The complete chloroplast genomes of two Mexican plants of the medicinal and toxic herb

Datura stramonium (Solanaceae)

I. M. De-la-Cruz¹ and J. Núñez-Farfán^{1*}

¹Department of Evolutionary Ecology, Institute of Ecology, National Autonomous University of Mexico (UNAM), Mexico City, Mexico, imda@ecologia.unam.mx (IMDA), farfan@unam.mx (JNF).

*E-mail of the corresponding author, J. Núñez-Farfán, farfan@unam.mx

Abstract

The annual herb, *Datura stramonium*, is a member of the Solanaceae family. In this study, we reported two chloroplast and mitochondrial genomes of two Mexican plants of *Datura stramonium*. Both chloroplast genomes contained 87 protein-coding genes, 157 genes, 7 rRNA genes and 62 tRNA genes. Both chloroplast genomes of *D. stramonium* had the same genome size (155,884 bp). Likewise, 196 genes, 162 protein coding-sequences and 3 rRNA were annotated for both mitochondrial genomes. The mitochondrion genome size for Ticumán was 330,791 bp and for Teotihuacan was 335,447 bp.

Keywords: Chloroplast, *Datura stramonium*, Mitogenome, Solanaceae, Organelle.

CHLOROPLAST GENOME ANNOUNCEMENT

The annual herb, *Datura stramonium*, is a member of the Solanaceae family (Castillo *et al.* 2020). This herb is well characterized because produces the highest concentration of tropane alkaloids within this family (Kohnen-Johannsen and Kayser 2019). In particular, scopolamine, atropine and anisodamine are the mainly secondary compounds of this species (Castillo *et al.* 2013; De-la-Cruz *et al.* 2019). These compounds have been used in medicine as an antiasthmatic, an anesthetic and for rheumatism (WHO 2015). Recently, the first two draft nuclear genomes of two Mexican plants of *D. stramonium* have been released (DDBJ/ENA/GenBank BioProject: PRJNA622882; De-la-Cruz *et al. in prep*). One nuclear genome corresponds to a plant collected in Ticumán, State of Morelos, Mexico (GenBank acc. JAAWWX000000000), and the other nuclear genome corresponds to a plant collected in Teotihuacán, State of Mexico, Mexico (GenBank acc. JAAWWY000000000). Here, we reported the chloroplast and the mitochondrial genomes of these two plants of *D. stramonium*. To this end, gDNA was extracted from fresh leaves with a modified CTAB mini-prep protocol (Doyle and Doyle 1987). A total of 200 ng of gDNA were used to construct paired-end libraries (PE) for Illumina HiSeq 4000 sequencing. The sequencing and libraries preparation were carried out in the QB3 Functional Genomics and Vincent J. Coates Sequencing Laboratories at the University of California, Berkeley. 323M PE raw sequences (2 X 150b; ~300 bp insert size) were obtained from Illumina sequencing; corresponding to 112 Gb and an average 30.85-fold genome coverage for the Ticumán individual, while 318M PE sequences corresponding to 110 Gb and 30.29-fold genome coverage were generated for the Teotihuacán individual. Trimming of Illumina sequences was carried out using a phred quality score > 30 in TRIMMOMATIC v0.32 (Bolger *et al.* 2014).

To assembly the chloroplast and mitochondrial genomes we used NOVOPlasty v3.8.2 (Dierckxsens *et al.* 2016). However, we did not used all the sequences mentioned above. Instead of that, we used the subsample option in NOVOPlasty. Using this option, the user limits the RAM memory that the program can use, and the program takes a fraction of the millions of PE reads based on the RAM memory allowed. In this way, using a max memory of 40 Gb, the program subsampled 15.10 % of the PE sequences in both datasets (2,168,304 sequences for the Teotihuacán plant and 1,545,934 for the Ticumán plant). Then, NOVOPlasty has to be initiated by a seed (Dierckxsens *et al.* 2016). Because we already had assembled the nuclear genomes of both plants, we used DIAMOND blast (E-value < 1e-5) (Buchfink and Huson 2015) to align the reference chloroplast genome of *D. stramonium* (Yang *et al.* 2014; GenBank acc. NC_018117) to the two *Datura* nuclear genome assemblies. Then, we retrieved the contigs from the *Datura* assemblies that had a match value of 100 % with the reference chloroplast. These contigs were used as a seed in NOVOPlasty to reconstruct completely the chloroplasts of our *D. stramonium* plants. The mitochondrial genomes were also assembled using the “mito_plant” option in NOVOPlasty (Dierckxsens *et al.* 2016). In this mode, each of the two chloroplast genomes were used as reference genomes and as seed we used the mitochondrial genome of *Capsicum annum* var. *glabriusculum*; NCBI, acc: MN196478). GeSeq was also used to annotate the mitochondrial genomes but in this case, we used as references, the annotations of the mitochondrial genomes of *C. annum* (acc. NC_024624), *Solanum lycopersicum* (acc. NC_035963) and *S. pennellii* (acc. NC_035964). tRNAscan-SE v2.0.5 was also used to find and annotate tRNA genes.

The *D. stramonium* chloroplasts were annotated as a circular molecule using GeSeq (Tillich *et al.* 2017). We used as references, the annotations of the *D. stramonium* and *Capsicum annum* chloroplast genomes (NC_018117 and NC_018552, respectively). tRNAscan-SE v2.0.5

was used to find and annotate tRNA genes. Both chloroplast genomes of *D. stramonium* (NCBI accessions; SAMN14687971 and SAMN14687972, Ticumán and Teotihuacán, respectively) had the same genome size (155,884 bp). The overall GC content was 38.59% for both genomes. Both chloroplast genomes contained 87 protein-coding genes, 157 genes, 7 rRNA genes and 62 tRNA genes. Likewise, the mitochondrial genomes were annotated as a circular molecule; 196 genes, 162 protein coding-sequences and 3 rRNA were annotated for both mitochondrial genomes. The mitochondrion genome size for Ticumán was 330,791 bp and for Teotihuacan was 335,447 bp (NCBI accessions; SAMN14691696, SAMN14691697, respectively).

To understand the phylogenetic relationship between *D. stramonium* and related species, the complete chloroplast of 9 species were aligned by MAFFT v7 (Kato et al. 2002). The evolutionary history was inferred by using the Maximum Likelihood method based on Tamura-Nei model in MEGA7.0 (Kumar et al. 2016). Bootstrap (BS) values were calculated from 1000 replicate analysis. The phylogenetic tree was divided into four groups, which were the *Solanum*, *Capsicum*, *Datura* and *Nicotiana* clades (Figure 1). As was expected, both genomes of *D. stramonium* were grouped with the chloroplast already published for this species and, occurred as a sister taxon with *Nicotiana*.

Figure 1. ML phylogenetic tree of *D. stramonium* with 9 species was constructed by chloroplast sequence. Numbers in the nodes are bootstrap values from 1000 replicates. *Solanum pimpinellifolium* was set as outgroup. NCBI accessions for chloroplast genomes = acc.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by CONACyT “Fronteras de la Ciencia” (#1527) and for PAPIIT UNAM grant #IG200717 to JNF. Ivan Mijail De-la-Cruz acknowledges the scholarship (# 283799) by CONACyT.

References

- Bolger, A. M., M. Lohse, and B. Usadel, 2014a Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*.30:2114-20.
- Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12(1), 59–60.
- Castillo, G., Calahorra-Oliart, A., Núñez-Farfán, J., Valverde, P. L., Arroyo, J., Cruz, L. L., and Tapia-López, R. (2019). Selection on tropane alkaloids in native and non-native populations of *Datura stramonium*. *Ecology and Evolution*, 9(18), 10176–10184.
- Castillo, G., Cruz, L. L., Hernández-Cumplido, J., Oyama, K., Flores-Ortiz, C. M., Fornoni, J., Valverde, P. L., and Núñez-Farfán, J. (2013). Geographic association and temporal variation of chemical and physical defense and leaf damage in *Datura stramonium*. *Ecological Research*, 28(4), 663–672.
- De-la-Cruz, I. M., Cruz, L. L., Martínez-García, L., Valverde, P. L., Flores-Ortiz, C. M., Hernández-Portilla, L. B., and Núñez-Farfán, J. (2020). Evolutionary response to herbivory: Population differentiation in microsatellite loci, tropane alkaloids and leaf trichome density in *Datura stramonium*. *Arthropod-Plant Interactions*, 14(1), 21–30.
- Dierckxsens, N., Mardulyn, P., and Smits, G. (2017). NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research*, 45(4), e18–e18.
- Doyle, J. J., and J. L. Doyle, 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*.19:11-15.
- Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14), 3059–3066.

- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, 33(7), 1870–1874.
- Kohnen-Johannsen, K. L., and O. Kayser, 2019 Tropane Alkaloids: chemistry, pharmacology, biosynthesis and Production. *Molecules*. 24:796.
- Yang, Y., Yuanye, D., Qing, L., Jinjian, L., Xiwen, L., and Yitao, W. (2014). Complete Chloroplast Genome Sequence of Poisonous and Medicinal Plant *Datura stramonium*: Organizations and Implications for Genetic Engineering. *PLoS ONE*, 9(11), e110656.
- WHO, 2015 Annex 1 19th WHO Model List of Essential Medicines; WHO: Geneva, Switzerland.

Supplemental material 1. Examples of: (a) Annotation of circularized molecule of the chloroplast genome of *Datura stramonium*

Ticumán. (b) Annotation of circularized molecule of the mitochondrial genome of *Datura stramonium* Teotihuacán.