



# Sequencing and *De Novo* Assembly of the Complete Chloroplast Genome of the Peruvian Carrot (*Arracacia xanthorrhiza* Bancroft)

Javier Santiago Alvarado,<sup>a</sup> Diane Hinojosa López,<sup>a</sup> Isaury Maldonado Torres,<sup>a</sup> María Margarita Meléndez,<sup>a</sup> Rosalinda Aybar Batista,<sup>a</sup> Vivek K. Raxwal,<sup>b</sup> Juan A. Negrón Berríos,<sup>a</sup>  Alok Arun<sup>a</sup>

Institute of Sustainable Biotechnology, Department of Science and Technology, Inter American University of Puerto Rico, Barranquitas, Puerto Rico<sup>a</sup>; Plant Molecular Biology Group, Central European Institute of Technology, Brno, Czech Republic<sup>b</sup>

**ABSTRACT** *Arracacia xanthorrhiza* is an important secondary food crop in South America and Puerto Rico. The lack of crop protection and improvement strategies leads to infections damaging the storage roots. Here, we report the annotated complete chloroplast genome sequence of *A. xanthorrhiza* as a step toward developing genomic resources for this crop.

The Peruvian carrot (*Arracacia xanthorrhiza* Bancroft), commonly known as apio, is a tuberous perennial crop in the family *Apiaceae* originating from the Andes (1–3). *A. xanthorrhiza* is a secondary food crop for over 100 million people, mainly in South America and Puerto Rico. Starch-rich storage roots are the main food product, the rootstock and leaves are used as animal feed, and the aerial stems are used as propagules (1, 4). The crop has low-input requirements and can be grown in a variety of frost-free tropical highland environments (2). However, in the absence of preventive management practices and proper handling, infections caused by pathogens severely damage the storage roots (5). The lack of certainty about the agents and the process of infection pose a major obstacle to understanding and preventing infection at its very onset. Despite the economic and agricultural importance of *A. xanthorrhiza*, any biotechnological strategy aimed at improving the yield or protecting the crop from pest damage is nonexistent, partly due to the lack of genomic resources. Here, we present the sequence and annotated chloroplast genomic DNA (cpDNA) of *A. xanthorrhiza*. Chloroplast genomes have been used extensively for DNA barcoding (6, 7), phylogenetic analyses (8), plastid transformation (9), and gene silencing studies (10). The availability of the chloroplast genome sequence of *A. xanthorrhiza* would be the first major step toward its bioengineering.

In the present study, cpDNA was extracted from leaf tissues of *A. xanthorrhiza* according to the protocol described in Shi et al. (11) with minor modifications. Approximately, 1 ng of cpDNA was used to prepare sequencing libraries using Illumina's NexteraXT DNA sample prep kit. Paired-end (2 × 150 bp) sequencing reactions were performed on Illumina's MiSeq platform at the Institute of Sustainable Biotechnology, IAUPR, Puerto Rico. The pair-end reads were trimmed for adapter and low-quality reads (Phred score <30) using Trim Galore ([http://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)). The resulting 7.9 million sequencing reads generated an estimated 1,101-fold chloroplast genome coverage. The *de novo* assembly of the filtered pair-end reads was performed using NOVOPlasty (12). Two annotation tools, CPGAVAS (13) and DOGMA (14), were used to annotate the cp genome. Finally, protein-coding genes were manually annotated.

**Received** 12 November 2016 **Accepted** 16 December 2016 **Published** 16 February 2017

**Citation** Alvarado JS, López DH, Torres IM, Meléndez MM, Batista RA, Raxwal VK, Berríos JAN, Arun A. 2017. Sequencing and *de novo* assembly of the complete chloroplast genome of the Peruvian carrot (*Arracacia xanthorrhiza* Bancroft). *Genome Announc* 5:e01519-16. <https://doi.org/10.1128/genomeA.01519-16>.

**Copyright** © 2017 Alvarado et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Alok Arun, [aloktransgenic@gmail.com](mailto:aloktransgenic@gmail.com).

J.S.A., D.H.L., and I.M.T. contributed equally to this work.

The current investigation suggests that the *A. xanthorrhiza* circular chloroplast genome is 143,989 base pairs (bp) in length with 37.48% G+C content. The cp genome is divided into (a) large single copy (LSC) (49,169 bp), (b) small single copy (SSC) (17,439 bp), and (c) two inverted regions (31,370 bp). The cp genome encoded 106 manually curated unique genes consisting of 71 protein-coding genes, seven rRNA genes, and 28 tRNA genes. Exons of five genes, namely, *ycf3*, *clpP*, *rps12*, *ndhA*, and *ndhB*, were interrupted with intron sequences. The genome architecture of the *A. xanthorrhiza* cp genome shows strong similarity to cp genome of closely related *Apiaceae* crop carrot, *Daucus carota* (15). The availability of the cp genome sequence of *A. xanthorrhiza* will be an asset for studying the genetic diversity of various apio landraces and their closely related wild relatives, and will contribute to a better understanding of genomics of this economically important crop.

**Accession number(s).** The chloroplast DNA genome sequence has been deposited in GenBank under accession number [KY117235](https://www.ncbi.nlm.nih.gov/nuccore/KY117235).

## ACKNOWLEDGMENTS

We thank Aramis S. Amaral for raising the plants and Jose A. Perez Melendez, IAUPR, PR and Nirmal Joshee, FVSU, GA for their suggestions on the manuscript. This study was supported by HSI-STEM grant (PO31C110181) to Juan A. N. Berrios and IAUPR Seed grant (06/242016) to Alok Arun.

## REFERENCES

- Hermann M, Heller J. 1997. Andean roots and tubers: ahipa, arracacha, maca and yacon. *In* Promoting the conservation and use of underutilized and neglected crops (IPGRI), Rome, Italy.
- Hodge WH. 1954. The edible arracacha. A little known root crop of the Andes. *Econ Bot* 8:195–221. <https://doi.org/10.1007/BF02984890>.
- Vavilov N. 1992. The important agricultural crops of pre-Colombian America and their mutual relationship. *In* Origin and geography of cultivated plants. Cambridge University Press, Cambridge.
- Dufour D, Hurtado J, Wheatley C. 1996. Characterization of starches from noncereal crops cultivated in tropical America: comparative analyses of starch behavior under different stress conditions, p. 11–15. *In* International Symposium on Cassava Starch and Starch Derivatives. Nanning, China.
- Thompson A. 1980. Reduction of losses during the marketing of arracacha (*Arracacia xanthorrhiza*). *Acta Hort* 116:55–60.
- Dong W, Liu H, Xu C, Zuo Y, Chen Z, Zhou S. 2014. A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: a case study on ginsengs. *BMC Genet* 15:138. <https://doi.org/10.1186/s12863-014-0138-z>.
- Larranaga N, Hormaza JI. 2015. DNA barcoding of perennial fruit tree species of agronomic interest in the genus *Annona* (Annonaceae). *Front Plant Sci* 6:589. <https://doi.org/10.3389/fpls.2015.00589>.
- Leliaert F, Tronholm A, Lemieux C, Turmel M, Depriest MS, Bhattacharya D, Karol KG, Fredericq S, Zechman FW, Lopez-Bautista JM. 2016. Chloroplast phylogenomic analyses reveal the deepest-branching lineage of the Chlorophyta, Palmophyllophyceae class. nov. *Sci Rep* 6:25367. <https://doi.org/10.1038/srep25367>.
- Hanson MR, Gray BN, Ahner BA. 2013. Chloroplast transformation for engineering of photosynthesis. *J Exp Bot* 64:731–742. <https://doi.org/10.1093/jxb/ers325>.
- Daniell H, Kumar S, Dufourmantel N. 2005. Breakthrough in chloroplast genetic engineering of agronomically important crops. *Trends Biotechnol* 23:238–245. <https://doi.org/10.1016/j.tibtech.2005.03.008>.
- Shi C, Hu N, Huang H, Gao J, Zhao YJ, Gao LZ. 2012. An improved chloroplast DNA extraction procedure for whole plastid genome sequencing. *PLoS One* 7:e31468. <https://doi.org/10.1371/journal.pone.0031468>.
- Dierckxsens N, Mardulyn P, Smits G. 24 October 2016. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res* <https://doi.org/10.1093/nar/gkw955>.
- Liu C, Shi L, Zhu Y, Chen H, Zhang J, Lin X, Guan X. 2012. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genomics* 13:715. <https://doi.org/10.1186/1471-2164-13-715>.
- Wyman SK, Jansen R, Boore J. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20:3252–3255.
- Ruhlman T, Lee SB, Jansen RK, Hostetler JB, Tallon LJ, Town CD, Daniell H. 2006. Complete plastid genome sequence of *Daucus carota*: implications for biotechnology and phylogeny of angiosperms. *BMC Genomics* 7:222. <https://doi.org/10.1186/1471-2164-7-222>.