



# Illumina Sequencing of Common (Short) Ragweed (*Ambrosia artemisiifolia* L.) Reproductive Organs and Leaves

Eszter Virág<sup>1\*</sup>, Géza Hegedűs<sup>2</sup>, Endre Barta<sup>3</sup>, Erzsébet Nagy<sup>1</sup>, Kinga Mátyás<sup>1</sup>, Balázs Kolics<sup>1</sup> and János Taller<sup>1</sup>

<sup>1</sup> Georgikon Faculty, Department of Plant Science and Biotechnology, University of Pannonia, Keszthely, Hungary, <sup>2</sup> Georgikon Faculty, Department of Economic Methodology, University of Pannonia, Keszthely, Hungary, <sup>3</sup> National Agricultural Research and Innovation Centre, Agricultural Biotechnology Institute, Gödöllő, Hungary

**Keywords:** Illumina sequencing, transcriptome, common ragweed, male flower, female flower

## INTRODUCTION

*Ambrosia artemisiifolia* L. (*A. artemisiifolia*, common ragweed) is one of the most aggressive, rapidly spreading and highly allergenic weeds found in many agricultural settings in the temperate zone. Chemical control of ragweed has some limitation in some crops, therefore it may cause reduced yields both in Europe and North America. The most affected are maize, sunflower, soya bean and pea in Europe (Chollet et al., 1999; Konstantinovic et al., 2005; Chauvel et al., 2006) and grains, tobacco and root crops in North America (Bassett and Crompton, 1975). Pollen allergy of common ragweed affects the human population's quality of life. Its pollen allergens are considered to be major elicitors of type I allergy during late summer and fall inducing respiratory distress such as allergic rhinitis and seasonal asthma but it is also linked to eczema, ear infections in children and sinusitis (bacterial infection of the sinuses) in adults (Dykewicz, 2003). The spread of this weed therefore causes severe agricultural and public health problems that are important to be solved globally.

For better understanding the genetic regulation of the common ragweed reproduction biology we sequenced the mRNA of flower tissues and leaves of different developmental stages using the Illumina platform. To this end different gender flowers, of this monoecious, dicotyledonous invasive weed were collected from a natural *Ambrosia* population of a highly infested West-Transdanubian region in Hungary.

The sequence data were assembled de novo to create a reference transcriptome for our future work for this species. Raw reads of the transcriptome assembly have been deposited to NCBI's Sequence Read Archive (SRA) database with the accession numbers SRR3995704 (male flower), SRR3995703 (female flower), and SRR3995705 (leaves). SRA accession: SRP08007. Bioproject ID: PRJNA335689. The Transcriptome Shotgun Assembly project has been deposited at DDBJ/ENA/GenBank under the accession GEZL00000000. The version described in this paper is the first version, GEZL01000000.

The presented data were used for the first time to determine the complete coding sequence and putative signal peptide of an Amb a 3 allergen isoform of *A. artemisiifolia* recently (Taller et al., 2016).

## MATERIALS AND METHODS

### Plant Materials

Three sample types, such as male (♂) and female (♀) flowers from initial to final developmental stages and leaves of different developmental stage and position of *A. artemisiifolia* plants were collected from beside an agricultural field in West-Transdanubian region of Hungary (46°

## OPEN ACCESS

### Edited by:

Joshua L. Heazlewood,  
University of Melbourne, Australia

### Reviewed by:

Rajandeep Sekhon,  
Clemson University, USA  
Manoj K. Sharma,  
Jawaharlal Nehru University, India  
Thiruvarangan Ramaraj,  
National Center for Genome  
Resources, USA

### \*Correspondence:

Eszter Virág  
virag@georgikon.hu

### Specialty section:

This article was submitted to  
Plant Genetics and Genomics,  
a section of the journal  
Frontiers in Plant Science

**Received:** 24 May 2016

**Accepted:** 22 September 2016

**Published:** 07 October 2016

### Citation:

Virág E, Hegedűs G, Barta E, Nagy E,  
Mátyás K, Kolics B and Taller J (2016)  
Illumina Sequencing of Common  
(Short) Ragweed (*Ambrosia*  
*artemisiifolia* L.) Reproductive Organs  
and Leaves. *Front. Plant Sci.* 7:1506.  
doi: 10.3389/fpls.2016.01506

44° 55.4" N, 17° 14' 20.1" E) during the whole flowering period from middle of July to the end of August. Samples were frozen immediately in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until further use. The highly pure total RNA from plant tissues were isolated using TaKaRa Plant RNA Extraction Kit according to manufacturer's instructions (Takara Bio Inc; Japan). Purity and concentration of all RNA samples were quantified spectrophotometrically using Agilent 2100 Bioanalyzer (Agilent Technologies; USA).

## Enrichment of mRNA, cDNA Synthesis and Library Preparation for Illumina HiSeq Paired-End Sequencing

For poly-A based mRNA enrichment and cDNA synthesis the Illumina TruSeq<sup>TM</sup> RNA sample preparation kit (Low-Throughput protocol) was used according to manufacturer's instructions. Briefly, 1.5  $\mu\text{g}$  of total RNA sample of male, female and leaf tissues were used for poly-A mRNA selection using streptavidin-coated magnetic beads. Two rounds of enrichment for poly-A mRNA was performed followed by thermal mRNA fragmentation. The cDNA was synthesized from enriched and chemically fragmented RNA using reverse transcriptase (SuperScript II) and random primers. cDNA was converted into double stranded (ds) DNA using the reagents supplied in the kit and  $\sim 0.5\text{--}1\mu\text{g}$  from each sample were used for the library preparation. The RNA-Seq was performed using Illumina HiSeq2000 system. The hybridization onto a flow cell, the dsDNA fragments were blunt-ended through an end-repair reaction and both ends were ligated to platform-specific double-stranded bar-coded adapters. The samples were run in one lane using multiple indexing adapters. For library amplification an adapter-selective PCR reaction was performed. In order to avoid the skewing of the library representation the number of PCR cycles was minimized to 15. The optimum cluster density of libraries was created by qPCR according to *qPCR Quantification Protocol Guide*. The size and purity of the samples were checked by Agilent 2100 Bioanalyzer (Agilent Technologies, USA). The PCR products were at 260 bp, approximately. The DNA libraries were multiplexed normalizing them to 10 nM.

## De Novo Assembling and Analysis of High Throughput Sequencing Data

Quality control and preprocessing of the  $2 \times 100$  bp raw reads was done with FastQC (Andrews, 2010) and are summarized in **Table 1**. After quality trimming at a Phred score  $\geq 28$  a de novo assembly of the combined three transcriptome sequencing datasets was performed using Trinity (Haas et al., 2013) with 25 k-mer size. The resulted contigs of the assembly were used later as the reference transcriptome of the common ragweed. The statistics of the assembly are summarized in **Table 2**. The reads from the samples were then mapped separately back to reference using the short read aligner program Bowtie (Langmead et al., 2009). The transcript abundances were estimated using the RSEM program (<http://deweylab.github.io/RSEM/>) with the scripts provided in the Trinity package. A collection of 162494 unigenes (average length of 391 bp) was generated and their annotation

**TABLE 1 | Quality control of raw reads obtained by HiSeq2000.**

Sample	Total raw reads	Read length	Average Phred score	GC%
♂ flower	24110256	100	36	52
♀ flower	23264636	100	36	54
leaf	24330693	100	37	46

**TABLE 2 | Statistics of transcriptome assembling.**

Attributes	Value
Total trinity genes	162494
Total trinity transcripts	229116
GC%	41.61
Contig N50	774
Median contig length	391
Total assembled reads	1377646

was performed using Trinotate suit (<https://trinotate.github.io/>). We found that 72.46 and 68,4 % of the unigenes had significant similarity with proteins in the National Center for Biotechnology Information (NCBI) non-redundant protein database (Nr) and Swiss-Prot database respectively.

## DIRECT LINK TO DEPOSITED DATA AND INFORMATION TO USERS

The raw reads are available in Fastq format at the following link <http://www.ncbi.nlm.nih.gov/sra/SRP080078>.

The transcriptome shotgun assembly is available in FASTA format at the following link <http://www.ncbi.nlm.nih.gov/nuccore/GEZL000000000>.

Users can download and use the data freely for research purpose only with quoting this paper as reference to the data.

## AUTHOR CONTRIBUTIONS

EV: wrote the manuscript, performed cDNA library preparation, bioinformatics and analysis on the data; EB: performed transcriptome assembling and bioinformatics; GH: performed statistical analysis and informatics operations; EN, BK, and KM: collected the samples and performed RNA isolation; JT: supervised the project and acquired funding.

## ACKNOWLEDGMENTS

This work was supported by the K100919, Postdoc-2014-44 and UNKP-16-4 grants of the Hungarian Scientific Research Fund (OTKA), Hungarian Academy of Sciences (MTA) and New National Excellence Programme, Ministry of Human Resources of the Hungarian Government. The authors acknowledge support from EU COST Action FA1203 "Sustainable management of *Ambrosia artemisiifolia* in Europe (SMARTER)."

## REFERENCES

- Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Bassett, I. J., and Crompton, C. W. (1975). The biology of canadian weeds. 11. *Ambrosia artemisiifolia* L. and *A. psilostachya* DC. *Can. J. Plant Sci.* 55, 463–476. doi: 10.4141/cjps75-072
- Chauvel, B., Dessaint, F., Cardinal-Legrand, C., and Bretagnolle, F. (2006). The historical spread of *Ambrosia artemisiifolia* L. in France from herbarium records. *J. Biogeogr.* 33, 665–673. doi: 10.1111/j.1365-2699.2005.01401.x
- Chollet, D., Drieu, Y., Molines, J., and Pauget, J. (1999). Comment lutter contre l'ambrosie à feuilles d'armoise. *Perspect. Agricoles* 250, 78–82.
- Dykewicz, M.S. (2003). 7. Rhinitis and sinusitis. *J. Allergy Clin. Immunol.* 111, S520–S529. doi: 10.1067/mai.2003.82
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084
- Konstantinovic, B., Meseldzija, M., Stojšin, V., Bagi, F., and Balaz, F. (2005). *Integral Control of Ambrosia artemisiifolia* L. in the Area of the City of Novi Sad [Serbia (Serbia and Montenegro)]. Belgrade: Savremena Poljoprivreda.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25. doi: 10.1186/gb-2009-10-3-r25
- Taller, J., Decsi, K., Farkas, E., Nagy, E., Mátyás, K. K., Kolics, B., et al. (2016). *De novo* Transcriptome Sequencing Based Identification of Amb a 3-like Pollen Allergen in Common Ragweed (*Ambrosia artemisiifolia*). *J. Bot. Sci.* 5, 12–16. Available online at: <http://www.rroij.com/open-access/de-novo-transcriptome-sequencing-based-identification-of-amb-a-3like-pollen-allergen-in-common-ragweed-ambrosia-artemisiifolia-.pdf>

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Virág, Hegedűs, Barta, Nagy, Mátyás, Kolics and Taller. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.