Draft Genome Sequences of *Clostridium* Strains Native to Colombia with the Potential To Produce Solvents

Juan Pablo Rosas-Morales,ª Ximena Perez-Mancilla,ª Liliana López-Kleine,ᵇ Dolly Montoya Castaño,ª
Diego Mauricio Riaño-Pachónᶜ

Institute of Biotechnology (IBUN), Universidad Nacional de Colombia, Bogotá D.C., Colombiaª; Statistics Department, Sciences School, Universidad Nacional de Colombia, Bogotá D.C., Colombiaᵇ; Brazilian Bioethanol Science and Technology Laboratory (CTBE), Brazilian Center for Research in Energy and Materials, Campinas, Brazilᶜ

J.P.R.-M. and X.P.-M. contributed equally to this work.

Genomes from four *Clostridium* sp. strains considered to be mesophilic anaerobic bacteria, isolated from crop soil in Colombia, with a strong potential to produce alcohols like 1,3-propanediol, were analyzed. We present the draft genome of these strains, which will be useful for developing genetic engineering strategies.

Address correspondence to Dolly Montoya Castaño, dmontoyac@unal.edu.co, or Diego Mauricio Riaño-Pachón, diego.riano@bioetanol.org.br.

The need for obtaining alternative energy sources has renewed the interest in the production of different biological chemicals and biofuels using inexpensive substrates and microorganisms like *Clostridium* (1). Researchers have thus joined forces for evaluating the use of several by-products in alternative technologies, like the use of lignocellulosic material to produce acids and alcohols, or microbial fermentation of glycerol to produce 1,3-propanediol (1,3-PD) (2, 3). The latter is extremely relevant as it is the monomer from a polymer of commercial interest known as polytrimethylene terephthalate (PTT) (4).

We have previously shown that several strains from our collection have promising total solvent yield from glucose and the ability to hydrolyze different polysaccharides (5). Some of them have also shown great potential to produce higher yields of 1,3-PD from glycerol than reference strains (6). The genomes of four of these strains isolated from crop soils (7), named *Clostridium* sp. IBUN 13A (potatoes), 62F (grass), 125C (soya), and 22A (potatoes), were sequenced in an Illumina HiSeq2000 instrument using the NExtera library preparation kit producing 3,211,515 paired-end sequences (2 × 101 bp) per strain on average. Preprocessing of reads was done with Clean_reads and Fastx-toolkit v0.0.14. Genome assembly was performed with Velvet v1.2.10 (8), and postassembly improvement was achieved using the tools IMAGE (9) and iCORN (10) from PAGIT toolkit (11).

Prediction of RNA genes was carried out with tRNAscan v1.21 for tRNA (12) and RNAmmer v1.2 for rRNA (13). Protein-coding genes were predicted using the MAKER2 pipeline (14), employing Glimmer v3.02 (15), GenemarkS v4.17, and Prodigal v2.6 (16) as *ab initio* gene predictors. The function assignation for the predicted genes was done using the RAST annotation server. Information about the genome of these four strains is summarized in Table 1.

All genes implicated in the reductive pathway for 1,3-PD production, including the two-component regulatory system, were found grouped in the same contig for three strains. In strain 62F, those protein-coding sequences were not detected in agreement with phenotypic observations shown in other studies (6). Hydrolases implied in polysaccharide degradation were identified in all strains, highlighting their biotechnological potential.

Metabolic reconstruction and gene finding from these genomes will lead to the development of a genetic engineering strategy for the increase of 1,3-PD production.

Nucleotide sequence accession numbers. The draft genome sequences of the four strains have been deposited at DDBJ/EMBL/GenBank under the accession numbers given in Table 1. In this paper, the first versions of the genomes are described.

TABLE 1 Summary of the *Clostridium* sp. draft genomes

Strain	Assembly length (bp)	% G+C content	No. of contigs	No. of genes	No. of proteins	Accession no.
IBUN13A	4,643,590	28.65	261	4,086	4,022	JZWG00000000
IBUN62F	3,836,807	28.77	85	3,356	3,288	JZWH00000000
IBUN125C	4,596,888	28.67	71	3,959	3,888	JZWF00000000
IBUN22A	4,607,385	28.67	208	4,124	4,063	JZWE00000000

ACKNOWLEDGMENTS

REFERENCES

1. **Gu Y, Jiang Y, Yang S, Jiang W**. 2014. Utilization of economical substrate-derived carbohydrates by solventogenic clostridia: pathway dissection, regulation and engineering. Curr Opin Biotechnol **29**:124–131. http://dx.doi.org/10.1016/j.copbio.2014.04.004.

2. **Koutinas AA, Vlysidis A, Pleissner D, Kopsahelis N, Lopez Garcia I, Kookos IK, Papanikolaou S, Kwan TH, Lin CSK**. 2014. Valorization of industrial waste and by-product streams via fermentation for the production of chemicals and biopolymers. Chem Soc Rev **43**:2587–2627. http://dx.doi.org/10.1039/c3cs60293a.

3. **Kaur G, Srivastava AK, Chand S**. 2012. Advances in biotechnological production of 1,3-propanediol. Biochem Eng J **64**:106–118. http://dx.doi.org/10.1016/j.bej.2012.03.002.

4. **Kurian JV**. 2005. A new polymer platform for the future—Sorona from corn derived 1,3-propanediol. J Polym Environ **13**:159–167. http://dx.doi.org/10.1007/s10924-005-2947-7.

5. **Montoya D, Arévalo C, Gonzales S, Aristizabal F, Schwarz WH**. 2001. New solvent-producing *Clostridium* sp. strains, hydrolyzing a wide range of polysaccharides, are closely related to *Clostridium butyricum*. J Ind Microbiol Biotechnol **27**:329–335. http://dx.doi.org/10.1038/sj/jim/7000193.

6. **Cárdenas DP, Pulido C, Aragón ÓL, Aristizabal FA, Suarez Z, Montoya Castaño D**. 2006. Evaluación de la producción de 1,3-propanodiol por cepas nativas de *Clostridium* sp. mediante fermentación a partir de glicerol USP y glicerol industrial subproducto de la producción de biodiésel. Rev Col Cienc Quím Farm **35**:120–137.

7. **Montoya D, Spitia S, Silva E, Schwarz WH**. 2000. Isolation of mesophilic solvent-producing clostridia from Colombian sources: physiological characterization, solvent production and polysaccharide hydrolysis. J Biotechnol **79**:117–126. http://dx.doi.org/10.1016/S0168-1656(00)00218-2.

8. **Zerbino DR, Birney E**. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res **18**:821–829. http://dx.doi.org/10.1101/gr.074492.107.

9. **Tsai IJ, Otto TD, Berriman M**. 2010. Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. Genome Biol **11**:R41. http://dx.doi.org/10.1186/gb-2010-11-4-r41.

10. **Otto TD, Sanders M, Berriman M, Newbold C**. 2010. Iterative correction of reference nucleotides (iCORN) using second generation sequencing technology. Bioinformatics **26**:1704–1707. http://dx.doi.org/10.1093/bioinformatics/btq269.

11. **Swain MT, Tsai IJ, Assefa SA, Newbold C, Berriman M, Otto TD**. 2012. A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs. Nat Protoc **7**:1260–1284. http://dx.doi.org/10.1038/nprot.2012.068.

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

13. **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.

14. **Holt C, Yandell M**. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics **12**:491. http://dx.doi.org/10.1186/1471-2105-12-491.

15. **Delcher AL, Bratke KA, Powers EC, Salzberg SL**. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics **23**:673–679. http://dx.doi.org/10.1093/bioinformatics/btm009.

16. **Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ**. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11**:119. http://dx.doi.org/10.1186/1471-2105-11-119.