Draft Genome Sequence and Annotation of the Synthetic Textile Dye-Decolorizing Strain *Bacillus amyloliquefaciens* AD20, Isolated from a Dye Waste Pond

Imade Y. Nsa,[a] Ayodeji A. Odunsi,[a] Olukayode O. Amund[a,b]

aDepartment of Microbiology, Faculty of Science, University of Lagos, Lagos, Nigeria
bDepartment of Biological Sciences, Faculty of Basic and Applied Sciences, Elizade University, Ilara-Mokin, Ondo State, Nigeria

ABSTRACT A limited number of *Bacillus amyloliquefaciens* genome sequences have been generated and are available in the public domain from soil, fermented foods, and plants. Here, we report the whole-genome sequence of *B. amyloliquefaciens* AD20, isolated from a dye pond with azo dye decolorization capabilities.

*B*acillus amyloliquefaciens*, a member of the *Bacillus subtilis* group, is a Gram-positive rod which produces α-amylase. *B. amyloliquefaciens* is associated with fighting root pathogens (1) and the breakdown of synthetic textile dyes (2–5). *B. amyloliquefaciens* AD20, isolated from a dye waste pond (DWP) near a textile factory, has dye-reducing capabilities.

To isolate culturable bacteria from the DWP, 1 g of sediment was inoculated into 99 ml of sterile distilled water to make an initial 10^{-2} dilution and further series up to 10^{-7}. Aliquots from 10^{-4} to 10^{-6} dilution tubes were inoculated onto tryptic soya agar (TSA) plates and incubated at 37°C for 24 h. After incubation, individual isolates selected from the mixed culture TSA plates were subcultured thrice on TSA to obtain pure cultures of the distinct bacterial isolates. Glycerol stocks of the isolates were stored at −70°C. *B. amyloliquefaciens* was one of the bacterial isolates from the screen that showed dye decolorization capabilities on a medium infused with the dyes.

To identify *B. amyloliquefaciens* AD20 by 16S rRNA gene sequencing (GenBank accession number MT365797) and for its whole-genome sequencing (WGS), genomic DNA was extracted according to the protocol for the ZR Quick-DNA fungal/bacterial miniprep kit (Zymo, USA) from approximately 20 mg of cells collected from an overnight culture on tryptic soya medium (37°C).

DNA quantification using a Qubit fluorometer and genome library preparation of the isolated DNA were conducted at FISABIO (Valencia, Spain). The DNA library preparation was performed using the Nextera XT library prep kit (FC-131-1024), and multiplexing was conducted using a Nextera XT index kit (FC-131-1096) (Illumina, USA). Sequencing of the paired-end reads was performed on an Illumina MiSeq instrument with 2 × 300-bp cycles.

The quality of the sequencing reads and draft genome were obtained using the QUAST version 5.0.2 evaluation tool (6, 7). In total, 5,467,278 paired-end raw reads were trimmed to 5,363,381 reads using FASTP version 0.20.1 (8), followed by assembly using Unicycler version 0.4.8.0 (9). The assembled contigs were annotated using the Rapid Annotations Subsystems Technology toolkit (RASTtk) version 1.073 (10, 11) on the Pathosystems Resource Integration Center (PATRIC) server version 3.6.9. Default parameters were used for all software.

The data obtained from RASTtk revealed 4,086 PATRIC protein-coding sequences, 3 rRNA genes, and 76 tRNA genes. The genome sequence comprised 4,003,399 bp in total, with an average short read coverage of 592.994×, and contained 28 contigs,

with a GC content of 46.32%. The longest contig was 1,098,879 bp, and it had an N_{50} value of 1,016,292 bp. There were 3,339 proteins with assigned functions and 747 hypothetical proteins. The reference genome used for functional annotation with PATRIC was *Bacillus amyloliquefaciens* strain Fito_F321. Plasmid-related sequences and *Bacillus* toxin genes were absent. Using PATRIC version 3.6.9 (12), the National Database of Antibiotic Resistant Organisms (NDARO) (13), and the Comprehensive Antibiotic Resistance Database (CARD) version 3.1.3 (14), built into RASTtk, lincomycin and tetracycline antibiotic resistance genes were identified. Significantly, using PATRICbrc, genes were identified which code for the dye decolorization enzymes flavin mononucleotide (FMN)-dependent NADH azoreductase (EC 1.7.1.6) (PLF_1386_00006331), azoreductase (EC 1.7.1.6) (PLF_1386_00005892), and the spore coat laccase CotA (EC 1.10.3.2) (PLF_1386_00311525).

Data availability. The draft genome sequence of *Bacillus amyloliquefaciens* AD20 was deposited in PATRICbrc under the accession number 1390.742, and its project data are available under BioProject accession number PRJNA734188, BioSample accession number SAMN19487240, and SRA accession number SRR14875335.

REFERENCES

1. Kulimushi PZ, Basime GC, Nachigera GM, Thonart P, Ongena M. 2018. Efficacy of Bacillus amyloliquefaciens as biocontrol agent to fight fungal diseases of maize under tropical climates: from lab to field assays in south Kivu. Environ Sci Pollut Res Int 25:29808–29821. https://doi.org/10.1007/s11356-017-9314-9.
2. Lu L, Zhao M, Li G-F, Li J, Wang T-N, Li D-B, Xu T-F. 2012. Decolorization of synthetic dyes by immobilized spore from Bacillus amyloliquefaciens. Catal Commun 26:58–62. https://doi.org/10.1016/j.catcom.2012.04.024.
3. Wang J, Lu L, Feng F. 2017. Improving the indigo carmine decolorization ability of a Bacillus amyloliquefaciens laccase by site-directed mutagenesis. Catalysts 7:275. https://doi.org/10.3390/catal7090275.
4. El-Bendary MA, Ezzat SM, Ewais EA, Al-Zalama MA. 2021. Optimization of spore laccase production by Bacillus amyloliquefaciens isolated from wastewater and its potential in green biodecolorization of synthetic textile dyes. Prep Biochem Biotechnol 51:16–27. https://doi.org/10.1080/10826068.2020.1786698.
5. Liu Y, Shao Z, Reng X, Zhou J, Qin W. 2021. Dye decolorization of a newly isolated strain Bacillus amyloliquefaciens W36. World J Microbiol Biotechnol 37:8. https://doi.org/10.1007/s11274-020-02974-4.
6. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics 34:i142–i150. https://doi.org/10.1093/bioinformatics/bty266.
7. 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.
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, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595.
10. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, III, Stevens R, Vonstein V, Wattam AR, Xia F. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep 5:8365. https://doi.org/10.1038/srep08365.
11. Davis JJ, Wattam AR, Aziz RK, Brettin T, Butler R, Butler RM, Chlenski P, Conrad N, Dickerman A, Dietrich EM, Gabbard JL, Gerdes S, Guard A, Kenyon RW, Machi D, Mao C, Murphy-Olson D, Nguyen M, Nordberg EK, Olsen GJ, Olson RD, Overbeek JC, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomas C, VanOeffelen M, Vonstein V, Warren AS, Xia F, Xie D, Yoo H, Stevens R. 2020. The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. Nucleic Acids Res 48:D606–D612. https://doi.org/10.1093/nar/gkz943.
12. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ, De Pascale G, Ejim L, Kalan L, King AM, Koteva K, Morar M, Mulvey MR, O'Brien JS, Pawlowski AC, Piddock LJV, Spanogiannopoulos P, Sutherland AD, Tang I, Taylor PL, Thaker M, Wang W, Yan M, Yu T, Wright GD. 2013. The Comprehensive Antibiotic Resistance Database. Antimicrob Agents Chemother 57:3348–3357. https://doi.org/10.1128/AAC.00419-13.
13. Sayers EW, Beck J, Brister JR, Bolton EE, Canese K, Comeau DC, Funk K, Ketter A, Kim S, Kimchi A, Kitts PA, Kuznetsov A, Lathrop S, Lu Z, McGarvey K, Madden TL, Murphy TD, O'Leary N, Phan L, Schneider VA, Thibaud-Nissen F, Trawick BW, Pruitt KD, Ostell J. 2020. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 48:D9–D16. https://doi.org/10.1093/nar/gkz899.
14. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen A-LV, Cheng AA, Liu S, Min SY, Miroshnichenko A, Tran H-K, Werfalli RE, Nasir JA, Oloni M, Speicher DJ, Florescu A, Singh B, Faltyn M, Hernandez-Koutoucheva A, Sharma AN, Bordeleau E, Pawlowski AC, Zubyk HL, Dooley D, Griffiths E, Maguire F, Winsor GL, Beiko RG, Brinkman FSL, Hsiao WWL, Domselaar GV, McArthur AG. 2020. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res 48:D517–D525. https://doi.org/10.1093/nar/gkz935.