



Data Article

Small RNA-seq dataset of wild type and 16C *Nicotiana benthamiana* leaves sprayed with naked dsRNA using the high-pressure spraying technique

Kübra Çalışır^a, Gabi Krczal^b, Veli Vural Uslu^{b,c,*}^a Molecular Biology and Genetics Department, Bilkent University, Ankara, Turkey^b Bioeconomy Department, RLP AgroScience GmbH, Neustadt an der Weinstraße, Germany^c Molecular and Applied Plant Sciences, Center for Organismal Studies, Heidelberg University, Heidelberg, Germany

ARTICLE INFO

Article history:

Received 8 June 2022

Revised 14 October 2022

Accepted 21 October 2022

Available online 27 October 2022

Dataset link: [GSE160110 \(Original data\)](https://www.ebi.ac.uk/bioproject/584448)

Keywords:

RNAi
dsRNA
sRNA-seq
exoRNA
SIGS
PTGS

ABSTRACT

Double-stranded RNA (dsRNA) applications have emerged as promising alternatives to chemical plant pesticides. It has been proposed that the protective effect of dsRNA is mediated by the RNA interference (RNAi) mechanism. Small RNAs (sRNAs) are one of the landmarks of RNAi mechanisms. Two classes of sRNAs appear upon RNAi, triggered by dsRNA: The cleavage products of the dsRNA mapping directly to the dsRNA sequence and the transitive sRNAs mapping to the target transcript outside of the dsRNA sequence. Therefore, the sRNA-seq data obtained from dsRNA-treated plants have been exclusively analysed in the context of the target genes and the outcome has been considered essential to evaluate the underlying mechanism of dsRNA mediated plant protection. Using high-pressure spraying technology (HPST), we have applied a GFP targeting 139bp-long dsRNA on wild type (WT) and GFP expressing (16C) *Nicotiana benthamiana* plants in biological triplicates. As a control, we applied water with HPST on 16C *N. benthamiana*. We have acquired sRNA-seq data on the treated and control leaves 5 days post spraying. In this dataset, we have expanded our sRNA-seq analysis

* Corresponding author.

E-mail address: velivural.uslu@agrosience.rlp.de (V.V. Uslu).

from the target GFP transgene sequence to the whole transcriptome of *N. benthamiana* to provide the community with a resource for the small RNA landscape after high-pressure spraying in 16C and WT samples. Furthermore, we have provided a comparison of sRNA landscape between WT and 16C lines.

© 2022 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Specifications Table

Subject	Agricultural and Biological Sciences
Specific subject area	Omics: Transcriptomics, Agronomy and Crop Science, Biotechnology
Type of data	Table, Chart, Graph
How the data were acquired	<ol style="list-style-type: none"> 1. miRVANA miRNA extraction kit has been used to extract small RNAs. 2. The samples have been sent to GeneXPro GmbH (Frankfurt/Germany) for RNA sequencing. 3. Small RNA libraries were prepared using TrueQuant Small RNAseq kit. 4. Illumina NextSeq500 was used for sequencing the sRNA-seq libraries (75 cycles) 5. The quality of the data has been checked by FastQC. sRNA-seq libraries were mapped onto CaMV35S promoter, 16C <i>N. benthamiana</i> GFP sequence, NOS terminator, which were obtained from [1] in the original paper. 6. For the current genome wide data analysis, we have performed a pseudo alignment using Kallisto tool (0.46.2) onto the reference transcriptome Niben101_annotation.transcripts.fasta.gz 7. We have performed comparisons between WT and 16C <i>N. benthamiana</i> sRNA signatures, both treated with dsRNA using statistical analysis run by Samtools tool on bam files.
Data format	Filtered raw reads Analysed RNA-seq files (excel)
Description of data collection	<ol style="list-style-type: none"> 1. dsRNA was produced using MEGAscript RNAi Kit (www.thermofisher.com) using a PCR amplified template as mentioned in [2]. 2. 200µl DsRNA- (20ng/µl) have been sprayed on 10-12cm high three healthy WT and three healthy 16C <i>N. benthamiana</i> leaves at 5-6 bar pressure. As a control 200µl water has been sprayed on two leaves of a single 16C <i>N. benthamiana</i> in the same conditions. Treated and control leaf samples have been harvested at 5 days post-spraying. Two leaves have been used for each biological replicates. High-pressure spraying technique creates a radial gradient of spraying. The highest pressure is in the center of the sprayed area and the pressure decreases away from the center. The very center of the sprayed area is generally necrotic and we removed these sites. In our set-up the visible sprayed area was 0.5cm and outside of this area was manually cut out by a razor to enrich the sample with sprayed areas as much as possible. Completely dead areas due to severe wounding in the center were excluded based on the autofluorescence under UV light. The size and the severity of necrotic areas were not visibly different between 16C and WT <i>N. benthamiana</i> and between water and dsRNA treatments. However, the younger leaves showed stronger necrosis, when compared the older leaves. Therefore, the collected leaves were a mix of young and old leaves. The size of the collected for each biological replicate areas did not exceed 1cm², approximately fitting to the lid area of 2ml Eppendorf tube. <p>Leaf samples have been crushed using metal beads in Qiagen Retsch TissueLyser for 1 minutes (30rps).</p>

(continued on next page)

Data source location	16C and WT seeds were originally provided by D.Baulcombe, Cambridge, UK. 16C line was first established in [3] and sequenced in [1]. The plants were grown at RLP AgroScience GmbH Greenhouse Chamber 5. RLP AgroScience GmbH, Breitenweg 71, 67435 Neustadt-Mussbach, Germany (49.3698408N,8.1871924E)
Data accessibility	All the raw data has been submitted to the GEO servers and accessible under Series GSE160110. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE160110 The series contains 7 datasets, which are given below: GSM4859998 dsRNA-midGFP-sprayed 16C leaf 5dps_bioReplicate_1 GSM4859999 dsRNA-midGFP-sprayed 16C leaf 5dps_bioReplicate_2 GSM4860000 dsRNA-midGFP-sprayed 16C leaf 5dps_bioReplicate_3 GSM4860001 dsRNA-midGFP-sprayed WT-Nb leaf 5dps_bioReplicate_1 GSM4860002 dsRNA-midGFP-sprayed WT-Nb leaf 5dps_bioReplicate_2 GSM4860003 dsRNA-midGFP-sprayed WT-Nb leaf 5dps_bioReplicate_3 GSM4860004 water-sprayed 16C leaf 5dps_bioReplicate_1
Related research article	Uslu VV, Bassler A, Krczal G, Wassenegger M. High-Pressure-Sprayed Double Stranded RNA Does Not Induce RNA Interference of a Reporter Gene. <i>Front Plant Sci</i> (2020);11:534391. PMID: 33391294

Value of the Data

- Here we provide sRNA-seq data from dsRNA treated WT and 16C *N. benthamiana*. In the original research article, this dataset was solely used for the sRNAs mapping to the GFP transgene in 16C plants [2]. However, this data contains sRNAs mapping to the coding genes.
- sRNA-seq data will be useful to researchers who use the WT and 16C *N. benthamiana* lines comparatively, especially in the context of exogenous RNA applications. The data shows differences between WT and 16C sRNA landscapes.
- sRNA-seq data, especially after dsRNA administration by spraying, is difficult to normalize [4] and we propose that such genome-wide sRNA-seq data could assist researchers to normalize the processed dsRNA sequences.

The data also hints at the minor differences between water-treated and dsRNA-treated 16C plants at a transcriptome-wide level.

It is known that GFP transgene in the 16C line is resistant to silencing. Comparison of WT and 16C sRNA-seq data may contribute to the understanding of this unique phenotype.

1. Data Description

In this manuscript, the sRNA-seq data from dsRNA and water treated 16C *N. benthamiana*, and dsRNA treated WT *N. benthamiana* has been analysed at the whole transcriptome level. First, we have provided the basic sequencing statistics using FastQC. The sample name and sample descriptions are provided together with the percentage of duplicated reads (%Dups), CG content (%GC), the average length of the reads (Length), and total read number in millions (M Seq) (Table 1). The average duplication score is 53.6%, with a standard deviation of 3.4%. The average GC content is 48%, with a standard deviation of 0.8%. The average read length is 41bp.

Our analysis shows that on average nine percent of the sRNA-seq reads map to the transcriptome (with a standard deviation of one percent). Nevertheless, the sRNA reads map on average 15900 distinct transcripts in *N. benthamiana* (Supplementary Table 1).

Once we performed the mapping of the sRNA-reads to *N. benthamiana* transcriptome, we obtained comparative information, which was extracted from bam files with Samtools tool. Differential gene analysis was performed comparing 16C samples to WT. 298 genes were differentially regulated (considered p-adjusted value < 0.05), of which 51 of them were upregulated and

Table 1

Sequencing Outcome Sequencing features of the dsRNA treated 16C samples, dsRNA treated WT samples and the water treated 16C sample. FastQC is used to obtain the ratio of duplicated reads (%Dups), the abundance of CG (%GC), the average length of the reads (Length), and the total read number in millions (M Seq).

Sample Name	Sample Description	% Dups	% GC	Length	M Seqs
SRR12898887	dsRNA-midGFP-sprayed 16C leaf 5dps_bioReplicate_1	57.50%	49%	42 bp	7.2
SRR12898888	dsRNA-midGFP-sprayed 16C leaf 5dps_bioReplicate_2	53.30%	48%	41 bp	7.5
SRR12898889	dsRNA-midGFP-sprayed 16C leaf 5dps_bioReplicate_3	56.70%	49%	41 bp	7.4
SRR12898890	dsRNA-midGFP-sprayed WT-Nb leaf 5dps_bioReplicate_1	55.90%	48%	42 bp	8.4
SRR12898891	dsRNA-midGFP-sprayed WT-Nb leaf 5dps_bioReplicate_2	52.10%	48%	40 bp	8.8
SRR12898892	dsRNA-midGFP-sprayed WT-Nb leaf 5dps_bioReplicate_3	51.50%	47%	41 bp	8.4
SRR12898893	water-sprayed 16C leaf 5dps_bioReplicate_1	48.00%	47%	38 bp	7.7

Table 2

sRNA-seq alignment to *N. benthamiana* transcriptome. Reads from 7 samples given in Table 1 are aligned to *N. benthamiana* transcriptome. The number of total reads, the number of reads mapped to the *N. benthamiana* transcriptome, the number of unmapped reads, and the number of genes after removal of genes with zero gene expression that sRNA reads map are provided.

Samples	# of Total Reads	# of Mapped Reads	# of Unmapped Reads	# of genes (no zero)
SRR12898887	7186316	694625	6491691	14275
SRR12898888	7511064	673718	6837346	16533
SRR12898889	7375715	789397	6586318	13431
SRR12898890	8363721	811478	7552243	20101
SRR12898891	8824820	710397	8114423	17347
SRR12898892	8434532	733795	7700737	17870
SRR12898893	7718149	598499	7119650	12177

247 genes were downregulated. A heatmap of the whole transcriptome was constructed based on log CPM values (Fig. 1). The heatmap shows that the dsRNA-treated 16C samples clustered with the water-treated 16C plants, rather than the dsRNA-treated WT samples, suggesting that the genetic background plays a more important role than the treatment. Most of the differentially expressed genes had low read counts. The differentially expressed genes were visualized by volcano plots (Fig. 2). The top 15 genes were annotated on the plot. The genes downregulated and upregulated in 16C when compared to WT have been provided in supplementary table 2 and supplementary table 3, respectively. Gene Ontology (GO) terms associated with the differentially expressed genes are also provided in these tables. We have analyzed the GO terms among upregulated and downregulated genes. Interestingly, several clusters appear in the GO term analysis among the genes downregulated in 16C, when compared to WT. For example, methyltransferase activity, RNA and DNA binding properties, and hydrolase activity are among the GO terms enriched for the genes downregulated in dsRNA-treated 16C, when compared to dsRNA-treated WT samples (Fig. 3). On the other hand, acetyltransferase activity and glycosyl transferase activity are among the few enriched GO term clusters associated with upregulated genes in 16C.

The analysis of the sRNA dataset at a transcriptome level is critical for understanding one of the unique features of the 16C line, which lack suppression of the GFP transgene even after tens of generations. We have previously shown that the 16C line contains, only sense strand sRNAs mapping to the GFP, which is an indication of degradation products rather than active post-transcriptional silencing of the GFP. Moreover, comparing the dsRNA sprayed WT and 16C samples, we did not detect any differences in the sRNA population, which map to GFP sequence. However, here the analysis points out that the 16C sRNA landscape differs from the WT sRNA landscape, irrespective of the dsRNA treatment.

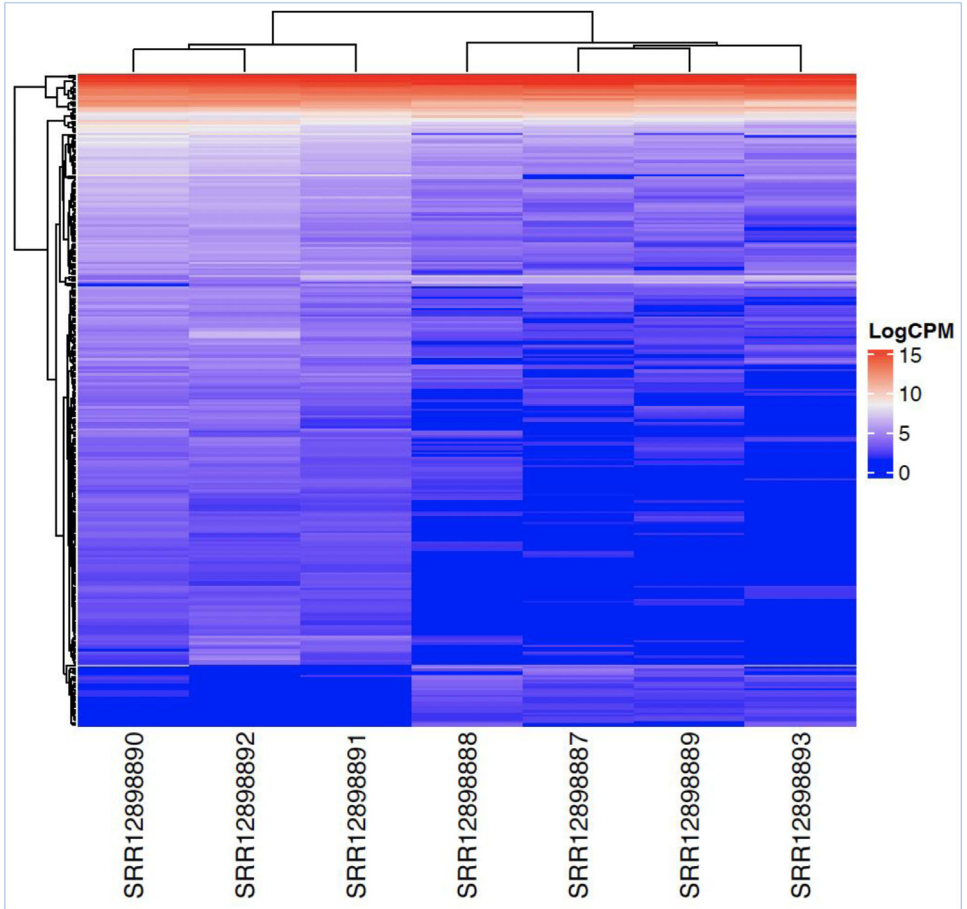


Fig. 1. Heatmap of differentially expressed genes The heatmap shows the clustering of the 298 differentially regulated genes (p -adjusted < 0.05) among the samples. The color code given on the right side indicates the counts per million in logarithmic scale of the differentially expressed genes for each sample.

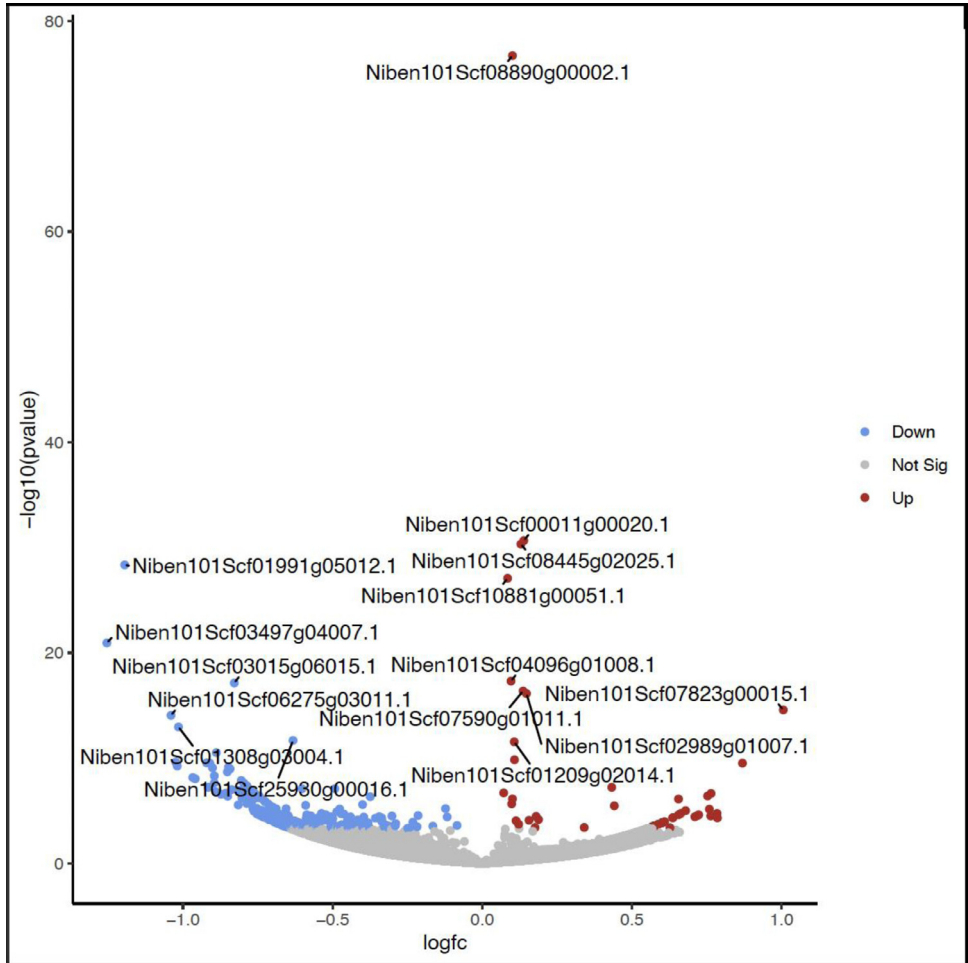


Fig. 2. The volcano plot and the IDs of the transcripts sRNAs abundance in 16C samples, mapping to the transcripts is compared with the WT samples. On the Y-axis, the p-value in the logarithmic scale is plotted, so that the transcripts in the upper part are more significantly deregulated than the ones in the bottom part. On the X-axis, the difference in fold change in expression level is plotted. Blue dots indicate the 247 downregulation (under-representation) and the 51 red dots indicate upregulation (over-representation) of the sRNAs mapping to the given transcripts in 16C.

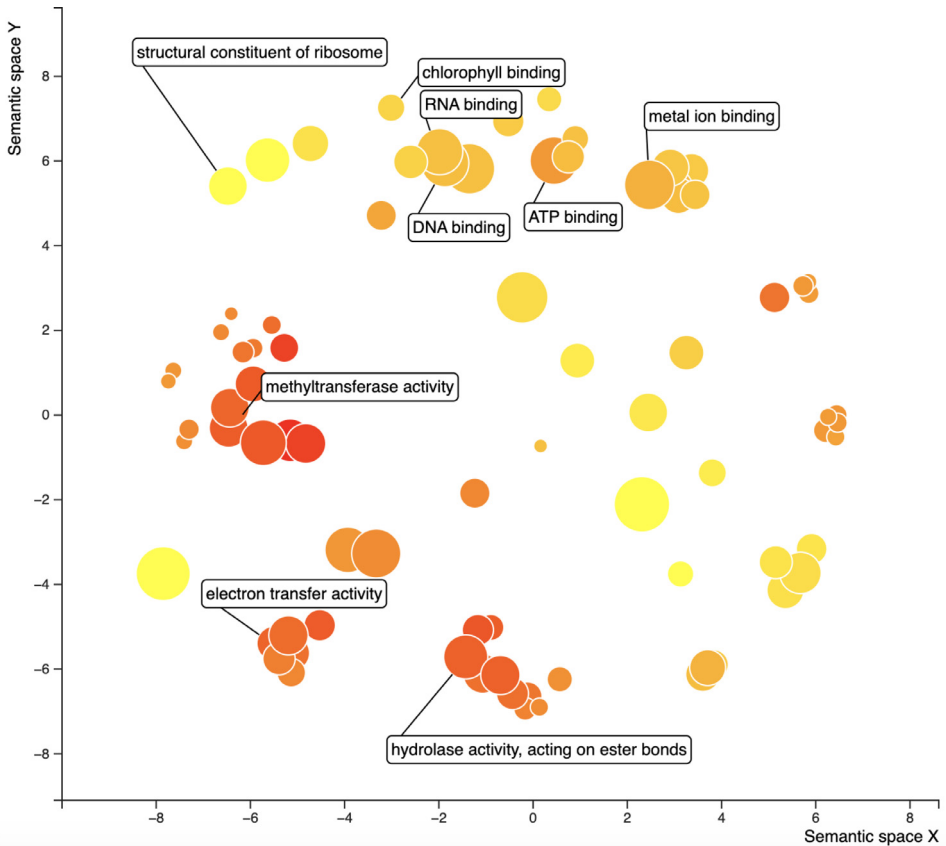


Fig. 3. GO term analysis of the genes underrepresented in 16C sRNA profiles. Gene ontology analysis by REVIGO has been shown for the transcripts underrepresented in 16C samples when compared to WT samples. The X- and Y-axis indicate semantic space distribution based on REVIGO annotations.

2. Experimental Design, Materials and Methods

2.1. Experimental design

N. benthamiana plants were grown at 28°C in long-day conditions, with an automated 48-hour watering cycle. 10-12cm tall *N. benthamiana* leaves were sprayed with 139bp long dsRNA, mapping to the mid-GFP region of 16C *N. benthamiana* [2]. We have also sprayed WT *N. benthamiana* plants to be able to assess whether the dsRNAs were taken up and processed by the intracellular RNAi machinery. However, by analyzing the distribution of the sRNA-seq reads, we could not detect any further processing. To our surprise, we could detect sense strand sRNAs mapping to the GFP in water sprayed 16C *Nicotiana benthamiana*, suggesting that the sRNA-sequencing can reflect the gene expression in the sprayed leaves. sRNA-seq data has been previously used to understand the transcriptomic response to nutrient stress in *Arabidopsis thaliana* [5]. Therefore, we expanded our analysis onto the whole transcriptome in *N. benthamiana* to compare water-treated 16C and dsRNA-treated 16C, as well as dsRNA treated WT and dsRNA treated 16C. We have detected reads mapping to over fifteen thousand genes on average, suggesting that it is possible to extract transcriptome information. An important experimental design detail we would like to point out is the biological replicates. In this work, we primarily

aimed at comparing 16C and WT *N.benthamiana* small RNA populations, which we perform by using 3 biological replicates each. However, the comparison is simply a ratio-based analysis and the ratios can be misleading if a baseline is not provided. Therefore, we provided water treated 16C dataset to the sample to draw a baseline for the transcript levels. Since single water treated 16C sample is not used for any statistical comparison, rather highlighting the baseline expression, we are confident that including the single water-treated sample will help analyze the data more meaningfully.

2.2. Limitations and future perspectives of the dataset

The transcript targets make up less than ten percent of each dataset. Therefore, we are confident that the dataset could also hint at the epigenetically regulated genomic sequences. The most abundant reads are 24nt long, which are known to involve in the transcriptional gene silencing mechanism [6]. However, a comprehensive annotation of the *N. benthamiana* genome is currently not available in our hands to fully investigate the nature of these 24nt sRNAs, in our dataset. In addition, in our dataset, we did not observe dramatic changes between water-treated and dsRNA-treated samples. Noteworthy that our experimental setup does not allow firm conclusions on the difference between water treated samples and dsRNA treated samples as we have a single sample for water treated 16C.

Another limitation of the dataset is the transcriptional wounding response. All the datasets are collected upon high-pressure spraying, which is known to induce wounding. Therefore, although the comparison of the datasets will mask the effect of wounding response, each dataset should contain signs of transcriptional changes related to wounding response.

2.3. Bioinformatic Analysis

Raw sequencing files, data quality was checked with the FastQC tool (Babraham Bioinformatics). The complete FastQC analysis is provided as a collection of charts in pdf format (Supplementary data 1). The sRNA-seq data was pseudo aligned and quantified with the Kallisto tool (version 0.46.2) to the reference transcriptome (Niben101_annotation.transcripts.fasta.gz) [7] [8]. Differential gene analysis was performed with the limma package in R, comparing 16C samples to WT. For significance, adjusted values were considered (adjusted p-value < 0.05). The heatmap was constructed with logCPM values using the ComplexHeatmap package. volcano plots to visualize differential representation of the transcript mapping sRNAs were obtained using the ggplot2 package. Gene ontology analyses were performed using REVIGO (<http://revigo.irb.hr/>) [9].

Ethics Statements

The experiments were conducted in *N. benthamiana* plants, which did not require any ethical statements.

CRedit Author Statement

Veli Vural Uslu: Conceptualization, Methodology, Execution of the Experiments, Writing – original draft preparation, Revision, Supervision; **Kübra Çalışır:** Data curation, Formal Analysis; **Gabi Krczal:** Conceptualization, Methodology, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

[GSE160110 \(Original data\)](#) (GEO).

Acknowledgments

This work was supported by the FELR Antrag-Nr: 137; RNA Stammapplikationsmethode

Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.dib.2022.108706](https://doi.org/10.1016/j.dib.2022.108706).

References

- [1] J.G. Philips, F. Naim, M.T. Lorenc, K.J. Dudley, R.P. Hellens, P.M. Waterhouse, The widely used *Nicotiana benthamiana* 16c line has an unusual T-DNA integration pattern including a transposon sequence, *PLoS One* 12 (2) (2017) e0171311.
- [2] V.V. Uslu, A. Bassler, G. Krczal, M. Wassenegger, High-Pressure-Sprayed Double Stranded RNA Does Not Induce RNA Interference of a Reporter Gene, *Front. Plant Sci.* 11 (2020) 534391.
- [3] M.T. Ruiz, O. Voinnet, D.C. Baulcombe, Initiation and maintenance of virus-induced gene silencing, *Plant Cell* 10 (6) (1998) 937–946.
- [4] V.V. Uslu, A. Dalakouras, V.A. Steffens, G. Krczal, M. Wassenegger, High-pressure sprayed siRNAs influence the efficiency but not the profile of transitive silencing, *Plant J.* 109 (5) (2022) 1199–1212.
- [5] G. Liang, Q. Ai, D. Yu, Uncovering miRNAs involved in crosstalk between nutrient deficiencies in Arabidopsis, *Sci. Rep.* 5 (2015) 11813, doi:[10.1038/srep11813](https://doi.org/10.1038/srep11813).
- [6] X.J. He, T. Chen, J.K. Zhu, Regulation and function of DNA methylation in plants and animals, *Cell Res.* 21 (3) (2011) 442–465.
- [7] N.L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification, *Nat. Biotechnol.* 34 (5) (2016) 525–527.
- [8] A. Bombarely, H.G. Rosli, J. Vrebalov, P. Moffett, L.A. Mueller, G.B. Martin, A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research, *Molecular plant-microbe interactions : MPMI* 25 (12) (2012) 1523–1530.
- [9] F. Supek, M. Bošnjak, N. Škunca, T. Šmuc, REVIGO summarizes and visualizes long lists of gene ontology terms, *PLoS One* 6 (7) (2011) e21800.