Genomics Data 11 (2017) 118–119

Contents lists available at ScienceDirect

Genomics Data

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

Data in Brief

De novo transcriptome assembly of heavy metal tolerant *Silene dioica*

Radim Cegan [a,*], Vojtech Hudzieczek [a], Roman Hobza [a,b]

[a] Department of Plant Developmental Genetics, Institute of Biophysics of the CAS, v.v.i., Kralovopolska 135, CZ-612 00 Brno, Czech Republic
[b] Centre of Plant Structural and Functional Genomics, Institute of Experimental Botany of the CAS, v.v.i., Slechtitelu 31, CZ-78371 Olomouc-Holice, Czech Republic

ARTICLE INFO

Article history:
Received 13 January 2017
Accepted 25 January 2017
Available online 26 January 2017

Keywords:
Silene dioica
RNA-Seq
Transcriptome
Heavy metal tolerance
Sex chromosomes

ABSTRACT

Silene dioica is a dioecious plant of the family *Caryophyllaceae*. In the present study, we used Illumina sequencing technology (MiSeq) to sequence, *de novo* assembly and annotate the transcriptomes of male and female copper tolerant *S. dioica* individuals. We sequenced the normalized mRNA of roots, shoots, flower buds and flowers for each sex. Raw reads of the transcriptome assembly project for *S. dioica* male and female individual have been deposited in NCBI's Sequence Read Archive (SRA) database with the accession number SRP094611. The Trinity and Detonate program was used to *de novo* assembly 92,347 transcripts for male and 94,757 transcripts for female transcriptome. The assembled transcriptome sequences for *S. dioica* male and female individuals can be accessed at NCBI with the following accession numbers: GFCG00000000 (male); GFCH00000000 (female). The obtained transcriptomic data will be useful for further studies focusing on copper tolerance, comparative transcriptome analysis with other *Silene* species and sex chromosomes evolution.

Specifications

Organism/cell line/tissue	*Silene dioica*/roots, shoots, flower buds and flowers
Sex	Male and female (separately)
Sequencer or array type	Illumina MiSeq
Data format	Raw data: FASTQ file
Experimental factors	*De novo* transcriptome assembly of *S. dioica* male and female individuals
Experimental features	Roots, shoots, flower buds and flowers of male and female *S. dioica* individuals were separately collected for RNA isolation, normalization, sequencing, *de novo* transcriptome assembly and annotation
Consent	N/A
Sample source location	Piesky, Slovakia, 48°49′04.5″N 19°07′52.6″E

1. Direct link to deposited data

https://www.ncbi.nlm.nih.gov/sra/SRP094611
https://www.ncbi.nlm.nih.gov/nuccore/GFCG00000000.1
https://www.ncbi.nlm.nih.gov/nuccore/GFCH00000000.1.

* Corresponding author.
E-mail address: cegan@ibp.cz (R. Cegan).

2. Introduction

Silene dioica (L.) Clairv. a genus of *Elisanthe* belonging to the family *Caryophyllaceae*, previously known as *Melandrium rubrum* is historically one of the first studied plants with adaptation to increased heavy metal soil concentration, namely copper [1]. This dioecious species possesses heteromorphic sex chromosomes with human-like sex determination (male XY, female XX; [2]). *S. dioica* and closely related species is extensively studied from various aspects such as sex chromosome evolution, sexual dimorphism, hybrid zone formation and sexually transmitted diseases (for review see [3]) but only limited transcriptomic data are available for heavy metal tolerance traits. We selected male and female individuals from the copper tolerant ecotype Piesky (Slovakia) to establish transcriptomic resources. We sequenced the normalized mRNA of roots, shoots, flower buds and flowers for each sex for subsequent studies. The sequence data were *de novo* assembled to create a reference transcriptome for this species. This article documents the public availability of complex transcriptomic resources of the copper tolerant dioecious plant, *S. dioica*.

3. Experimental design, materials and methods

3.1. Plant material

S. dioica seeds were collected at an old copper mine dump locality Piesky (Slovakia; 48°49′04.5″N 19°07′52.6″E). The seeds were germinated on a mist bench. After germination, seedlings were transferred to an aeroponic propagator, filled with modified Hoagland's nutrient solution [4]. For the experiments, an aeroponic culture in growth chamber

http://dx.doi.org/10.1016/j.gdata.2017.01.004

Table 1
S. dioica sequencing and read processing.

Sex	Raw reads	Trimmed reads (%)	Merged reads (%)	Accession number
T17 - male	5,706,442	98.73	83.66	SRR5079457
T18 - female	7,145,611	98.59	81.22	SRR5079458

(25 °C, 12 h photoperiod, 75% humidity) was used. These conditions were maintained during the cultivation and the solution was changed twice a week.

3.2. RNA isolation, library preparation, and RNA sequencing

Before RNA extraction, the sex of *S. dioica* plants was determined using sex specific primers [5]. Total RNA was extracted from male and female individual using the NucleoSpin RNA Plant kit (Macherey-Nagel, Germany) according to the standard protocol separately for roots, leaves, flower buds and flowers. Extracted RNA was quantified using a NanoDrop 2000 (Thermo Scientific). Samples from the same individual were pooled in equal ratio to well represent male and female transcriptome. Male and female total RNA (25 μg each) were provided to GATC Biotech Konstanz for construction of random primed normalized cDNA libraries with poly(A)+ selection and sequencing. Poly(A)+ RNA isolated from the total RNA samples were fragmented with ultrasound (2 pulses of 30 s at 4 °C). First-strand cDNA synthesis was primed with a N6 randomized primer. The Illumina TruSeq sequencing adapters were then ligated to the 5′ and 3′ ends of the cDNA. The cDNA was finally amplified with PCR using proofreading enzyme. Normalization was carried out by one cycle of denaturation and reassociation of the cDNA. Reassociated ds-cDNAs were separated from the remaining ss-cDNAs (normalized cDNA) by passing the mixture over a hydroxylapatite column. After hydroxylapatite chromatography, the ss-cDNAs were PCR amplified. For Illumina sequencing, the normalized cDNA in the size range of 400–600 bp was eluted from preparative agarose gels. The resulting libraries were sequenced on Illumina MiSeq (PE, 2 × 250), using standard protocol.

3.3. Data pre-processing, de novo transcriptome assembly and annotation

The quality of sequencing reads was inspected in FastQC [6]. Read trimming on quality (Q30) and sequencing adaptors removal were done with Trimmomatic −0.32 [7]. Due to library preparation and sequencing strategy, quality trimmed reads for both datasets were further merged by FLASH tool v1.2.9 [8]. Transcriptome sequencing and read processing are summarized in Table 1. Merged reads for both datasets were assembled separately using Trinity v2.0.6 [9]. Further, we performed several transcriptome assembly quality assessments based on the recommendations of Trinity developer web pages. Firstly, to obtain basic statistical information over assemblies, we ran the Trinity package utility script *TrinityStats.pl*. Then we ran RSEM-EVAL from DETONATE software package [10]. Based on RSEM-EVAL *contig impact score* we removed all contigs with negative score from the assemblies as recommended in Li et al., [10] and reran *TrinityStats.pl* script and RSEM-EVAL. The results of these assessments are summarized in Table 2.

Table 2
Statistics of *S. dioica* transcriptome assembly after RSEM-EVAL filtering.

Feature	T17 - male	T18 - female
Total trinity genes	80,903	83,591
Total trinity transcripts	92,347	94,757
GC%	40.17	40.46
Median contig length	761	726
Predicted proteins	56,849	57,976

Finally, male reference transcriptome consisted of 92,347 transcripts and female 94,757 transcripts. Functional annotation of transcript assemblies was performed using the Blast2GO [11] and coding regions within transcripts were identified by TransDecoder [12] (Table 2.).

Conflict of interest

The authors declare no conflicts of interest.

Acknowledgement

This work was financially supported by the Czech Science Foundation (grant no. 13-34962P). Computational resources were provided by the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085, provided under the programme “Projects of Large Research, Development, and Innovations Infrastructures”.

References

[1] S. Prát, Die Erblichkeit der Resistenz gegen Kupfer. Ber. Deut. Bot. Ges. 52 (1934) 65–67, http://dx.doi.org/10.1111/j.1438-8677.1934.tb01655.x.
[2] B. Vyskot, R. Hobza, Gender in plants: sex chromosomes are emerging from the fog. Trends Genet. 20 (2004) 432–438, http://dx.doi.org/10.1016/j.tig.2004.06.006.
[3] G. Bernasconi, J. Antonovics, A. Biere, D. Charlesworth, L.F. Delph, D. Filatov, et al., Silene as a model system in ecology and evolution. Heredity 103 (2009) 5–14, http://dx.doi.org/10.1038/hdy.2009.34.
[4] H. Schat, W.M.T. Bookum, Genetic control of copper tolerance in *Silene vulgaris*. Heredity 68 (1992) 219–229, http://dx.doi.org/10.1038/hdy.1992.35.
[5] R. Hobza, A. Widmer, Efficient molecular sexing in dioecious *Silene latifolia* and *S. dioica* and paternity analysis in F1 hybrids. Mol. Ecol. Resour. 8 (2008) 1274–1276, http://dx.doi.org/10.1111/j.1755-0998.2008.02344.x.
[6] S. Andrews, FastQC a quality-control tool for high-throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastq.
[7] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30 (2014) 2114–2120, http://dx.doi.org/10.1093/bioinformatics/btu170.
[8] T. Magoc, S.L. Salzberg, FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27 (2011) 2957–2963, http://dx.doi.org/10.1093/bioinformatics/btr507.
[9] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, et al., Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29 (2011) 644–652, http://dx.doi.org/10.1038/nbt.1883.
[10] B. Li, N. Fillmore, Y. Bai, M. Collins, J.A. Thomson, R. Stewart, et al., Evaluation of de novo transcriptome assemblies from RNA-Seq data. Genome Biol. 15 (2014)http://dx.doi.org/10.1186/s13059-014-0553-5.
[11] A. Conesa, S. Gotz, J.M. Garcia-Gomez, J. Terol, M. Talon, M. Robles, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21 (2005) 3674–3676, http://dx.doi.org/10.1093/bioinformatics/bti610.
[12] B.J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P.D. Blood, J. Bowden, et al., De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8 (2013) 1494–1512, http://dx.doi.org/10.1038/nprot.2013.084.