Draft Genome Sequence of *Curtobacterium* sp. Strain ER1/6, an Endophytic Strain Isolated from *Citrus sinensis* with Potential To Be Used as a Biocontrol Agent

Leandro Maza Garrido,[a] João Marcelo Pereira Alves,[b] Liliane Santana Oliveira,[b] Arthur Gruber,[b] Gabriel Padilla,[a] Welington Luiz Araújo[a]

NAP/BIOP, Departamento de Microbiologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, Biomédicas II, Cidade Universitária, São Paulo, São Paulo, Brazil[a]; Departamento de Parasitologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, Biomédicas II, Cidade Universitária, São Paulo, São Paulo, Brazil[b]

Herein, we report a draft genome sequence of the endophytic *Curtobacterium* sp. strain ER1/6, isolated from a surface-sterilized *Citrus sinensis* branch, and it presented the capability to control phytopathogens. Functional annotation of the ~3.4-Mb genome revealed 3,100 protein-coding genes, with many products related to known ecological and biotechnological aspects of this bacterium.

Address correspondence to Arthur Gruber, argruber@usp.br, or Welington Luiz Araújo, wlaraujo@usp.br.

The plant environment is an important habitat for *Curtobacterium* species that may interact with their host plant, inducing disease symptoms or controlling phytopathogens. This genus was first described by Yamada and Komagata (1) and includes eight recognized *Curtobacterium* species (*C. albidum, C. ammoniigenes, C. citreum, C. flaccumfaciens, C. ginsengisoli, C. herbarum, C. luteum,* and *C. pusillum*) within the phylum *Actinobacteria*. The citrus endophytic *Curtobacterium* sp. strain ER1/6 was isolated from *Citrus sinensis* (sweet orange) and previously identified by fatty acid methyl ester (FAME-MIDI) as *C. flaccumfaciens* (2). Further analysis showed that this strain is a divergent genotype which lacks the ability to induce disease symptoms in beans and ornamental seedlings. In addition, a previous study showed that this endophytic bacterium may inhibit the citrus variegated chlorosis (CVC) symptoms caused by *Xylella fastidiosa* (3). Thus, based on its phenotypic and physiological properties, as well as phylogenetic distinctiveness, this citrus endophytic strain might represent a novel species in the *Curtobacterium* genus.

The genomic DNA, isolated from an overnight culture in tryptic soy broth (TSB) using a commercial Wizard genomic DNA purification kit (Promega Co.), was sequenced using the Illumina HiSeq 2000 platform in order to reach about 300-fold depth of coverage. A total of 1,074,582,228 bp were generated from a 100-bp paired-end run, with 97.5% of bases showing quality scores above 20. Paired-end data sets merging, trimming, and *de novo* assembly were performed using Geneious (version 7.1.9; Biomatters). The final assembly comprised 19 large contigs (>1,000 bp) totaling 3,368,973 bp in length and presenting a G+C content of 72.2%.

The assembled genomic sequence was submitted to a functional annotation pipeline constructed with the EGene2 platform (4). Gene predictions were performed using Glimmer 3.02 (5), tRNAscan-SE 1.3 (6), and RNAmmer 1.2 (7). Translated products were submitted to several analyses, including BLAST similarity searches against NCBI's nonredundant (nr) database, identification of protein motifs on InterPro (8), and Conserved Domain Database (CDD) (9), orthology mapping using eggNOG (10) and KEGG Orthology (KO), KEGG pathway mapping (11), and GO (Gene Ontology) term mapping (12). We found a total 3,239 genes, including 3,100 protein-coding genes, 87 noncoding RNAs (ncRNAs), and 47 tRNAs, comprising a complete set for the 20 amino acids. In addition, genes for rRNAs of the 5S, 16S and 23S subunits were also found. A total of 1,259 and 2,279 protein sequences were successfully classified in KO and eggNOG orthologous groups, respectively. A survey of secondary metabolites pathways using antiSMASH 2.0 (13) revealed six putative gene clusters (two terpenes, one type III-polyketide synthases [PKS], one siderophore, one bacteriocin, and one unknown).

Accession number(s). This communication describes the first version of a whole-genome shotgun project of the citrus endophytic *Curtobacterium* strain ER1/6. The nucleotide sequence and annotation data have been deposited at DDBJ/EMBL/GenBank under the accession number MJAK00000000.

FUNDING INFORMATION

REFERENCES

1. **Yamada K, Komagata K.** 1972. Taxonomic studies of *Coryneform* bacteria V. Classification of *Coryneform* bacteria. J Gen Appl Microbiol **18:** 417–431. http://dx.doi.org/10.2323/jgam.18.417.

2. **Araújo WL, Marcon J, Maccheroni W, Jr, Van Elsas JD, Van Vuurde JW, Azevedo JL.** 2002. Diversity of endophytic bacterial populations and their interaction with *Xylella fastidiosa* in citrus plants. Appl Environ Microbiol **68:**4906–4914. http://dx.doi.org/10.1128/AEM.68.10.4906-4914.2002.

3. **Lacava PT, Li W, Araújo WL, Azevedo JL, Hartung JS.** 2007. The endophyte *Curtobacterium flaccumfaciens* reduces symptoms caused by *Xylella fastidiosa* in *Catharanthus roseus*. J Microbiol **45:**388–393.

4. **Durham AM, Kashiwabara AY, Matsunaga FT, Ahagon PH, Rainone F, Varuzza L, Gruber A.** 2005. EGene: a configurable pipeline generation system for automated sequence analysis. Bioinformatics **21:**2812–2813. http://dx.doi.org/10.1093/bioinformatics/bti424.

5. **Delcher AL, Harmon D, Kasif S, White O, Salzberg SL.** 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res **27:**4636–4641. http://dx.doi.org/10.1093/nar/27.23.4636.

6. **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. http://dx.doi.org/10.1093/nar/25.5.0955.

7. **Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res **35:**3100–3108. http://dx.doi.org/10.1093/nar/gkm160.

8. **Mitchell A, Chang HY, Daugherty L, Fraser M, Hunter S, Lopez R, McAnulla C, McMenamin C, Nuka G, Pesseat S, Sangrador-Vegas A, Scheremetjew M, Rato C, Yong SY, Bateman A, Punta M, Attwood TK, Sigrist CJ, Redaschi N, Rivoire C, Xenarios I, Kahn D, Guyot D, Bork P, Letunic I, Gough J, Oates M, Haft D, Huang H, Natale DA, Wu CH, Orengo C, Sillitoe I, Mi H, Thomas PD, Finn RD.** 2015. The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Res **43:**D213–D221. http://dx.doi.org/10.1093/nar/gku1243.

9. **Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH.** 2015. CDD: NCBI's conserved domain database. Nucleic Acids Res **43:**D222–D226. http://dx.doi.org/10.1093/nar/gku1221.

10. **Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, Jensen LJ, von Mering C, Bork P.** 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res **44:**D286–D293. http://dx.doi.org/10.1093/nar/gkv1248.

11. **Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M.** 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res **44:**D457–D462. http://dx.doi.org/10.1093/nar/gkv1070.

12. **Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R.** 2009. QuickGO: a Web-based tool for gene ontology searching. Bioinformatics **25:**3045–3046. http://dx.doi.org/10.1093/bioinformatics/btp536.

13. **Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R.** 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res **39:**W339–W346. http://dx.doi.org/10.1093/nar/gkr466.