library kit was used to prepare both nucleic acids for sequencing. The Oxford Nanopore (ONP) MinION (FLO-MIN106D, R9) platform generated 1.6 Gb raw reads from 200 ng DNA using a PCR sequencing kit (SQK-PSK004, ONP, UK).

Fastp v0.19.6 (8) was used to trim and describe the quality of all short reads, with default settings, generating 18 million DNA and 144 million RNA high-quality

reads. Long-read data were base called and quality controlled using GUPPY v3.2.2-GPU (ONP) (9), generating 400,000 high-quality reads, as counted by NanoPlot (10).

SPAdes v3.12.0 (11), using a kmer length of 127 bp, 20 threads, 256 Gb of RAM, and the--nanopore switch, was used to assemble high-quality genomic DNA (gDNA) short and long reads into a hybrid assembly, EP01v1.6.4. This 33.7-Mb assembly contained 2,902 contigs and had an N_{50} value of 158 kb and a GC content of 48.4% (QUAST v4.3 [12]). It contained 3,474 of 4,046 (85.9%) Eurotiomycetes benchmarking universal single-copy ortholog (BUSCO) genes (BUSCO v4.0.1 [13]). Trinity v2.3.2, with default settings, was used to assemble a 126-Mb transcriptome from the RNAseq reads (14). HiSat2 v2.1.0 was used to map these reads to EP01v1.6.4 (15, 16). Funannotate v1.7.1 (17), using the clean, sort, mask, predict, and annotate tools, and Blast2Go v5.2.5 (18), using default settings, annotated EP01v1.6.4. This annotation included *ab initio* gene models from Augustus v3.3.2 (19), derived using --singlestrand=true --cds=on --codingseq=on switches, and the *Aspergillus nidulans* model, as well as aligned RNAseq reads and transcriptomic evidence, to produce 12,503 predicted gene models.

Data availability. Data are available in GenBank under BioProject accession number PRJNA589713 and accession number JAACFV000000000.

REFERENCES

1. Elix JA. 2014. A catalogue of standardized chromatographic data and biosynthetic relationships for lichen substances. Published by the author, Canberra, Australia.
2. Goga M, Elečko J, Marcinčinová M, Ručová D, Bačkorová M, Bačkor M. 2020. Lichen metabolites: an overview of some secondary metabolites and their biological potential, p 175–209. *In* Mérillon J-M, Ramawat KG (ed), Co-evolution of secondary metabolites. Springer, Cham, Switzerland.
3. Sancho LG, Green TGA, Pintado A. 2007. Slowest to fastest: extreme range in lichen growth rates supports their use as an indicator of climate change in Antarctica. Flora 202:667–673. https://doi.org/10.1016/j.flora.2007.05.005.
4. Calchera A, Grande FD, Bode HB, Schmitt I. 2019. Biosynthetic gene content of the "perfume lichens" Evernia prunastri and Pseudevernia furfuracea. Molecules 24:203. https://doi.org/10.3390/molecules24010203.
5. Wang Y-Y, Liu B, Zhang X-Y, Zhou Q-M, Zhang T, Li H, Yu Y-F, Zhang X-L, Hao X-Y, Wang M, Wang L, Wei J-C. 2014. Genome characteristics reveal the impact of lichenization on lichen-forming fungus Endocarpon pusillum Hedwig (Verrucariales, Ascomycota). BMC Genomics 15:34. https://doi.org/10.1186/1471-2164-15-34.
6. Park S-Y, Choi J, Lee G-W, Park C-H, Kim JA, Oh S-O, Lee Y-H, Hur J-S. 2014. Draft genome sequence of Endocarpon pusillum strain KoLRILF000583. Genome Announc 2:e00452-14. https://doi.org/10.1128/genomeA.00452-14.
7. González-Mendoza D, Argumedo-Delira R, Morales-Trejo A, Pulido-Herrera A, Cervantes-Díaz L, Grimaldo-Juarez O, Alarcón A. 2010. A rapid method for isolation of total DNA from pathogenic filamentous plant fungi. Genet Mol Res 9:162–166. https://doi.org/10.4238/vol9-1gmr680.
8. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560.
9. Wick RR, Judd LM, Holt KE. 2019. Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol 20:129. https://doi.org/10.1186/s13059-019-1727-y.
10. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics 34:2666–2669. https://doi.org/10.1093/bioinformatics/bty149.
11. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
12. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
13. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351.
14. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652. https://doi.org/10.1038/nbt.1883.
15. 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.
16. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
17. Palmer JM, Stajich JE. 2019. Funannotate v1.7. Zenodo. https://doi.org/10.5281/zenodo.3594559.
18. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676. https://doi.org/10.1093/bioinformatics/bti610.
19. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res 34:W435–W439. https://doi.org/10.1093/nar/gkl200.