Complete sequence of wild Physalis philadelphica chloroplast genome

Isaac Sandoval-Padilla, Jessica Pérez-Alquicira, María del Pilar Zamora-Tavares, Aarón Rodríguez, Moisés Cortés-Cruz, Gabriela Alcalá-Gómez and Ofelia Vargas-Ponce

*Maestría en Biotecnología y Manejo de Recursos Naturales y Agrícolas, Centro Universitario de Ciencias Biológicas y Agropecuarias, Universidad de Guadalajara, Zapopan, México; Cátedras CONACYT–Universidad de Guadalajara, Laboratorio Nacional de Identificación y Caracterización Vegetal (LanVeg), Departamento de Botánica y Zoología, Centro Universitario de Ciencias Biológicas y Agropecuarias, Universidad de Guadalajara, Zapopan, México; Laboratorio Nacional de Identificación y Caracterización Vegetal (LanVeg), Consejo Nacional de Ciencia y Tecnología (CONACYT), Universidad de Guadalajara, Zapopan, México; Instituto de Botánica, Depto. Botánica y Zoología, Centro Universitario de Ciencias Biológicas y Agropecuarias, Universidad de Guadalajara, Zapopan, México; Centro Nacional de Recursos Genéticos, Laboratorio de ADN y Genómicas, Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, Tepatitlán de Morelos, México; Doctorado en Biotecnología, Ecología y Manejo de Recursos Naturales y Agrícolas, Centro Universitario de Ciencias Biológicas y Agropecuarias, Universidad de Guadalajara, Zapopan, México

**ABSTRACT**

*Physalis philadelphica* Lam. has horticultural importance because of its edible fruits. Cultivated and wild populations grow in Mexico. In this study, the complete plastome nucleotide sequence of wild plants was generated using the IonTorrent PGM sequencing technology. The plastome size was 156,804 bp and displayed the typical circular quadruplicate structure, consisting of a pair of inverted repeat regions (25,595 bp) separated by a large single copy region (87,131 bp) and a small single copy region (18,483 bp). The chloroplast genome included 80 protein–coding genes, four rRNAs, and 31 tRNAs. The phylogenetic analysis based on 19 Solanaceae chloroplast genomes recovered a clade with all *Physalis* species. This work revealed the importance of the plastome sequence to solve infrageneric phylogenetic relationships.

The genus *Physalis* L. (Solanaceae) groups 90 species and Mexico is the centre of diversification (Zamora-Tavares et al. 2015). Sixty-five species grow in Mexico, but *Physalis philadelphica* Lam. has commercial value. Its fruits are part of the staple diet of the Mexicans (Zamora-Tavares et al. 2015). Wild populations of *Physalis philadelphica* grow nationwide, across different environments (Martínez et al. 2017). It is cultivated in different regions of the country throughout the year, and in countries of the Americas, Europe, and Asia is present as an experimental crop (Zamora-Tavares et al. 2015).

Domestication reduces genetic variation but wild populations restore it (Casas et al. 2007). Across Angiosperms, chloroplast DNA has a conserved genomic structure with a high mutation rate in some regions (Ravi et al. 2008). As a result, plastome sequences are useful to understand relationships between domesticates and wild (Daniell et al. 2016). Plastome variation between cultivated and wild ancestors, as well as their comparison with other wild relatives, have been documented in plants of economic importance such as *Onyza* spp. (O. barthii, A. Chev, O. glaberimia, Steud., O. nivara, S.D. Sharma & Shastr, O. rufigon, Griff., O. sativa, L., Tong et al. 2016) and *Helianthus annus* L. (Makarenko et al. 2016). It is our interest to make these comparisons in *P. philadelphica* to elucidate if the species has undergone different evolutionary patterns. In the present study, we sequenced and assembled the complete plastome of wild plants of *P. philadelphica*. Until now, GenBank has three available plastomes from *Physalis* (gene bank accession numbers *P. angulata* L. NC039457, *P. peruviana* L. MH019242, and *P. pruinosa* L. NC039458).

Fresh leaves from wild plants of *Physalis philadelphica* were harvested in San Antonio de los Vázquez, Ixtlahuacán del Río, Jalisco, México (20°49’29.14”N, 103°39’55.5”O). cpDNA was isolated based on Shi et al. (2012) and was stored at the Laboratorio Nacional de Identificación y Caracterización Vegetal (LanVeg) at the Universidad de Guadalajara (Voucher: 021118ISP). A total of 1,685,803 single-end reads were generated using the IonTorrent PGM technology (Thermo Fisher Scientific, Carlsbad, CA) and these reads were de novo assembled into contigs using SPAdes 3.12.0 (Bankevich et al. 2012). The assembled genomic sequences were filtered with Bowtie2 2.3.5 (Langmead and Salzberg 2012) followed by the annotation using DOGMA (Wyman et al. 2004). tRNAs were predicted with tRNAscan SE 2.0.3.

**CONTACT** Ofelia Vargas-Ponce, vargasofelia@gmail.com, Instituto de Botánica, Depto. Botánica y Zoología, Centro Universitario de Ciencias Biológicas y Agropecuarias, Universidad de Guadalajara, Ap. postal 139, Zapopan, México; Laboratorio Nacional de Identificación y Caracterización Vegetal (LanVeg), Consejo Nacional de Ciencia y Tecnología (CONACYT), Universidad de Guadalajara, Ap. postal 139, Zapopan, México

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
The circular representation of the *P. philadelphica* plastome (GenBank MN192191) was obtained with OGDraw 1.3.1 (Greiner et al. 2019).

The chloroplast genome size of *Physalis philadelphica* was 156,804 bp. The genome displayed the typical circular quadripartite structure. It consisted of a pair of inverted repeated (IR) regions (25,595 bp) separated by two single copy regions, one large (LSC, 87,131 bp) and one small (SSC, 18,483 bp). Plastome nucleotide composition was 30.87% adenine, 19.06% cytosine, 18.45% guanine, and 31.66% thymine. General guanine-cytosine content was 37.5%, but the IR had 43.06% each, and the LSC and SSC regions showed 35.56 and 31.35%, respectively. The plastome included 115 genes, 80 protein-coding, four rRNAs, and 31 tRNAs. Twenty-two duplicate genes were present in the IR regions: four rRNAs, seven tRNAs, nine protein-coding, and two pseudogenes. Additionally, there were 15 genes harbouring introns, 13 with two and two with one.

Within Solanaceae, the phylogenetic position of *Physalis philadelphica* was estimated. The phylogenetic tree was constructed based on 17 Solanaceae chloroplast genomes available on GenBank, using *Ipomoea batatas* (L.) Lam. (Convolvulaceae) as outgroup. Maximum Likelihood (ML) analysis was performed using RA × ML (Stamatakis, 2014) and 1,000 bootstraps iterations. The phylogenetic estimation was performed with the evolutionary model GTR + I + G, according to the estimate of jModelTest 2.1.10 (Darriba et al. 2012). At the tribe level, the cladogram recovered the same relationships of Solanaceae as did Särkinen et al. (2013). Subtribes lochrominae and Physalidinae had a sister group relationship (Deanna et al. 2019). The phylogenetic hypothesis supported *Physalis* as a monophyletic group (Figure 1). Even though, the statistical support was moderated, this work revealed the usefulness of the plastome sequence to solve infrageneric phylogenetic relationships.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Funding**

This work was supported by UDG and CONACyT-Laboratorio Nacional de Identificación y Caracterización Vegetal (LaniVeg) [Grant no. 293833], Universidad de Guadalajara [Grant Prosi-2018 to PZT and OVP], and CONACyT-México for master scholarship for graduate studies at Maestría en Biotecnología y Manejo de Recursos Naturales y Agrícolas (BIMARENA) [Grant no. 928518 to ISP].

**ORCID**

Ofelia Vargas-Ponce http://orcid.org/0000-0003-4139-503X

**References**

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prijibelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Compl Biol. 19:455–477.

Casas A, Otero-Arnaiz A, Pérez-Negrón E, Valiente-Banuet A. 2007. In situ management and domestication of plants in Mesoamerica. Ann Bot. 100:1101–1115.
Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences. Methods Mol Biol. 1962:1–14.
Daniell H, Lin CS, Yu M, Chang WJ. 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 17:134.
Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods. 9:772.
Deanna R, Larter MD, Barboza GE, Smith SD. 2019. Repeated evolution of a morphological novelty: a phylogenetic analysis of the inflated fruiting calyx in the Physalideae tribe (Solanaceae). Am J Bot. 106:270–210.
Greiner S, Lehwark P, Bock R. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 47:W59–W64.
Langmead B, Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods. 9:357–359.
Makarenko MS, Usatov AV, Markin NV, Azarin KV, Gorbachenko OF, Usatov NA. 2016. Comparative genomics of domesticated and wild sunflower: complete chloroplast and mitochondrial genomes. J Biol Sci. 16:71–75.
Martinez M, Vargas-Ponce O, Rodriguez A, Chiang F, Ocegueda S. 2017. The Solanaceae family in Mexico. Bot Sci. 95:1–15.
Ravi V, Khurana JP, Tyagi AK, Khurana P. 2008. An update on chloroplast genomes. Plant Syst Evol. 271:101–122.
Särkinen T, Bohs L, Olmstead RG, Knapp S. 2013. A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. BMC Evol Biol. 13:214.
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.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30:1312–1313.
Tong W, Kim TS, Park YJ. 2016. Rice chloroplast genome variation architecture and phylogenetic dissection in diverse Oryza species assessed by whole-genome resequencing. Rice (N Y). 9:57.
Wyman SK, Jansen RK, Boone JL. 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics. 20:3252–3255.
Zamora-Tavares P, Vargas-Ponce O, Sánchez-Martínez J, Cabrera-Toledo D. 2015. Diversity and genetic structure of the husk tomato (Physalis philadelphica Lam.) in Western Mexico. Genet Resour Crop Evol. 62:141–153.