genomeA
Journals.ASM.org

Draft Genome Sequence of *Bacillus thuringiensis* Strain DAR 81934, Which Exhibits Molluscicidal Activity

Aisuo Wang,[a,b] Julie Pattemore,[a,b] Gavin Ash,[a,b] Angela Williams,[c] James Hane[c]

School of Agricultural and Wine Sciences, Charles Sturt University, Wagga Wagga, NSW, Australia[a]; EH Graham Centre for Agricultural Innovation, Charles Sturt University, Wagga Wagga, NSW, Australia[b]; Black Box Informatics, Perth, WA, Australia[c]

Bacillus thuringiensis **has been widely used as a biopesticide for a long time. Its molluscicidal activity, however, is rarely realized. Here, we report the genome sequence of *B. thuringiensis* strain DAR 81934, a strain with molluscicidal activity against the pest snail *Cernuella virgata*.**

Received 5 December 2012 **Accepted** 25 February 2013 **Published** 21 March 2013

Citation Wang A, Pattemore J, Ash G, Williams A, Hane J. 2013. Draft genome sequence of *Bacillus thuringiensis* strain DAR 81934, which exhibits molluscicidal activity. Genome Announc. 1(2):e00175-12. doi:10.1128/genomeA.00175-12.

Address correspondence to Gavin Ash, GAsh@csu.edu.au.

Bacillus thuringiensis is widely used as a biopesticide due to its capability to produce insecticidal crystal proteins (ICPs) (1). However, a new *B. thuringiensis* strain, DAR 81934, isolated in our laboratory, demonstrated molluscicidal activity toward the pest snail *Cernuella virgata* (data not shown), which confirms the findings of other researchers (2, 3). To further study the genetic basis of this *B. thuringiensis* strain, we sequenced its whole genome. The strain has been lodged in the Plant Pathology Herbarium (DAR) culture collection, Orange, NSW, Australia (strain DAR 81934).

The genome of *B. thuringiensis* DAR 81934 (Bt 81934) was sequenced at the Australian Genome Research Facility (AGRF) using an Illumina HiSeq 2000 instrument. The total novel isolate reads were aligned to three reference genomes (those of *B. thuringiensis* serovar Konkukian strain 97-27, *B. thuringiensis* strain Al Hakam, and *B. thuringiensis* BMB171) via the Burrows-Wheeler Aligner (BWA) (parameters: -o3 -e3 -d5 -i5 -R50) (4). Local realignment around indels was performed with the Genome Analysis Toolkit (GATK) v1.5.20 (5). Regions of >50× coverage were used as the basis for contigs. Contigs were first scaffolded via read-pairing relationships with SSPACE 2.0 (6) and then scaffolded via Optimal Syntenic Layout of unfinished assemblies (OSLay) (7). Scaffolding gaps were closed using the Beijing Genomics Institute (BGI) GapCloser 1.2. Novel isolate reads that were not aligned to the three reference genomes were *de novo* assembled with Velvet 1.2 using a kmer length of 41 bp (8). All the assemblies were combined via HaploMerger (9). As haplotype merging can potentially introduce single nucleotide polymorphism (SNP)-like assembly errors at merged sites (10), raw reads were back-aligned to the final assembly and the sequence consensus was confirmed via GATK (5).

The 5.94-Mb genome of Bt 81934 contains two components: a 5.69-Mb chromosome sequence (scaffolds 1 to 9), and a 0.25-Mb plasmid sequence (scaffolds 10 to 11). The average G+C content of the chromosome sequence is 33.67%, while that of the plasmid sequence is 32.75%. Protein-coding genes were predicted *in silico* via GeneMark-S (11), producing 6,042 genes, with 5,797 genes in the chromosome and 248 genes in the plasmid. tRNA and rRNA genes were identified by tRNAscan-SE (12) and RNAmmer (13), respectively. The whole genome contains 73 tRNA genes and 24 rRNA genes (all in scaffold 1). Bt toxin genes were predicted via BtToxinScanner (14). Two *cry* genes were identified in scaffold 10. Additionally, the BtToxin_Scanner database of *cry*, *cyt*, and *vip* genes was compared to the predicted protein dataset and scaffold sequences via BLASTp and tBLASTn, respectively. This produced matches to a further four toxin candidate genes in scaffolds 1 and 3.

In summary, this is the first report for the genome sequence of a *B. thuringiensis* strain with molluscicidal activity. The genome data indicate that Bt 81934 harbors *cry* and *vip* genes not only in the plasmid sequence, but also in the chromosome sequence. The availability of the genome data will facilitate the understanding of Bt endotoxin protein production and the genetic basis of its molluscicidal activity.

Nucleotide sequence accession numbers. The draft genome sequence for Bt 81934 has been included in the GenBank Whole-Genome Shotgun (WGS) database under the accession no. ANPK01000001 to ANPK01000083.

ACKNOWLEDGMENT

This research was funded by Grains Research & Development Corporation (GRDC) in Australia (grant no. UCS00013).

REFERENCES

1. **Höfte H, Whiteley HR.** 1989. Insecticidal crystal proteins of *Bacillus thuringiensis*. Microbiol. Rev. **53**:242–255.
2. **Bahy AA, Salem HH, Wang XM, Huang TH, Xie QD, Zhang XY.** 2010. Effect of *Bacillus thuringiensis* var. *israelensis* endotoxin on the intermediate snail host of *Schistosoma japonicum*. Curr. Res. Bacteriol. **3**:37–41.
3. **Halima HS, Bahy AA, Huang TH, Xie QD.** 2006. Molecular characterization of novel *Bacillus thuringiensis* isolate with molluscicidal activity against the intermediate host of schistosomes. Biotechnology **5**:413–420.
4. **Li H, Durbin R.** 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics **26**:589–595.
5. **DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler**

D, Daly MJ. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. **43**:491–498.
6. **Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W.** 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics **27**: 578–579.
7. **Richter DC, Schuster SC, Huson DH.** 2007. OSLay: optimal syntenic layout of unfinished assemblies. Bioinformatics **23**:1573–1579.
8. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. **18**:821–829.
9. **Huang S, Chen Z, Huang G, Yu T, Yang P, Li J, Fu Y, Yuan S, Chen S, Xu A.** 2012. HaploMerger: reconstructing allelic relationships for polymorphic diploid genome assemblies. Genome Res. **22**:1581–1588.
10. **Schwartz R.** 2010. Theory and algorithms for the haplotype assembly problem. Commun. Inf. Syst. **10**:23–38.
11. **Besemer J, Lomsadze A, Borodovsky M.** 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. **29**:2607–2618.
12. **Lowe TM, Eddy SR.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25**: 955–964.
13. **Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35**:3100–3108.
14. **Ye W, Zhu L, Liu Y, Crickmore N, Peng D, Ruan L, Sun M.** 2012. Mining new crystal protein genes from *Bacillus thuringiensis* on the basis of mixed plasmid-enriched genome sequencing and a computational pipeline. Appl. Environ. Microbiol. **78**:4795–4801.