

Genome Sequence of “*Candidatus Frankia datiscae*” Dg1, the Uncultured Microsymbiont from Nitrogen-Fixing Root Nodules of the Dicot *Datisca glomerata*

Tomas Persson,¹ David R. Benson,² Philippe Normand,³ Brian Vanden Heuvel,⁴ Petar Pujic,³ Olga Chertkov,⁵ Hazuki Teshima,⁵ David C. Bruce,⁵ Chris Detter,⁵ Roxanne Tapia,⁵ Shunsheng Han,⁵ James Han,⁶ Tanja Woyke,⁶ Sam Pitluck,⁶ Len Pennacchio,⁶ Matt Nolan,⁶ Natalia Ivanova,⁶ Amrita Pati,⁶ Miriam L. Land,⁷ Katharina Pawlowski,^{1*} and Alison M. Berry⁸

Department of Botany, Stockholm University, 10691 Stockholm, Sweden¹; Department of Molecular & Cell Biology, University of Connecticut, Storrs, Connecticut 06269-3125²; Université Lyon 1, Université de Lyon, CNRS, Ecologie Microbienne UMR5557, Cedex 69622 Villeurbanne, France³; Department of Biology, Colorado State University, Pueblo, Colorado 81001⁴; Los Alamos National Laboratory, Los Alamos, New Mexico 87544-2587⁵; DOE Joint Genome Institute, Walnut Creek, California 94598⁶; Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831⁷; and Department of Plant Sciences, University of California, Davis, California 95616⁸

Received 26 September 2011/Accepted 28 September 2011

Members of the noncultured clade of *Frankia* enter into root nodule symbioses with actinorhizal species from the orders Cucurbitales and Rosales. We report the genome sequence of a member of this clade originally from Pakistan but obtained from root nodules of the American plant *Datisca glomerata* without isolation in culture.

Frankia strains are filamentous, sporulating, aerobic actinobacteria that induce nitrogen-fixing root nodules on about 220 plant species from eight families in three orders, Fagales, Rosales, and Cucurbitales. Phylogenetic analyses place symbiotic *Frankia* strains into three distinct clusters (7). Genomes of two members of cluster I have been sequenced, *Frankia alni* strain ACN14a (7.50 Mbp; GenBank accession number NC_008278) and the *Casuarina*-infective *Frankia* sp. strain HFPCcI3 (5.43 Mbp; GenBank accession number NC_007777). Three genomes of cluster III representatives have been sequenced, including *Frankia* sp. strains EAN1pec (8.9 Mbp; GenBank accession number NC_009921), EUN1f (9.35 Mbp; GenBank accession number NC_014666), and EuI1c (8.8 Mbp; GenBank accession number ADGX00000000).

Strains of cluster II, which represents the basal clade of the symbiotic *Frankia* strains (2, 7), cannot be cultured despite numerous attempts. Cluster II strains enter into nitrogen-fixing root nodule symbioses with actinorhizal species in the orders Cucurbitales (Datisceae and Coriariaceae) and Rosales (Rosaceae, and *Ceanothus* in the Rhamnaceae). Here, we announce the first genome sequence of a strain from this cluster using DNA from *Frankia* isolated from root nodules (3) of the American suffruticose endemic *Datisca glomerata* (C. Presl) Baill. (Durango root). The *Frankia* strain was originally sampled from soil in Pakistan that was used to infect *Datisca cannabina* and *Coriaria nepalensis* (5), and homogenates of the nodules were later used to inoculate and repeatedly reinoculate *D. glomerata* in greenhouses at Göttingen and Stockholm Universities. Based on the recommendation of Murray and

Stackebrandt (6), we propose naming this strain “*Candidatus Frankia datiscae*” Dg1.

The finished genome of Dg1 was generated at the Joint Genome Institute using a combination of Illumina (1) and 454 (4) technologies. One Illumina GAII shotgun library, one 454 Titanium standard library, and one paired-end 454 library were constructed and sequenced. For finishing, gaps and misassemblies were resolved by sequencing cloned PCR fragments.

The genome sequence of Dg1 has 5,323,186 bp with a GC content of 70% and 78% coding bases. It has one circular chromosome with 4,579 genes of which 4,202 encode proteins, 56 encode structural RNAs, 2 encode rRNA operons, and 325 represent pseudogenes. This is the smallest *Frankia* genome sequenced so far, slightly smaller than that of strain CcI3. It has been suggested that the genome of CcI3 underwent a process of reduction while the strain's capability of saprotrophic growth was reduced (8). The small genome size of Dg1 would also fit that hypothesis, considering that cluster II strains have not been cultured. Moreover, Dg1 has 325 pseudogenes, compared to 50 in CcI3, 128 in EAN1pec, and 12 in ACN14a, suggesting ongoing genome degradation. This degradation does not appear to involve genes encoding enzymes involved in glycolysis or amino acid, purine, or pyrimidine biosynthesis.

Nucleotide sequence accession number. The “*Candidatus Frankia datiscae*” Dg1 genome sequence and annotation data have been deposited in GenBank under accession number NC_015656.

The work conducted by the U.S. Department of Energy Joint Genome Institute was supported by the Office of Science of the U.S. Department of Energy under contract no. DE-AC02-05CH11231. The isolation of the genomic DNA was supported by a grant from the Swedish Research Council FORMAS to K.P.

* Corresponding author. Mailing address: Dept. of Botany, Stockholm University, 10691 Stockholm, Sweden. Phone: 46 8 16 3772. Fax: 46 8 16 5525. E-mail: pawlowski@botan.su.se.

REFERENCES

1. **Bennett, S.** 2004. Solexa Ltd. *Pharmacogenomics* **5**:433–438.
2. **Clawson, M. L., A. Bourett, and D. R. Benson.** 2004. Assessing the phylogeny of *Frankia*-actinorhizal plant nitrogen-fixing root nodule symbioses with *Frankia* 16S rRNA and glutamine synthetase gene sequences. *Mol. Phylog. Evol.* **31**:131–138.
3. **Lundquist, P.-O., and K. Huss-Danell.** 1991. Nitrogenase activity and amounts of nitrogenase proteins in a *Frankia-Alnus incana* symbiosis subjected to darkness. *Plant Physiol.* **95**:808–813.
4. **Margulies, M., et al.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**:376–380.
5. **Mirza, M. S., S. S. Hameed, and A. D. L. Akkermans.** 1994. Genetic diversity of *Datisca cannabina*-compatible *Frankia* strains as determined by sequence analysis of the PCR-amplified 16S ribosomal RNA gene. *Appl. Environ. Microbiol.* **60**:2371–2376.
6. **Murray, R. G. E., and E. Stackebrandt.** 1995. Taxonomic note: implementation of the provisional status *Candidatus* for incompletely described prokaryotes. *Int. J. Syst. Bacteriol.* **45**:186–187.
7. **Normand, P., et al.** 1996. Molecular phylogeny of the genus *Frankia* and related genera and emendation of the family Frankiaceae. *Int. J. Syst. Bacteriol.* **46**:1–9.
8. **Normand, P., et al.** 2007. Genome characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant biogeography. *Genome Res.* **17**:7–15.