Data in Brief 54 (2024) 110309

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Localization data of the T-DNA insertion site in Arabidopsis line SALK_146824C

Elena Belykh[a,*], Kirill Yadrikhinskiy[b], Maksim Golubev[b], Nikolay Belykh[b], Ilya Velegzhaninov[a], Elena Garmash[a]

[a] Institute of Biology, Komi Scientific Center, Ural Division RAS, Kommunisticheskaya 28, 167982 Syktyvkar, Russia
[b] Pitirim Sorokin Syktyvkar State University, Oktyabrsky prosp., 55, Syktyvkar, Russia

ARTICLE INFO

Article history:
Received 21 November 2023
Revised 14 February 2024
Accepted 5 March 2024
Available online 10 March 2024

Dataset link: For Data on localization of the T-DNA insertion site in Arabidopsis line SALK_146824C (Original data)

Keywords:
Arabidopsis thaliana
T-DNA
Mutant line
VTC2

ABSTRACT

The present work reports on the determination of localization of T-DNA insertion in *Arabidopsis thaliana* SALK_146824C line. The line is characterized as deficient in GDP-l-galactose phosphorylase 1 (*VTC2*) gene encodes a protein involved in ascorbate biosynthesis. Primer pairs allow to distinguishing SALK_146824C from wild type plants and to the exact localization of the insertion were designed used. Sanger sequencing confirmed the location of the T-DNA insertion in the sixth exon of the gene. RT-PCR data shows nearly undetectable levels of *VTC2* mRNA expression level in SALK_146824C line as compared with wild type plants of Columbia-0 line. Localization of the primer pair upstream and downstream the T-DNA insertion did not affect the expression values.

* Corresponding author.
E-mail address: belykh@ib.komisc.ru (E. Belykh).

https://doi.org/10.1016/j.dib.2024.110309

Specifications Table

Subject	Biology
Specific subject area	Plant biology
Data format	Raw, Analyzed
Type of data	Table, Figure
Data collection	End-point PCR was performed to determine the presence and localization of T-DNA insertion, Sanger sequencing was used to confirm the exact localization of the T-DNA insertion, RT-PCR was employed to examine VTC2 expression level.
Data source location	Institute of Biology of Komi Scientific Centre, Ural Branch of Russian Academy of Sciences, Russia
Data accessibility	Belykh, Elena; Garmash, Elena; Velegzhaninov, Ilya; Yadrikhinskiy, Kirill; Belykh, Nikolay; Golubev, Maksim (2023), "For Data on localization of the T-DNA insertion site in Arabidopsis line SALK_146824C", Mendeley Data, V1, doi: 10.17632/fkmd7b7h8z.1 https://data.mendeley.com/datasets/fkmd7b7h8z/1

1. Value of the Data

- A number of Arabidopsis mutant lines with T-DNA insertions are used for plant genetic studies. Data on exact localization of T-DNA insertion makes possible the correct distinguishing mutant plants from wild type plant with use of PCR. Also, it makes easier RT-PCR primers design.
- Complex use of procedure including PCR and Sanger sequencing allows to localize the T-DNA insertion with the accuracy suitable in case of point-to-point genome editing tasks and has low costs.
- RT-PCR results show nearly undetectable levels of *VTC2* mRNA expression in SALK_146824C mutant compared with wild-type plants. The expression level shows no dependency from primer pair localization (upstream and downstream the insertion) and from the light exposure.

2. Data Description

A substantial number of Arabidopsis mutant lines, including SALK, SAIL, GABI-KAT and some other collections [1], are obtained with the use of T-DNA insertion. SALK lines [2] is a collection of mutants of *Arabidopsis thaliana* Col-0 ecotype, obtained through *Agrobacterium tumefaciens* induced insertion of transfer-DNA (pROK 2- T-DNA transformation vector, containing NPTII kanamycin-resistance selectable marker gene, driven by the CaMV 35S promoter) [1,3]. The exact location of T-DNA insertion is not known for all of lines and some T-DNA lines contain more than one insertion into the genome [4]. This lack of data complicates the identification of mutant plants, point-to-point genome editing tasks and the design of PCR primers, especially in loss-of-function mutants where the effectiveness of the mutagenesis depends on T-DNA insertion localization [5]. Furthermore, the estimation of the expression level of a knockout gene with the RT-PCR may depend on the localization of the expected PCR fragments [6].

The aim of our study was to determine the precise location of the T-DNA insertion within the AT4G26850 gene in Arabidopsis SALK_146824C line. AT4G26850 gene, which encodes the enzyme GDP-l-galactose phosphorilase 1 gene (GGP), was the only gene found to be homozygous and polymorphized in the studied line [7]. GGP catalyzes the first committed step in the l-galactose synthesis pathway of ascorbate (AsA), or Vitamin C, an important multifunctional antioxidant compound involved in stress tolerance [8]. In *A. thaliana*, GGP is encoded by the paralogous genes VITAMIN C2 (*VTC2*) and *VTC5*, but their expression levels and tissues specificity are slightly different, with *VTC2* playing a more important role in AsA biosynthesis [9]. Knockout of *VTC2* in EMS-induced ascorbate-deficient mutants caused a dramatic decrease in AsA levels and plant growth [10].

Table 1
A list of primer sets.

Name	Sequence (5′->3′)	Name	Sequence (5′->3′)
End-point PCR primers		RT-PCR primers	
V1_f V1_r	GCCTAGCCCCTGAACAAACA TGTCGTCTGTCTTCTCTGGT	RT_PCR_1f RT_PCR_1r	TGATAGTGGCAAAAGCACCCA TGACTGCTTGCCTCAAAGGA
V2_f V2_r	TGGGCATCTTCACCAGCTTC GCCTTAACGGTACGTTTTGGAA	RT_PCR_2f RT_PCR_2r	GGACCTCCATGGGTTCTGTT GAGGTTACTGCTCTCGCCTT
V3_f V3_r	GCCTGTTTCCAAAAACAATCACC TGCTCAGCTTAACGAGGGTC	*AT2G28390*	AACTCTATGCAGCATTTGATCCACT TGATTGCATATCTTTATCGCCATC
V4_f V4_r	ACTGAAGGACAAGGCACTCG ACCTTTTCACCCTTGCTCATCT	*AT4G34270*	GTGAAAACTGTTGGAGAGAAG CAATCAACTGGATACCCTTTCGCA
V5_f V5_r	TCATAACCAGAGCTTGTGGTGG GAGGTTACTGCTCTCGCCTT		
T-DNA_r	CTTCCTTTCTCGCCACGTTC		

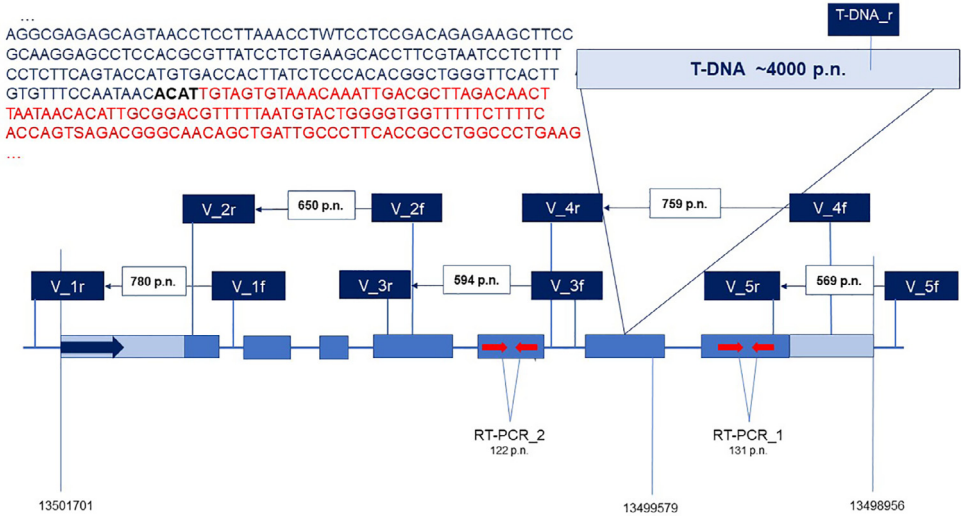

Fig. 1. The mapping scheme of the *VTC2* gene and the localization of the primer pairs. The sequences of the primers used are provided in the Table 1., and the length of the expected PCR product is indicated between the correspondent primer names for the end-point PCR and below the primer names for RT-PCR. The result of the amplicon sequencing corresponds to the *VTC2* gene (in blue letters) and the T-DNA insertion (in red letters).

The exact localization of the T-DNA insertion in the SALK_146824C Arabidopsis mutant is not available or is contradictory [3]. To address this, five primer pairs V1-V5 (Table 1, Figs. 1 and 2) were used to roughly localize the T-DNA insertion. However, the expected size products were not found for the V4 pair in case of the SALK_146824C mutant. Subsequently, the V4_f primer, together with the T-DNA_r primer located on the T-DNA insertion, was used to check the insertion position. This resulted in the generation of a product approximately 700 p.n. The obtained amplicon was cleaned up and sequenced. The alignment of the consensus sequence showed its total identity to the AT4G26850 GGP gene and to the pROK2 plasmid (Fig. 1). This corresponds to the T-DNA insertion localization at the position 13,499,579 in the 6th exon of the AT4G26850 gene.

The use of RT-PCR1 primer pair, located upstream the insertion, showed a very low expression level of *VTC2* mRNA in SALK_146824C mutant compared with the Col-0 plants (Fig. 3). High light exposure resulted in a three-time increase in the expression level of *VTC2* two hours after in wild-type plants, followed by a gradual reduction to the baseline of the *VTC2* relative tran-

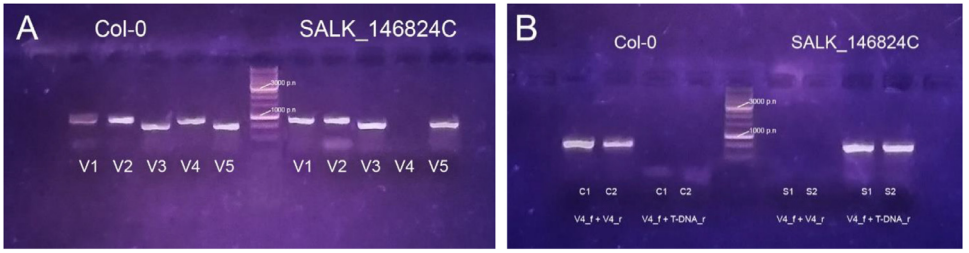

Fig. 2. The results of the end-point PCR showed the absence of PCR products for V4 fragment in the SALK_146824C line compared to the Col-0 plants (A). Conversely, the presence of a PCR product for the V4 primer pair and the absence of a PCR product for the T-DNA insertion were observed in the Col-0 plants instead of the SALK_146824C line (B).

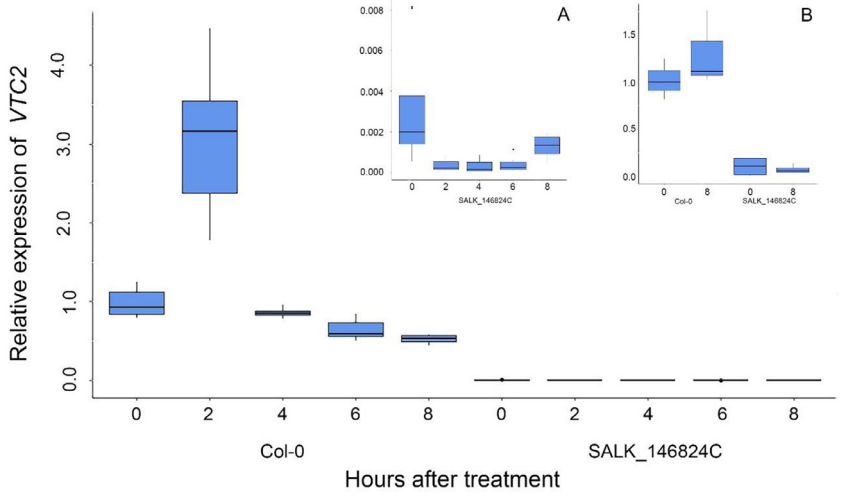

Fig. 3. The Assessment of the *VTC2* relative expression level in Arabidopsis lines Col-0 and SALK_146824C grown under 90 µmol m^{-2} s^{-1} (0 h) and after exposure to high light (400 µmol m^{-2} s^{-1}) using the RT_PCR_1 primer pair. The results, presented in the upper right corner, show A – the scaled relative transcript level of *VTC2* in the leaves of SALK_146824C using the RT_PCR_1 primer pair, and B – the relative transcript level of *VTC2* in the leaves of Col-0 and SALK_146824C using the RT_PCR_2 primer pair.

script level in leaves of Col-0 line. At the same time, no changes were noted in the mRNA level in leaves of SALK_146824C line (Fig. 3A). A similar trend was observed for the RT-PCR2 primer pairs located downstream of the T-DNA insertion, where expression level of *VTC2* mRNA in the leaves of SALK_146824C was significantly lower than in the case of wild-type plants.

3. Experimental Design, Materials and Methods

The *Arabidopsis thaliana* SALK_146824C T-DNA insertion line was obtained from Arabidopsis Biological Resource Center [7], while Colambia-0 (WT) was acquired from the Nottingham Arabidopsis Stock Centre (NASC, Nottingham University, UK). Seeds of both lines were sown in a pre-wetted substrate (perlite:vermiculite:soil (1:1:2)), stratified at 4 °C in the dark for 96 h and then maintained under controlled conditions of a 22 °C temperature, 10-hour day length (from 7:00 to 17:00), and a photosynthetic active photon flux density of 90 µmol photons m^{-2} s^{-1}. The four-week-old Arabidopsis plants at growth stage 1.14; [11] were subsequently exposed to high light (400 µmol m^{-2} s^{-1}) for 8 h.

The total DNA and RNA isolation from rosette leaves (growth stage 1.14; [11]) involved several steps. Genomic DNA was extracted ~~with~~ using the Sorb–GMO–B kit (Sintol, Russia) according to the manufacturer's instructions. PCR reactions were carried out with Screen Mix (Evrogen, Russia) on a T-100 thermal cycler (Bio-Rad, USA). The resulting PCR products were then separated on a 1% agarose gel, and the amplicons were purified using the Cleanup Standard kit (Evrogen, Russia) as per the supplier's instructions. Primer pairs for both end-point PCR and real-time PCR were designed using the Primer-BLAST online tool [12] (Table 1). Sanger sequencing was performed with the GenSeq kit (Sintol, Russia) and the genetic analyzer NANOPHOR 05 (Sintol, Russia). The nucleotide sequences of the GDP-l-galactose phosphorylase 1 gene and the pROK2 plasmid sequence were obtained from the NCBI and TAIR databases correspondently. Multiple alignments of nucleotide sequences were conducted using ClustalO in the UGENE program [13].

Total RNA was isolated using an Aurum Total RNA Mini Kit (Bio-Rad), reverse-transcription applied using oligo-dT primers and an MMLV RT Kit (Evrogen, Russia). qRT-PCR assays were performed using qPCRmix-HS SYBR (Evrogen, Russia) on CFX96 thermal cycler (Bio-Rad, USA); detailed protocol of RT-PCR could be found in [14]. AT2G28390 and AT4G34270 [15] were used for target gene expression normalization. Relative expression values were obtained using the $2^{-\Delta\Delta CT}$ method [16]. Effectivity of RT_PCR_1 and RT_PCR_2 primer pairs were 2.07 and 2.03 correspondently.

Limitations

None.

Ethics Statement

The authors have read and follow the ethical requirements for publication in Data in Brief and confirm that the current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

Data Availability

For Data on localization of the T-DNA insertion site in Arabidopsis line SALK_146824C (Original data) (Mendeley Data).

CRediT Author Statement

Elena Belykh: Investigation, Writing – review & editing, Visualization; **Kirill Yadrikhinskiy:** Investigation; **Maksim Golubev:** Software; **Nikolay Belykh:** Software; **Ilya Velegzhaninov:** Methodology, Writing – review & editing; **Elena Garmash:** Writing – review & editing, Supervision.

Acknowledgements

The study was performed using the equipment of the Collective Usage Center (CUC) "Molecular Biology", Institute of Biology, Komi Science Centre, Ural Branch RAS. This research was partly funded by the Russian Science Foundation (RSF: grant number 22–24–01082) and within the Ministry of Science and Higher Education projects of Institute of Biology of Komi Science Center of the Ural Branch of the Russian Academy of Science 122040600024–5 and 122040600021–4.

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.

References

[1] R.C. O'Malley, C.C. Barragan, J.R. Ecker, A user's guide to the Arabidopsis T-DNA insertional mutant collections, Methods Mol. Biol. 1284 (2015) 323–342, doi:10.1007/978-1-4939-2444-8_16.
[2] The Salk Institute Genomic Analysis Laboratory http://signal.salk.edu/.
[3] J.M. Alonso, A.N. Stepanova, T.J. Leisse, C.J. Kim, H. Chen, P. Shinn, D.K. Stevenson, J. Zimmerman, P. Barajas, R. Cheuk, C. Gadrinab, C. Heller, A. Jeske, E. Koesema, C.C. Meyers, H. Parker, L. Prednis, Y. Ansari, N. Choy, H. Deen, M. Geralt, N. Hazari, E. Hom, M. Karnes, C. Mulholland, R. Ndubaku, I. Schmidt, P. Guzman, L. Aguilar-Henonin, M. Schmid, D. Weigel, D.E. Carter, T. Marchand, E. Risseeuw, D. Brogden, A. Zeko, W.L. Crosby, C.C. Berry, J.R. Ecker, Genome-wide insertional mutagenesis of Arabidopsis thaliana, Science 301 (2003) 653–657.
[4] B. Pucker, N. Kleinbölting, B. Weisshaar, Large scale genomic rearrangements in selected Arabidopsis thaliana T-DNA lines are caused by T-DNA insertion mutagenesis, BMC Genomics 22 (2021) 599, doi:10.1186/s12864-021-07877-8.
[5] Y.H. Wang, How effective is T-DNA insertional mutagenesis in Arabidopsis? J. Biochem Tech. 1 (2008) 11–20.
[6] M.C. Rodríguez, A. Wawrzynska, A. Sirko, Intronic T-DNA insertion in Arabidopsis NBR1 conditionally affects wild–type transcript level, Plant Signal Behav. 9 (2014) e975659.
[7] Arabidopsis Biological Resource Center ABRC (https://abrc.osu.edu/).
[8] C.H. Foyer, T. Kyndt, R.D. Hancock, Vitamin C in Plants: novel concepts, new perspectives, and outstanding Issues, Antioxid. Redox Signal 32 (2020) 463–485, doi:10.1089/ars.2019.7819.
[9] J. Dowdle, T. Ishikawa, S. Gatzek, S. Rolinski, N. Smirnoff, Two genes in Arabidopsis thaliana encoding GDP-L-galactose phosphorylase are required for ascorbate biosynthesis and seedling viability, Plant J. 52 (2007) 673–689.
[10] B. Lim, N. Smirnoff, C.S. Cobbett, J.F, Golz Ascorbate-Deficient vtc2 mutants in Arabidopsis do not exhibit decreased growth, Front. Plant Sci. 7 (2016) 1025, doi:10.3389/fpls.2016.01025.
[11] D.C. Boyes, A.M. Zayed, R. Ascenzi, A.J. McCaskill, N.E. Hoffman, K.R. Davis, J. Gorlach, Growth, stage-based phenotypic analysis of Arabidopsis: a model for high throughput functional genomics in plants, Plant Cell 13 (2001) 1499–1510.
[12] J. Ye, G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen, Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction, BMC Bioinformatics 13 (2012) 134, doi:10.1186/1471-2105-13-134.
[13] K. Okonechnikov, O. Golosova, M. Fursov, Unipro UGENE: a unified bioinformatics toolkit, Bioinformatics 28 (2012) 1166–1167, doi:10.1093/bioinformatics/bts091.
[14] E.V. Garmash, O.V. Dymova, E.V. Silina, R.V. Malyshev, E.S. Belykh, M.A. Shelyakin, I.O. Velegzhaninov, *AOX1a* Expression in *Arabidopsis thaliana* affects the state of chloroplast photoprotective systems under moderately high light conditions, Plants 11 (2022) 3030, doi:10.3390/plants11223030.
[15] T. Czechowski, M. Stitt, T. Altmann, M.K. Udvardi, W.R. Scheible, Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis, Plant Physiol. 139 (2005) 5–17.
[16] K.J. Livak, T.D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method, Methods 25 (2001) 402–408.