High-Quality Genome Assembly of *Nannochloris desiccata* 2437 and Its Associated Bacterial Community

Claire K. Sanders, Thomas C. Biondi, Wyatt Eng, Yuliya A. Kunde, Blake T. Hovde, Taraka Dale

Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico, USA

Claire K. Sanders and Thomas C. Biondi contributed equally to this work. Author order was determined by seniority.

**ABSTRACT**

High-quality genome sequences were generated for the nonaxenic marine microalga *Nannochloris desiccata* UTEX 2437 and eight of its associated environmental bacterial species. *N. desiccata* UTEX 2437 is diploid, and its 20.738-Mbp nuclear genome sequence is assembled in 29 contigs.

Microalgae are being widely investigated for the production of biofuels and nutraceuticals. The industry is continuously searching for species with phenotypes that include fast growth and high lipid yields, particularly those grown in brackish or marine waters. A microbial mat was collected from a salt marsh in Laguna Figueroa, Baja California, Mexico, and desiccated. In 1984, a eukaryotic alga from this mat was reconstituted and deposited into the UTEX culture collection as *Chlorella desiccata* 2437, this culture includes a natural microbial community (1). An axenic isolate from UTEX 2437, UTEX 2526, was previously sequenced (NCBI GenBank accession number JAGTXX000000000). Phylogenetic analysis revealed UTEX 2526 to reside in the genus *Nannochloris* (14).

*Nannochloris desiccata* 2437 was purchased from UTEX, maintained on f/2 agar plates, then grown in liquid silicate-free, modified f/2 medium (Sanders et al., submitted). The culture was illuminated with 300 μmoles photons m⁻² s⁻¹ with a 16/8-h light/dark cycle and maintained in a 1% CO₂ atmosphere. Cell pellets were stored at −80°C, then thawed, washed, and embedded into 1% low melting point (LMP) agarose plugs. Protoplasting solution was used to remove the cell wall, followed by lysis using proteinase K and digestion using beta-agarase I to release the genomic DNA (gDNA) into the solution. The DNA was subsequently purified using a high salt:phenol:chloroform:isoamyl alcohol protocol and concentrated using AMPure PB beads. gDNA was fragmented using the Megaruptor 2 instrument with a target size of 20 kbp. Libraries were constructed using the PacBio Express low DNA input HiFi template prep protocol, size selected using diluted AMPure PB beads, which removed all DNA fragments of <3 kbp, and sequenced on a PacBio Sequel instrument using chemistry v3.0 and DNA polymerase v3.0. Two single-molecule real-time (SMRT) cells 1M LR were used per library. Twenty-hour movies were recorded, and HiFi reads were extracted using PacBio’s ccs v4.2.0 module (https://github.com/PacificBiosciences/ccs) (details in the work of Sanders et al., submitted).

Based on the knowledge that this alga species is diploid from the previously assembled UTEX 2526, HiFi reads were assembled using the diploid aware assembler HiFiasm v0.12-r304, with default parameters (2). Bacterial reads were assembled using Flye v2.8.2-b1689 with the –meta parameter. The resulting contigs were binned using two different binning tools: METABAT2 v2.12.1 (3) and MaxBin2 v2.2.7 (4). The resulting bins were combined into consensus bins using DASTool v1.1.2 (5). The consensus bins were classified using GTDB-Tk (6), the completeness was assessed using BUSCO (7), and the relative abundance was calculated using the CheckM tools “coverage”
and “profile” (8). Eukaryotic and prokaryotic reads in the sample were separated by aligning all of the UTEX 2437 reads (minimap2 v2.17-r941 with the parameter “-ax asm20”) with the axenic assembly of UTEX 2526. The aligned (algal) reads and unaligned (nonalgal) reads were then separated using SAMtools view v1.9 with the parameters -f 4 and -f 4, respectively.

The nuclear genome assembly of *N. desiccata* 2437 is 20.576 Mbp with a GC content of 45.0%. There were 24 contigs with a maximum size of 2.689 Mbp and a contig *N*<sub>dp</sub> size of 1.522 Mbp. Organelle genomes were assembled as complete circular contigs, with a 40,238-bp mitochondrial genome and a 93,009-bp chloroplast genome. Genome annotation was performed using BRAKER2 v2.1.5 (9, 10). Functional gene motifs and domains were added using InterProScan v5.26-65.0-intel-2017b (11) against the CDD and TIGRFAM databases. Gene functions were then allocated using the best BLASTP (12) match to the UniProt Swiss-Prot database (13).

Bacterial reads were assembled, resulting in the identification of the presence of eight species with 39.5% to 99.2% completeness, four having benchmarking universal single-copy ortholog (BUSCO) scores of >90%. This includes the complete, 3,220,578-bp, circular genome sequence of *Erythrobacter* sp.

**Data availability.** All genome sequences were deposited at NCBI under BioProject accession number PRJNA704951. The reads for *N. desiccata* UTEX 2437 were deposited in the NCBI SRA under accession number SRR15813383. The bacterial assembly accession numbers are listed in Table 1, and the algal genome assembly can be found under GenBank accession number GCA_019202925.

**ACKNOWLEDGMENT**

This work was funded by the U.S. Department of Energy Bioenergy Technologies Office Annual Operating Plan project NL0025841.

### REFERENCES

1. Margulis L, Hinkle G, Mckhann H, Moynihan B, Brown W. 1988. *Mychos-nastes desiccatus* BROWN sp. nova (Chlorococcales, Chlorophyta)—an intertidal alga forming achronrophylous desiccation-resistant cysts. Arch Hydrobiol Suppl Algal Stud 78:425–446.
2. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods 18:170–175. https://doi.org/10.1038/s41592-020-01056-5.
3. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. Peer J 7:e7359. https://doi.org/10.7717/peerj.7359.
4. Wu YW, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenome datasets. Bioinformatics 32:605–607. https://doi.org/10.1093/bioinformatics/btv638.
5. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol 3:836–843. https://doi.org/10.1038/s41564-018-0171-1.
6. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. Bioinformatics 36:1925–1927. https://doi.org/10.1093/bioinformatics/btz848.
7. Simão 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.
8. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.
9. Brunet I, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genom Bioinform 3:qaa108. https://doi.org/10.1093/narg/bqaa108.
10. Gremme G. 2013. Computational gene structure prediction. PhD dissertation. Universität Hamburg, Hamburg, Germany. https://ediss.sub.uni-hamburg.de/handle/ediss/4964

11. Blum M, Chang HY, Chuguransky S, Grego T, Kandasamy S, Mitchell A, Nuka G, Paysan-Lafosse T, Qureshi M, Raj S, Richardson L, Salazar GA, Williams L, Bork P, Bridge A, Gough J, Haft DH, Letunic I, Marchler-Bauer A, MI H, Natale DA, Necci M, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman A, Finn RD. 2021. The InterPro protein families and domains database: 20 years on. Nucleic Acids Res 49:D344–D354. https://doi.org/10.1093/nar/gkaa977.

12. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

13. The UniProt Consortium. 2021. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res 49:D480–D489. https://doi.org/10.1093/nar/gkaa1100.

14. Sanders CK, Hanschen ER, Blondi TC, Hovde BT, Kunde YA, Eng WL, Kwon T, Dale T. 2022. Phylogenetic analyses and recassification of the oleaginous marine species Nannochloris sp. “desiccata” (Trebouxiophyceae, Chlorophyta), formerly Chlorella desiccata, supported by a high-quality genome assembly. J Phycol 58:436–448. https://doi.org/10.1111/jpy.13242.