Genome Sequence of *Aeromicrobium erythreum* NRRL B-3381, an Erythromycin-Producing Bacterium of the *Nocardioidaceae*

Erin A. Harrell, Eric S. Miller
Department of Plant & Microbial Biology, North Carolina State University, Raleigh, North Carolina, USA

*Aeromicrobium erythreum* NRRL B-3381 has a 3,629,239-bp circular genome that has 72% G+C content. There are at least 3,121 coding sequences (CDSs), two rRNA gene operons, and 47 tRNAs. The genome and erythromycin (*ery*) biosynthetic gene sequences provide resources for metabolic and combinatorial engineering of polyketides.

At*throbacter* sp. strain NRRL B-3381 was part of a 1970 U.S. patent issued for an erythromycin process (1). Unlike other erythromycin processes, the NRRL B-3381 strain (isolated from Lajas Valley, Cabo Rojo, Puerto Rico) was notable for producing only erythromycin A and not the related compounds erythromycin B and C. From a large collection of industrially relevant actinobacteria, Sydney Brenner, then of the Medical Research Council (MRC) Molecular Genetics Unit, initiated a research program to genetically manipulate this nonfilamentous bacterium for polyketide combinatorial chemistry (2, 3). Since then, strain NRRL B-3381 has been taxonomically reclassified as the type genus and species *Aeromicrobium erythreum* (4), methods of plasmid transformation and gene disruption were developed (3, 5), and cosmid clones of *ery* (erythromycin) genes were isolated and sequenced (6). Although interesting metabolic manipulations of *A. erythreum* have been performed (7, 8), extensive uses of its polyketide synthase and other erythromycin biosynthesis genes have not been reported. Access to the complete genome sequence of the NRRL B-3381 strain may facilitate macrolide antibiotic development and other biotechnological uses of this and related Actinobacteria (9).

Total DNA was prepared using the Qiagen Gentra Puregene yeast/bact kit with overnight cultures of *A. erythreum* (collection strain designated AR18) grown at 30°C in 2xYT medium with shaking. Ten micrograms of purified DNA was processed with the strain designated AR18) grown at 30°C in 2xYT medium with shaking. Ten micrograms of purified DNA was processed with the University Genomic Sciences Laboratory and then analyzed by Pacific Biosciences (PacBio) 10-kbp library kit in the NC State Laboratory; George Yuan and Roberto Lleras of Pacific Biosciences; the NCBI Genomes Laboratory; and then analyzed by the National Center for Biotechnology Information (NCBI) GenBank database. The genome assembly was done with PacBio SMRT HGAP 2/Quiver using 75,640 single-molecule real-time (SMRT) RS II sequencing. Genome assembly and other biotechnological uses of this and related Actinobacteria (9).

The 3.6-Mbp genome was annotated with the NCBI Prokaryotic Genome Annotation Pipeline (10), with some manual curation. There are 3,121 CDSs, 270 pseudogenes needing further analysis, two rRNA operons (16S, 23S, and 5S), at least 47 tRNAs, and one potential transfer-messenger RNA (tmRNA). Methylation kinetics revealed m6A (N6-methyladenine), primarily at CTCCAG and CTGGAG (a BpmI-like site).

Sequences of the erythromycin-related genes, including those encoding the methyltransferase resistance enzyme (*ermR*) and the 65-kbp *ery* gene cluster, are essentially as previously reported (2, 6) (accession no. AY626585). The *ery* gene cluster includes three polyketide (6-desoxyerythronolide B) synthase modules (*eryAI—AIII*) and the methyltransferases, dehydratase, isomerase, sugar transferases, etc. leading to erythromycin *Aeromicrobium ma*- rinum DSM 15272 (accession no. CM001024) encodes the most proteins currently orthologous to proteins of *A. erythreum*, and orthologs are also found from *Nocardioidea simplex VKM Ac-203SD (11).

The available *A. erythreum* NRRL B-3381 genome sequence should provide a resource for comparative and evolutionary genomics and, as suggested by Sydney Brenner 30 years ago, facilitate metabolic and combinatorial engineering of polyketide biosynthesis in a genetically tractable unicellular actinobacterium.

**Nucleotide sequence accession number.** The genome sequence has been deposited in GenBank with accession no. CP011502. The version described in this paper is the first version.

**ACKNOWLEDGMENTS**

We thank Cory Dashiell and Jenn Schaff of the NCSU Genomic Sciences Lab; George Yuan and Roberto Lleras of Pacific Biosciences; the NCBI Prokaryotic Genomes Annotation Pipeline group; Mark Weber of Ferma-logic, Inc. for discussions; and Sydney Brenner for wit, wisdom, and innovation.

**REFERENCES**


