JB Journals.ASM.org

GENOME ANNOUNCEMENT

Complete Genome Sequence of *Amycolatopsis mediterranei* S699 Based on *De Novo* Assembly via a Combinatorial Sequencing Strategy

Biao Tang,[a] Wei Zhao,[a,b] Huajun Zheng,[c] Ying Zhuo,[d] Lixin Zhang,[d] and Guo-Ping Zhao[a,c,e,f]

State Key Laboratory of Genetic Engineering, Department of Microbiology, School of Life Sciences and Institute of Biomedical Sciences, Fudan University, Shanghai, China[a]; China HYK Gene Technology Company Ltd., Guangdong, China[b]; Shanghai-MOST Key Laboratory of Disease and Health Genomics, Chinese National Human Genome Center at Shanghai, Shanghai, China[c]; CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China[d]; Department of Microbiology and Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong SAR, China[e]; and Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China[f]

The genome of *Amycolatopsis mediterranei* S699 was resequenced and assembled *de novo*. By comparing the sequences of S699 previously released and that of *A. mediterranei* U32, about 10 kb of major indels was found to differ between the two S699 genomes, and the differences are likely attributable to their different assembly strategies.

Isolated in 1957 at St. Raphael, France (8), *Amycolatopsis mediterranei* original strain ATCC 13685 synthesizes a collection of rifamycin complexes, while a single fermentation product of rifamycin B, whose derivatives are effectively used to treat mycobacterial infections (11), could be obtained only if sodium diethyl barbiturate was added to the medium (9). In 1975, a mutant strain designated ATCC 21789 that is able to produce a sole product of rifamycin B without the addition of barbiturate was isolated (11), and a related strain, S699 (5), has been widely used in laboratory research ever since (1, 2, 4, 6).

In 2011, the whole-genome sequence of *A. mediterranei* S699 (CP002896) presenting a 10,236,779-bp chromosome, assembled by mapping the shotgun sequences to the 10,236,715-bp genome of a rifamycin SV-producing strain, *A. mediterranei* U32 (CP002000) (14), was released (10). In this project, an S699 strain obtained from Giangarlo Lancini at Lepetit Laboratories, Geranzano, Italy, was resequenced via a combinatorial strategy and *de novo* assembly.

A total of 589,827 reads were generated using a Roche 454 GS FLX Titanium system and were assembled using the Newbler Program (version 2.3), which resulted in 67 contigs with an average size of 152,345 bp. Sanger-based sequencing was employed to facilitate gap closing, to amend the low-quality regions (score < 40), and to verify the variations between the draft sequence and the corresponding genome regions of other *A. mediterranei* strains. Finally, a consensus sequence containing 10,246,920 bp with an estimated error rate of <0.5 per 100,000 bases providing 31-fold coverage was acquired.

Beside a few different sequences predicted by Glimmer (3) and Genemark (7), most of the S699 genome was annotated by BLASTN based on the annotation of U32 followed by BLASTP functional assignment using KEGG and NR databases. A total of 9,227 protein-coding gene loci (*AMES_CDS*s) with an average length of 989 bp were identified.

Compared to the S699 (CP002896) genome, 218 single nucleotide polymorphisms (SNPs) and 51 indels were found in our S699 (CP003729) sequence. All 12 large indels (>40 bp and mostly repeat sequences) are insertions compared to not only the S699 (CP002896) genome but also that of U32. Except for the three insertions that failed to be specifically amplified by PCR, we confirmed that all of the other nine insertions, including a 7.5-kb insertion, were present in the genomes of ATCC 13685 and ATCC 21789. Therefore, we propose that the major indel variations between these two S699 genomic sequences could be caused by their different assembly strategies (10).

In comparison to the U32 genome, 234 SNPs and 48 indels were identified in our S699 (CP003729) genome. In particular, in the *rif* cluster, unlike that of U32, the Rif-Orf16 (cytochrome P450) encoded by gene locus *AMES_0651* essential for the conversion of SV to B is a wild-type protein, as is the case in other rifamycin B-producing strains (13, 14). In addition, a frameshift of an aminotransferase (Rif-Orf9) was found at the 3′ region of its encoding *AMES_0639* gene locus, but it may not affect the production of rifamycin (1, 12).

Nucleotide sequence accession number. The genome sequence of *A. mediterranei* S699 in this project has been deposited in the GenBank database under accession number CP003729.

ACKNOWLEDGMENTS

This work was supported by grants from the National Natural Science Foundation of China (30830002) and the Postdoctoral Science Foundation of China (2012M510787). L.Z. is an awardee for the National Distinguished Young Scholar Program in China.

REFERENCES

1. **August PR, et al.** 1998. Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the rif biosynthetic gene cluster of *Amycolatopsis mediterranei* S699. Chem. Biol. 5:69–79.
2. **Chiao JS, Xia T, Mei BG, Jin ZK, Gu WL.** 1996. Rifamycin SV and related ansamycins, regulation of biosynthesis, p 477–498. *In* Vining LC, Stuttard C (ed), Genetics and biochemistry of antibiotic production. Butterworth-Heinemann, Boston, MA.
3. **Delcher AL, Harmon D, Kasif S, White O, Salzberg SL.** 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. 27: 4636–4641.
4. **Doi-Katayama Y, et al.** 2000. Thioesterases and the premature termina-

Received 23 July 2012 **Accepted** 6 August 2012

Address correspondence to Guo-Ping Zhao, gpzhao@sibs.ac.cn.

B.T. and W.Z. contributed equally to this article.

doi:10.1128/JB.01295-12

tion of polyketide chain elongation in rifamycin B biosynthesis by *Amycolatopsis mediterranei* S699. J. Antibiot. (Tokyo) **53**:484–495.
5. **Kim CG, et al.** 1992. Formation of 3-amino-5-hydroxybenzoic acid, the precursor of mC7N units in ansamycin antibiotics, by a new variant of the shikimate pathway. J. Am. Chem. Soc. **114**:4941–4943.
6. **Lal R, et al.** 1995. Rifamycins: strain improvement program. Crit. Rev. Microbiol. **21**:19–30.
7. **Lukashin AV, Borodovsky M.** 1998. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. **26**:1107–1115.
8. **Margalith P, Beretta G.** 1960. Rifomycin. XI. taxonomic study on *Streptomyces mediterranei* nov. sp. Mycopathol. Mycol. Appl. **13**:321–330.
9. **Margalith P, Pagani H.** 1961. Rifomycin. XIV. Production of rifomycin B. Appl. Microbiol. **9**:325–334.
10. **Verma M, et al.** 2011. Whole genome sequence of the rifamycin B-producing strain *Amycolatopsis mediterranei* S699. J. Bacteriol. **193**:5562–5563.
11. **White RJ, Martinelli E, Lancini G.** 1974. Ansamycin biogenesis: studies on a novel rifamycin isolated from a mutant strain of *Nocardia mediterranei.* Proc. Natl. Acad. Sci. U. S. A. **71**:3260–3264.
12. **Xu J, Wan E, Kim CJ, Floss HG, Mahmud T.** 2005. Identification of tailoring genes involved in the modification of the polyketide backbone of rifamycin B by *Amycolatopsis mediterranei* S699. Microbiology **151**:2515–2528.
13. **Yuan H, et al.** 2011. Two genes, rif15 and rif16, of the rifamycin biosynthetic gene cluster in *Amycolatopsis mediterranei* likely encode a transketolase and a P450 monooxygenase, respectively, both essential for the conversion of rifamycin SV into B. Acta Biochim. Biophys. Sin. **43**:948–956.
14. **Zhao W, et al.** 2010. Complete genome sequence of the rifamycin SV-producing *Amycolatopsis mediterranei* U32 revealed its genetic characteristics in phylogeny and metabolism. Cell Res. **20**:1096–1108.