



ELSEVIER

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

CrossMark

Data Article

Illumina sequencing of the chloroplast genome of common ragweed (*Ambrosia artemisiifolia* L.)Erzsébet Nagy^a, Géza Hegedűs^b, János Taller^a, Barbara Kutasy^a, Eszter Virág^{a,*}^a University of Pannonia, Georgikon Faculty, Department of Plant Science and Biotechnology, Keszthely, Hungary^b University of Pannonia, Georgikon Faculty, Department of Economic Methodology, Keszthely, Hungary

ARTICLE INFO

Article history:

Received 7 July 2017

Received in revised form

21 September 2017

Accepted 4 October 2017

Available online 7 October 2017

Keywords:

Illumina sequencing

Chloroplast genome

cpDNA

Common ragweed

Ambrosia artemisiifolia

ABSTRACT

Common ragweed (*Ambrosia artemisiifolia* L.) is the most widespread weed and the most dangerous pollen allergenic plant in large areas of the temperate zone. Since herbicides like PSI and PSII inhibitors have their target genes in the chloroplast genome, understanding the chloroplast genome may indirectly support the exploration of herbicide resistance and development of novel control methods. The aim of the present study was to sequence and reconstruct for the chloroplast genome of *A. artemisiifolia* and establish a molecular dataset. We used an Illumina MiSeq protocol to sequence the chloroplast genome of isolated intact organelles of ragweed plants grown in our experimental garden. The assembled chloroplast genome was found to be 152,215 bp (GC: 37.6%) in a quadripartite structure, where 80 protein coding genes, 30 tRNA and 4 rRNA genes were annotated in total. We also report the complete sequence of 114 genes encoded in *A. artemisiifolia* chloroplast genome supported by both MIRA and Velvet *de novo* assemblers and ordered to *Helianthus annuus* L. using the Geneious software.

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

* Corresponding author.

E-mail address: virag@georgikon.hu (E. Virág).

Specifications Table

Subject area	<i>Biology</i>
More specific subject area	<i>Chloroplast genome of common ragweed</i>
Type of data	<i>Table, figure</i>
How data was acquired	<i>2 × 300 Illumina MiSeq sequencing</i>
Data format	<i>Raw reads in FASTAQ, complete cp genome in FASTA</i>
Experimental factors	<i>5 g young leaves were collected from young about 20 cm tall plants, and incubated for 48 h at 4 °C in dark</i>
Experimental features	<i>Complete chloroplast genome of <i>Ambrosia artemisiifolia</i></i>
Data source location	<i>Keszthely-city, Hungary</i>
Data accessibility	<i>Information and complete data are accessible in the NCBI under BioProject and BioSample ID: PRJNA383307, SAMN06761249. The raw reads are available in Fastq format in the NCBI SRA database at the following link https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?run=SRR6050242. Complete chloroplast genome is available in GenBank under accession number: MF362689; https://www.ncbi.nlm.nih.gov/nucleotide/MF362689</i>

Value of the data

- Common ragweed is one of the most aggressive invasive weed species and the most dangerous pollen allergenic plant in large areas of the temperate zone.
- Understanding the chloroplast genome of this species may indirectly support chemical control of it, since a large part of herbicides have their target genes in the chloroplast genome e.g. triazine-derivatives [1], diphenylethers [2] or the redox active Paraquat [3].
- The reported data mean an important source for further chloroplast derived investigations like phylogenetic, photosynthetic or oxidative metabolism studies of the species.

1. Data

Intact chloroplasts were isolated from young leaves of *Ambrosia artemisiifolia*. Followed by cpDNA isolation and sequencing. The raw reads are available in Fastq format in the SRA database under the accession SRR6050242. The assembled chloroplast genome and annotated genes are available through NCBI nucleotide (MF362689).

2. Experimental design, materials and methods

2.1. Plant material and isolation of cpDNA

Seeds of an *A. artemisiifolia* plant grown in our experimental garden were sown on peat, and plants were grown in pots under greenhouse conditions.

In total, 5 g leaf tissue was collected from young, about 20 cm tall plants. To avoid high level starch accumulation the harvested leaves were incubated in Parafilm-sealed Petri dishes for 48 h at 4 °C in dark before chloroplast preparation. Chloroplast was isolated using the Chloroplast Isolation kit (Sigma-Aldrich, USA) according to the instructions of the manufacturer. The intact chloroplasts were

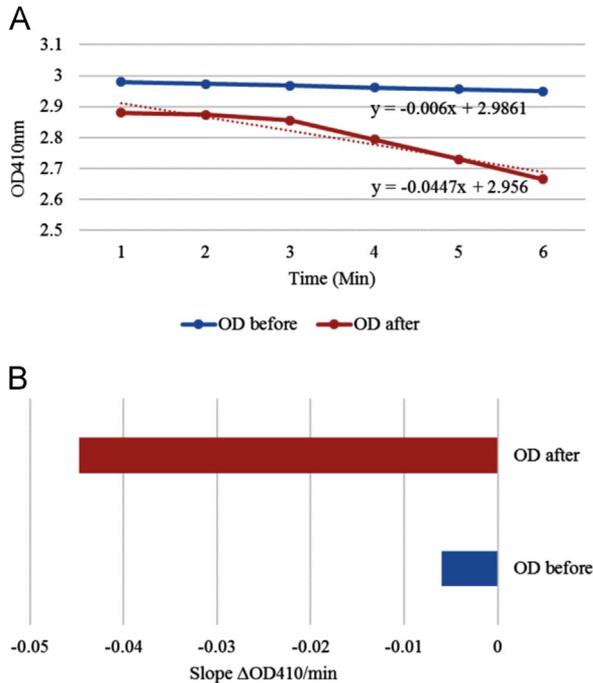


Fig. 1. Analysis of the *Ambrosia* chloroplast preparation. Graph A: The degree of integrity of prepared chloroplast. It is assessed by comparing the rate of ferricyanid reduction upon illumination (at 410 nm) before (blue) and after (orange) osmotic shock. Graph B: Bars representing the slopes of the lines in graph A. The differences of slope values indicated that 81% of isolated chloroplast was intact and suitable for cpDNA extraction.

separated from the broken ones by centrifugation on top of 40/80% Percoll® gradient. To calculate the percentage of intact chloroplasts the ferricyanide photoreduction procedure was used [4]. The reduction of ferricyanid was measured spectrophotometrically at 410 nm. The percentage of intact chloroplasts of the preparation was assessed by comparing the rates of ferricyanide photoreduction with and without osmotic shock of the chloroplasts using the following formula:

$$\frac{B-A}{B} \cdot 100\% = \% \text{ intact chloroplasts}$$

where A and B are the change in absorbance at 410 nm as a function of time (min) without and with osmotic shock measured by spectrophotometer. Analysis indicated that the 81% of the *Ambrosia* chloroplast preparation was intact and suitable for cpDNA extraction (Fig. 1).

Isolation of cpDNA was performed as described by Nascimento Vieira et al. [5] with the following modification: after the addition of potassium acetate the sample was kept on ice for 2 hours. Then the procedure was continued according to reference till DNA dissolution in nuclease-free water.

2.2. Library preparation and sequencing

An Illumina paired-end cpDNA library (average insert size of 500 bp) was constructed using the Illumina TruSeq library preparation kit according to manufacturer's protocol. The cpDNA library was sequenced with 2×300 bp on MiSeq platform (Illumina, USA).

Table 1

Classification of genes after chloroplast genome reconstruction. The annotated genes were categorized according to their function. Nominations: underlined: contains one intron, underlined bold: **contains more than one intron**.

Group of genes	Genes
Protein genes	
ATP synthase	<i>atpA atpB atpE <u>atpF</u> atpH atpI</i>
Cytochrome b/f complex	<i>petA <u>petB</u> <u>petD</u> petG petL petN</i>
Large subunit of RuBisCO	<i>rbcl</i>
NADH dehydrogenase	<i><u>ndhA</u> <u>ndhB</u> ndhC ndhD ndhE ndhF ndhG ndhH ndhI ndhJ ndhK</i>
Photosystem I.	<i>psaA psaB psaC psal psaj</i>
Photosystem II.	<i>psbA psbB psbC psbD psbE psbF psbH psbI psbJ psbK psbL psbM psbN psbT psbZ</i>
Photosystem I assembly protein	<i><u>ycf3</u> ycf4</i>
Proteins of unknown function	<i>ycf1 ycf2 ycf15</i>
Ribosomal proteins	
Large subunit	<i><u>rpl2</u> rpl14 rpl16 rpl20 rpl22 rpl23 rpl32 rpl33 rpl36</i>
Small subunit	<i>rps2 rps3 rps4 rps7 rps8 rps11 rps12 rps14 rps15 <u>rps16</u> rps18 rps19</i>
RNA polymerase	<i>rpoA rpoB <u>rpoC1</u> rpoC2</i>
Translation factor	<i>infA</i>
Other genes	<i>accD cemA <u>clpP</u> ccs matK</i>
RNA genes	
Ribosomal RNAs	<i>rrn4.5 rrn5 rrn16 rrn23</i>
Transfer RNAs	<i><u>trnA-UGC</u> trnC-GCA trnD-GUC trnE-UUC trnF-GAA trnJ-M-CAU trnG-GCC <u>trnG-UCC</u> trnH-GUG trnI-CAU <u>trnI-GAU</u> trnK-UUU trnL-CAA <u>trnL-UAA</u> trnL-UAG trnM-CAU trnN-GUU trnP-UGG trnQ-UUG trnR-ACG trnR-UCU trnS-GCU trnS-GCU trnS-UGA trnT-GGU trnT-UGU trnV-GAC <u>trnV-UAC</u> trnW-CCA trnY-GUA</i>

2.3. Chloroplast genome assembly

Prior to the de novo assembly of cp genome quality control of the raw paired-end reads (972,060 reads) were done using FastQC [6]. Based on FastQC report the trimming of low quality sequences (quality score < 20; Q20) were filtered out by using a self-developed application, GenoUtils, written in Visual Studio integrated developmental environment with C#. The remaining high quality paired end reads (864,583 reads) were assembled. To create full-length contiguous sequences without the guidance of a reference genome, we obtained *de novo* assembly by applying the overlap-based genome assembler MIRA (version 4.0.2) [7] and Velvet (version 1.2.10) [8]. The assembled contigs were ordered against the complete cp genome of *Helianthus annuus* L. as reference using the Geneious (version 9.1.6) (<http://www.geneious.com>) software [9].

2.4. Gene annotation

The web-based program Dual OrganellarGenoMe Annotator (DOGMA, <http://dogma.cccb.utexas.edu/>) [10] was used to annotate the assembled genome using default parameters to predict protein coding genes, as well as tRNA and rRNA genes. The previously reported *A. artemisiifolia* transcriptome dataset [11] was used to identify the coding regions of cp genes [2]. Subsequently, BLASTN was used to further identify intron-containing gene positions by searching the *de novo* assembled cp genome.

The size of the complete chloroplast genome of *A. artemisiifolia* was found to be 152,215 bp (GC: 37.6%). The cp genome exhibited a quadripartite structure consisting of LSC and SSC regions of 84,399 bp and 17,958 bp respectively, separated by a pair of inverted repeats (IRa and IRb) each being 24,929 bp. A total of 114 genes were annotated including 80 protein coding genes, 30 tRNA genes, and 4 rRNA genes. Six of the protein coding genes and the 3' exon of rps12 are duplicated in the IR regions. Seven of the tRNA genes and all four rRNA genes are also duplicated in the IR regions. The presence of

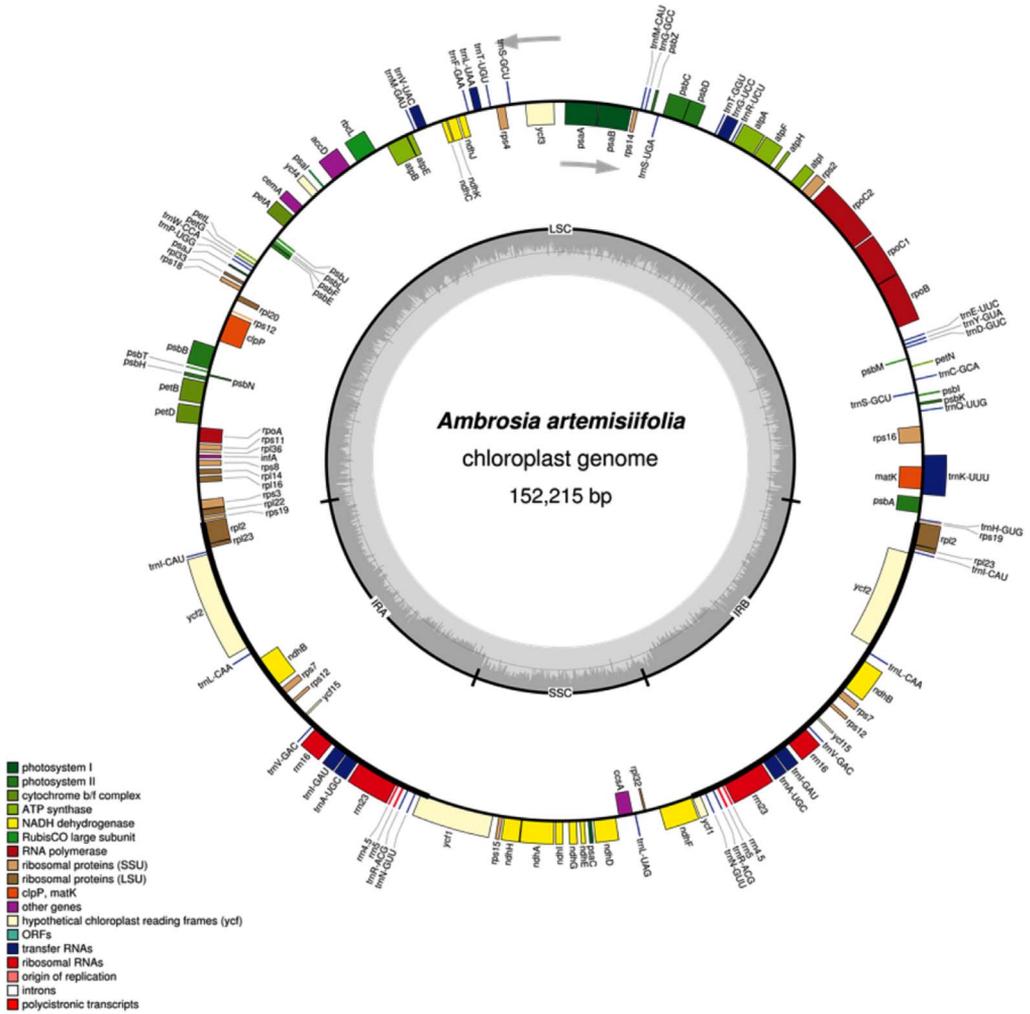


Fig. 2. Physical map of *Ambrosia artemisiifolia* cp genome. The graphical organization was created by OGDRAW [12].

one or two introns were identified in 16 genes, which include 10 protein coding genes and six tRNA genes (Table 1, Fig. 2).

Acknowledgements



This work was supported through the OTKA-K100919 (Hungarian Scientific Research Fund), Postdoc-2014-44 grant of the Hungarian Academy of Sciences (MTA) and the ÚNKP-2016-4-IV of the New National Excellence Program of the Ministry of Human Capacities.

Transparency document. Supplementary material

Transparency document associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.dib.2017.10.009>.

References

- [1] E.P. Fuerst, M.A. Norman, Interactions of herbicides with photosynthetic electron transport, *Weed Sci.* 39 (1991) 458–464.
- [2] B.C. Tripathy, R. Oelmüller, Reactive oxygen species generation and signaling in plants, *Plant Signal. Behav.* 7 (2012) 1621–1633.
- [3] H.G. Sedigheh, M. Mortazavian, D. Norouzian, M. Atyabi, A. Akbarzadeh, K. Hasanpoor, M. Ghorbani, Oxidative stress and leaf senescence, *BMC Res. Notes* 4 (2011) 477.
- [4] R. Lilley, M. Fitzgerald, K. Rienits, D. Walker, Criteria of intactness and the photosynthetic activity of spinach chloroplast preparations, *New Phytol.* 75 (1975) 1–10.
- [5] L. do Nascimento Vieira, H. Faoro, H.P. de Freitas Fraga, M. Rogalski, E.M. de Souza, F. de Oliveira Pedrosa, R.O. Nodari, M. P. Guerra, An improved protocol for intact chloroplasts and cpDNA isolation in conifers, *PLoS One* 9 (2014) e84792.
- [6] S. Andrews, *FastQC: A Quality Control Tool for High Throughput Sequence Data*, 2010.
- [7] B. Chevreux, T. Pfisterer, B. Drescher, A.J. Driesel, W.E. Müller, T. Wetter, S. Suhai, Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs, *Genome Res.* 14 (2004) 1147–1159.
- [8] D.R. Zerbino, E. Birney, Velvet: algorithms for de novo short read assembly using de Bruijn graphs, *Genome Res.* 18 (2008) 821–829.
- [9] M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data, *Bioinformatics* 28 (2012) 1647–1649.
- [10] S.K. Wyman, R.K. Jansen, J.L. Boore, Automatic annotation of organellar genomes with DOGMA, *Bioinformatics* 20 (2004) 3252–3255.
- [11] E. Virág, G. Hegedűs, E. Barta, E. Nagy, K. Mátyás, B. Kolics, J. Taller, Illumina sequencing of common (short) ragweed (*Ambrosia artemisiifolia* L.) reproductive organs and leaves, *Front. Plant Sci.* 7 (2016).
- [12] M. Lohse, O. Drechsel, S. Kahlau, R. Bock, OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets, *Nucleic Acids Res.* 41 (2013) W575–W581.