



# Complete Genome Sequence of the Carotenoid-Producing Strain *Gordonia ajoucoccus* A2

Soon Jae Kwon,<sup>a,b</sup> Yong Jin Choi,<sup>a,b</sup> Ju Min Kim,<sup>c,d</sup>  Pyung Cheon Lee<sup>a,b</sup>

<sup>a</sup>Department of Molecular Science and Technology, Ajou University, Woncheon-dong, Yeongtong-gu, Suwon, South Korea

<sup>b</sup>Department of Applied Chemistry and Biological Engineering, Ajou University, Woncheon-dong, Yeongtong-gu, Suwon, South Korea

<sup>c</sup>Department of Energy Systems Research, Ajou University, Suwon, South Korea

<sup>d</sup>Department of Chemical Engineering, Ajou University, Suwon, South Korea

Ju Min Kim and Pyung Cheon Lee contributed equally to this work. Author order was determined on the basis of seniority.

**ABSTRACT** *Gordonia ajoucoccus* strain A2, isolated from crude oil-contaminated soils, synthesizes yellow keto- $\gamma$ -carotene from various *n*-alkanes as the sole carbon source. Its complete genome sequence consists of a single circular chromosome (5,090,254 bp, 67.3% G+C content). Seven putative genes were identified supporting the proposed keto- $\gamma$ -carotene pathway of *G. ajoucoccus* A2.

Previously (1), we isolated and characterized a new strain, *Gordonia ajoucoccus* A2, which can degrade *n*-alkanes of variable chain lengths from oil-contaminated soil. *G. ajoucoccus* A2 produced carotenoids when grown in media containing *n*-alkanes as the sole carbon sources.  $\gamma$ -Carotene and keto- $\gamma$ -carotene are the main carotenoids in *G. ajoucoccus* A2 (1). Carotenoids are widely used as cosmetic ingredients, antioxidants, and food or feed additives (2–4). Studying *G. ajoucoccus* A2—including whole-genome sequencing—to validate it as a carotenoid producer is worthwhile. These genomic data can form the basis for elucidation of carotenoid biosynthesis mechanisms and for metabolic engineering of *Gordonia* strains (5–8) suitable for large-scale fermentative production of other biotechnologically important carotenoids from industrial discharge or other sources of *n*-alkanes.

The strain was isolated from a local Korean chemical company (Seok-Rim Chemicals) and aerobically cultured at 30°C in a 125-ml serum bottle containing 20 ml of mineral salt medium (1) with 0.5% (wt/vol) fructose as the sole carbon source for 3 days. Genomic DNA was extracted via a genomic DNA kit (MacroGen, South Korea) with RNase A treatment and sequenced in single-molecule real-time (SMRT) sequencing cells using PacBio RS II SMRT technology (Pacific Biosciences, CA) on the HiSeq 2000 platform (Illumina, USA). Both procedures were performed by DNA Link, Inc. (Seoul, South Korea). The sequencing libraries were prepared using the SMRTbell template prep kit 1.0 (PacBio RS II) and the TruSeq Nano DNA kit (Illumina). All software was run with default settings unless stated otherwise. After subread filtering of the PacBio raw data, 141,490 long reads with an average length of 11,140 bp (total, 1,130,407,021 bp; genome coverage, >220-fold; mean read score, 0.829) were generated and *de novo* assembled in the Canu v1.3 (9) assembler with the parameter genomeSize=5m. The overlapping regions at both ends of one contig were trimmed to create unique stretches on both ends using Circlator (10) (b2r\_length\_cutoff=60000, 100000, or 200000). The resulting one contig was error corrected in Quiver (11) for three cycles. The error-corrected assembly was further polished using Pilon v1.22 (--fix bases) (12) with trimmed paired-end reads (total, 8,082,491 reads; genome coverage, >126-fold), which were obtained from 2 × 251-bp paired-end reads (8,427,944 reads totaling 2,115,400,630 bp) in Sickle v1.33 (<https://github.com/najoshi/sickle>). The assembly sta-

**Citation** Kwon SJ, Choi YJ, Kim JM, Lee PC. 2020. Complete genome sequence of the carotenoid-producing strain *Gordonia ajoucoccus* A2. Microbiol Resour Announc 9:e00662-20. <https://doi.org/10.1128/MRA.00662-20>.

**Editor** Julie C. Dunning Hotopp, University of Maryland School of Medicine

**Copyright** © 2020 Kwon et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Ju Min Kim, [jumin@ajou.ac.kr](mailto:jumin@ajou.ac.kr), or Pyung Cheon Lee, [pclee@ajou.ac.kr](mailto:pclee@ajou.ac.kr).

**Received** 8 June 2020

**Accepted** 21 August 2020

**Published** 10 September 2020

tistics were calculated using stats.sh from BBmap v38.68 (<https://sourceforge.net/projects/bbmap/>). Genome annotation and gene prediction were conducted using the Prokaryotic Genome Annotation Pipeline (PGAP) (13).

The genome consists of a single 5,090,254-bp circular chromosome with 67.3% G+C content. Annotation revealed 4,825 coding DNA sequences and 110 encoded RNAs (9 rRNAs and 47 tRNAs). Seven genes for keto- $\gamma$ -carotene biosynthesis were predicted in the genome, namely, one gene encoding geranylgeranyl diphosphate synthase (CrtE), one encoding phytoene synthase (CrtB), two encoding phytoene dehydrogenase (CrtI), one encoding lycopene  $\gamma$ -cyclase (CrtL), and two genes encoding  $\gamma$ -carotene ketolase (CrtO).

**Data availability.** This whole-genome shotgun project has been deposited in DDBJ/EMBL/GenBank under the accession number [CP052884.1](#). The SRA/DRA/ERA accession numbers are [SRR11612832](#) (PacBio) and [SRR11612831](#) (Illumina). The BioSample and BioProject numbers are [SAMN14731745](#) and [PRJNA628123](#), respectively.

## ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (grant number 2012M1A2A2026562) and the Priority Research Centers Program through the National Research Foundation of Korea (grant number 2019R1A6A11051471). This research was also partially supported by the Ajou University research fund. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

## REFERENCES

- Kim JH, Kim SH, Yoon JH, Lee PC. 2014. Carotenoid production from *n*-alkanes with a broad range of chain lengths by the novel species *Gordonia ajoucoccus* A2T. *Appl Microbiol Biotechnol* 98:3759–3768. <https://doi.org/10.1007/s00253-014-5516-y>.
- Lee P, Schmidt-Dannert C. 2002. Metabolic engineering towards biotechnological production of carotenoids in microorganisms. *Appl Microbiol Biotechnol* 60:1–11. <https://doi.org/10.1007/s00253-002-1101-x>.
- Kim SH, Lee JM, Kim SC, Park CB, Lee PC. 2014. Proposed cytotoxic mechanisms of the saffron carotenoids crocin and crocetin on cancer cell lines. *Biochem Cell Biol* 92:105–111. <https://doi.org/10.1139/bcb-2013-0091>.
- Kim SH, Kim MS, Lee BY, Lee PC. 2016. Generation of structurally novel short carotenoids and study of their biological activity. *Sci Rep* 6:21987. <https://doi.org/10.1038/srep21987>.
- Choi JY, Lee K, Lee PC. 2019. Complete genome sequence of the carotenoid-producing *Deinococcus* sp. strain AJ005. *Microbiol Resour Announc* 8:e01245-19. <https://doi.org/10.1128/MRA.01245-19>.
- Choi JY, Lee K, Lee PC. 2019. Characterization of carotenoid biosynthesis in newly isolated *Deinococcus* sp. AJ005 and investigation of the effects of environmental conditions on cell growth and carotenoid biosynthesis. *Mar Drugs* 17:705. <https://doi.org/10.3390/md17120705>.
- Lee JH, Kim JW, Lee PC. 2018. Complete genome sequence of *Flavobacterium kingsejongi* WV39, a type species of the genus *Flavobacterium* and a microbial C40 carotenoid zeaxanthin producer. *J Biotechnol* 266:9–13. <https://doi.org/10.1016/j.jbiotec.2017.11.012>.
- Kim JW, Choi BH, Kim JH, Kang H-J, Ryu H, Lee PC. 2018. Complete genome sequence of *Planococcus faecalis* AJ003<sup>T</sup>, the type species of the genus *Planococcus* and a microbial C30 carotenoid producer. *J Biotechnol* 266:72–76. <https://doi.org/10.1016/j.jbiotec.2017.12.005>.
- 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>.
- Hunt M, De Silva N, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol* 16:294. <https://doi.org/10.1186/s13059-015-0849-0>.
- Chin C-S, 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>.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.
- Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. *Nucleic Acids Res* 44: 6614–6624. <https://doi.org/10.1093/nar/gkw569>.