Data in Brief 48 (2023) 109047

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Arabidopsis transcriptome dataset of the response of imbibed wild-type and glucosinolate-deficient seeds to nitrogen-containing compounds

Mailen Ortega-Cuadros[a,b,1], Sophie Aligon[a,1], Nubia Velasquez[a], Jerome Verdier[a], Philippe Grappin[a,*]

[a] *Institut Agro, University Angers, INRAE, IRHS, SFR 4207 QuaSaV, Angers F-49000, France*
[b] *Institute of Biology, University of Antioquia, Calle 67 N° 53-108, Medellín 050010, Colombia*

ARTICLE INFO

Article history:
Received 28 December 2022
Revised 24 February 2023
Accepted 2 March 2023
Available online 13 March 2023

Dataset link: Arabidopsis transcriptome dataset of the response of imbibed wild-type and glucosinolate-deficient seeds to nitrogen-containing compounds (Original data)

Keywords:
Seed imbibition
Glucosinolates
Potassium thiocyanate
Potassium nitrate
Arabidopsis
RNA-seq

ABSTRACT

The presented RNAseq data were obtained from *Arabidopsis* seeds dry and 6h imbibed to describe, in wild-type and glucosinolate (GSL)-deficient genotypes, the response at the RNA level to nitrogen compounds, *i.e.*, potassium nitrate (KNO_3, 10mM), potassium thiocyanate (KSCN, 8µM). The *cyp79B2 cyp79B3 (cyp79B2/B3)* double mutant deficient in Indole GSL, the *myb28 myb29 (myb28/29)* double mutant deficient in aliphatic GSL, the quadruple mutant *cyp79B2 cyp79B3 myb28 myb29 (qko)* deficient in total GSL in the seed and the WT reference genotype in Col-0 background were used for the transcriptomic analysis. Total ARN was extracted using NucleoSpin® RNA Plant and Fungi kit. Library construction and sequencing were performed with DNBseq™ technology at Beijing Genomics Institute. FastQC was used to check reads quality and mapping analysis were made using a quasi-mapping alignment from Salmon. Gene expression changes in mutant seeds compared to WT were calculated using DESeq2 algorithms. This comparison with the *qko, cyp79B2/B3* and *myb28/29* mutants made it possible to identify 30220, 36885 and 23807 differentially expressed genes

* Corresponding author.
E-mail address: philippe.grappin@agrocampus-ouest.fr (P. Grappin).
Social media: @IRHS_Seed_lab (J. Verdier)
[1] Authors have contributed equally to this work.

https://doi.org/10.1016/j.dib.2023.109047

(DEGs), respectively. Mapping rate result was merge into a single report using MultiQC; graphic results were illustrated through Veen diagrams and volcano plots. Fastq raw data and count files from 45 samples are available in the repository Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) and can be consulted with the data identification number GSE221567 at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE221567.

Specifications Table

Subject	Agricultural and Biological Sciences
Specific subject area	Omics: Transcriptomics Plant Science: Plant Physiology
Type of data	Tables Figures
How the data were acquired	Mutants and WT seeds were collected from *Arabidopsis* Col-0 genetic background. Briefly, RNA extraction was made from dry -and 6h imbibed seeds and RNA samples were sent to Beijing Genomics Institute (BGI, https://www.bgi.com). The library construction was made by BGI and paired-end reads were sequenced using the sequencing platform DNBseq™. Bioinformatic analyses were performed using Salmon (version 0.14.1) [1], FastQC [2] and DESeq2 [3] programs. Results were illustrated using MultiQC tool [4], the online Venn Diagram tool (http://bioinformatics.psb.ugent.be/webtools/Venn/) and volcano plots using ggplot2 (version 3.4.0) [5] package in R studio (version 4.2.0).
Data format	Raw, processed and, analyzed reads
Description of data collection	Glucosinolate mutants (*qko, cyp79B2/B3* and *myb28/29*) and WT seeds were used in post-harvested dry state or in 6h imbibed conditions containing water or KNO_3 10mM or KSCN 8µM. Incubation was done in a controlled room chamber under 16 h photoperiod at 22°C/20°C (day/night) and 70% relative humidity. Samples were collected for RNA extraction. Sequencing was performed by BGI, and 217 Gb of data were generated corresponding to clean raw sequences. After quality control with FastQC [2] and MultiQC tools [4], high quality reads were mapped to reference transcriptome from *Arabidopsis* Araport 11 [6] using the Salmon algorithm [1]. The output allowed to calculated the respective quantification and gene expression changes at transcriptome level (*i.e.*, count file).
Data source location	Institution: Growth chambers located at the Institut de Recherche en Horticulture et Semences City: Beaucouzé Country: France GPS coordinates: 47°28′37.7"N 0°36′42.1"W
Data accessibility	Repository name: NCBI GEO [7] Data identification number: GSE221567 Direct URL to data: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE221567 Repository name: Mendeley Data [8] Direct URL to data: https://data.mendeley.com/datasets/cggjbsmw7d/1

Value of the Data

- These data contain transcriptomic response of seed to nitrogen compounds using GSL deficient mutants. The dataset is providing new insights and allows understanding of the regulatory mechanisms related to the contribution of nitrogen metabolism in seed quality.

- These data are useful resource for seed biologist community. Likewise, researchers interested on plant breeding and genetics would identify interesting molecular markers.
- The scientific community can use these data to identify at the transcriptional level the signaling contribution of nitrogen compounds during seedling emergence. Also, researchers can identify key targets of GSL and nitrate for seed vigor and specifically the contribution of GSL metabolism during this process.

Objective

Nitrogen containing metabolites such as GSL [9] are well known as key player in plant growth and development [10]. The goal of this study was to investigate at the transcriptome level the contribution of GSL in stimulating seed germination. The used of a GSL mutant collection will help to clarify the possible role of indole -and aliphatic GSL at the first stage of the plant life cycle.

1. Data Description

The dataset describes the response of the seeds to KNO_3 and to KSCN at the transcriptional level using a collection of GSL mutants. We sequenced transcripts from 45 samples corresponding to *Arabidopsis* seeds in Col-0 genetic background: the wild type (WT) and three GSL mutant lines (*qko: cyp79B2 cyp79B3 myb28 myb29* quadruple mutant; *cyp79B2/B3: cyp79B2 cyp79B3* double mutant; *myb28/29: myb28 myb29* double mutant). Four experimental conditions were analysed with dry and 6h imbibed seeds and with KNO_3 and KSCN treatments in three biological replicates. All fastq raw data files (and count) are available in the NCBI Sequence Read Archive (SRA) database under the repository name NCBI GEO with the data identification number GSE221567 (https: //www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE221567) [7]. The quality of raw reads was checked with FastQC [2]; subsequently, filtered reads were mapped to the reference transcriptome Araport11 from *Arabidopsis* [6] using Salmon algorithm [1]. The mapping summary can be found in Table 1 and the report with all general statistics is available at the Supplementary Table 1 [8].

Table 1
Summary of mapping rates obtained with MultiQC tool [4] following Salmon quasi-mapping algorithm [1]. Wild-type and GSL mutant *Arabidopsis* seeds (*qko, cyp79B2/B3* and *myb28/29*) in dry state, or 6h imbibed state in water or in KNO_3 10mM or KSCN 8µM. REP: biological replicate. % Aligned: % Mapped reads; M Aligned: Mapped reads (millions); M Seqs: Total Sequences (millions).

Sample Name	% Aligned	M Aligned	M Seqs
Col_dry_REP1	92.4%	22.2	24.1
Col_dry_REP2	91.5%	22.0	24.1
Col_dry_REP3	91.2%	22.0	24.2
Col_water_REP1	93.1%	22.4	24.1
Col_water_REP2	94.8%	22.8	24.0
Col_water_REP3	93.6%	22.6	24.1
Col_KNO3_REP1	92.9%	22.3	24.0
Col_KNO3_REP2	92.3%	22.3	24.1
Col_KNO3_REP3	94.5%	22.7	24.0
Col_KSCN_REP1	93.3%	22.5	24.1
Col_KSCN_REP2	93.5%	22.5	24.0
Col_KSCN_REP3	93.7%	22.6	24.2
qko_dry_REP1	91.8%	22.1	24.1
qko_dry_REP2	94.4%	22.7	24.0
qko_dry_REP3	93.7%	22.5	24.0
qko_water_REP1	92.8%	22.4	24.1

(*continued on next page*)

Table 1 (*continued*)

Sample Name	% Aligned	M Aligned	M Seqs
qko_water_REP2	92.6%	22.3	24.1
qko_water_REP3	90.7%	21.4	23.6
qko_KNO3_REP1	92.9%	22.3	24.0
qko_KNO3_REP2	92.7%	22.3	24.1
qko_KNO3_REP3	87.7%	18.1	20.6
qko_KSCN_REP1	92.4%	22.2	24.1
qko_KSCN_REP2	92.6%	22.3	24.1
qko_KSCN_REP3	90.6%	21.9	24.1
cyp_dry_REP1	93.3%	22.5	24.1
cyp_dry_REP2	92.1%	22.2	24.1
cyp_dry_REP3	92.1%	22.1	24.0
cyp_water_REP1	94.5%	22.7	24.0
cyp_water_REP2	93.8%	22.5	24.0
cyp_water_REP3	95.4%	22.9	24.0
cyp_KNO3_REP1	94.0%	22.6	24.0
cyp_KNO3_REP2	93.9%	22.6	24.1
cyp_KNO3_REP3	94.6%	22.7	24.0
cyp_KSCN_REP1	93.3%	22.4	24.0
cyp_KSCN_REP2	94.3%	22.7	24.0
myb_dry_REP1	93.6%	22.5	24.0
myb_dry_REP2	93.2%	22.5	24.1
myb_water_REP1	93.2%	22.4	24.1
myb_water_REP2	93.7%	22.6	24.1
myb_water_REP3	93.1%	22.4	24.0
myb_KNO3_REP1	93.3%	22.4	24.0
myb_KNO3_REP2	92.6%	22.2	24.0
myb_KSCN_REP1	92.1%	22.1	24.0
myb_KSCN_REP2	92.6%	22.3	24.1
myb_KSCN_REP3	92.2%	22.2	24.0

Table 2
Numbers of DEGs ($\log_2$ FC > 1 or < -1 and Benjamini- Hochberg adjusted p-value < 0.05) identified from the comparison between the respective mutant lines (*qko; cyp79B2/B3; myb28/29*) and the WT seed samples. Conditions: dry and 6h-imbibed seed in water; KNO_3 10mM; KSCN: KSCN 8µM. (n): Quantity of genes associated to each expression change.

	Summary				
Mutant lines	Expression change	Water (n)	KNO_3 (n)	KSCN (n)	Dry (n)
qko	Down-regulated	3877	3131	4145	4569
	Non-significant	20120	21654	19639	18714
	Up-regulated	3590	2802	3803	4303
cyp79B2/B3	Down-regulated	5584	4322	5187	2859
	Non-significant	16381	18766	17082	21233
	Up-regulated	5622	4499	5318	3494
myb28/29	Down-regulated	3162	2087	3799	3278
	Non-significant	21392	23681	20381	21086
	Up-regulated	3033	1819	3407	3222

From dry or imbibed seeds in water, KNO_3 or KSCN conditions, genes expression changes in mutants compared to WT were identified. All analyses were obtained according to DEseq2 statistical parameters [3]. From this analysis, the number of differentially expressed genes (DEGs) were described in Table 2 and quantification of expression changes were graphically represented using volcano plots (Fig. 1) [5].

Common DEGs from the GSL mutants are represented by Venn diagrams (Fig. 2).

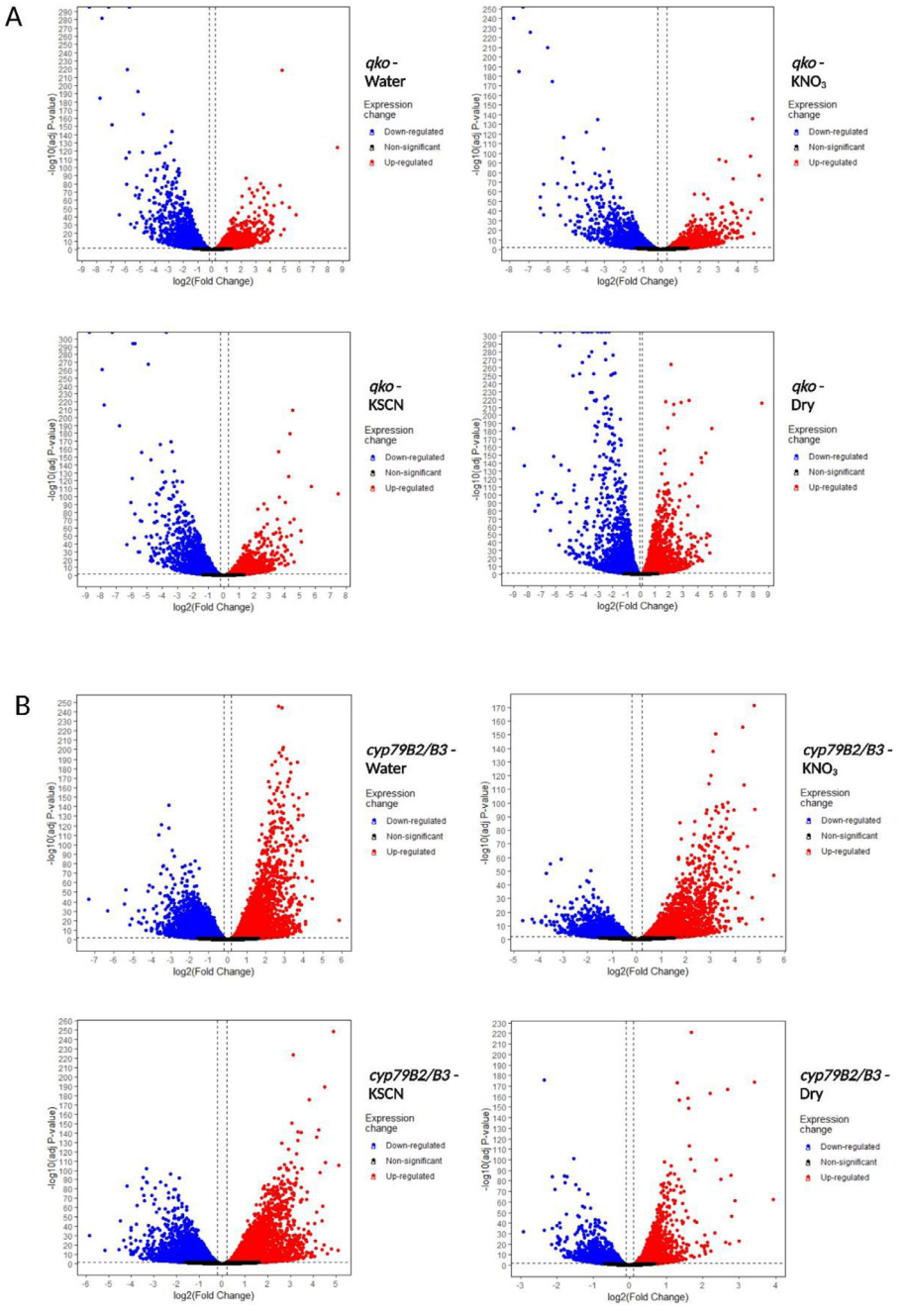

Fig. 1. Volcano plots quantifying differential expression for identified DEGs in *qko* (A), *cyp79B2/B3* (B) and *myb28/29* (C) mutants compared to WT. Differential expressions were shown for dry seed and imbibed seeds in water, KNO_3 and KSCN conditions respectively. These results were obtained using DESeq2 algorithm [3]. The $\log_2$ Fold change (Log_2FC) corresponds to expression changes from comparison ratio between mutant versus WT genotypes. Colors scale: Blue represents down-regulated genes, red up-regulated genes and black not significantly expression change. Adjusted p-value threshold for significance $p < 0.05$.

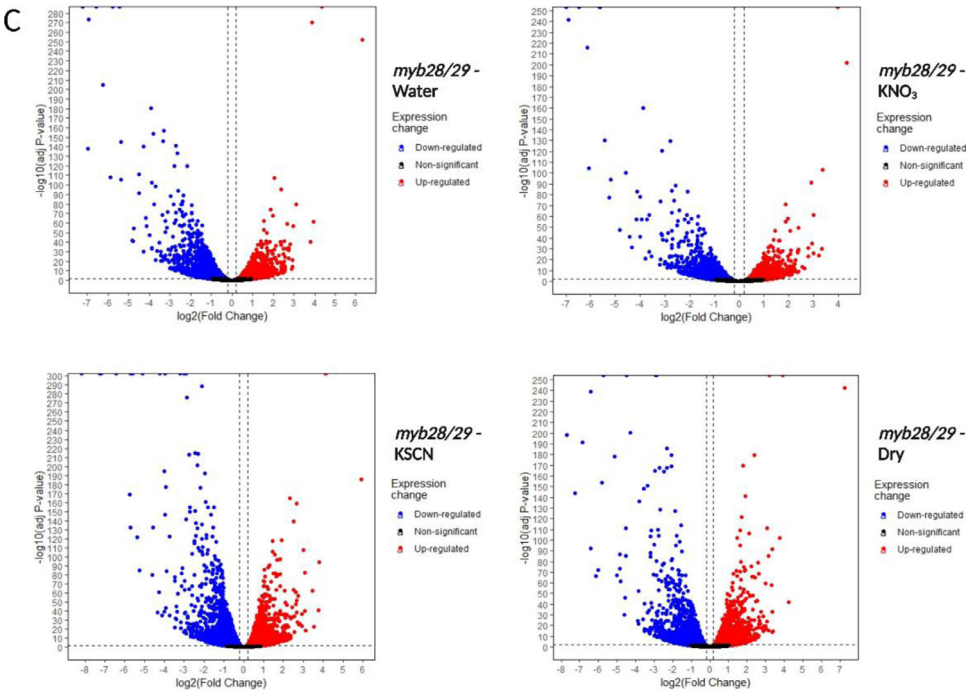

Fig. 1. Continued

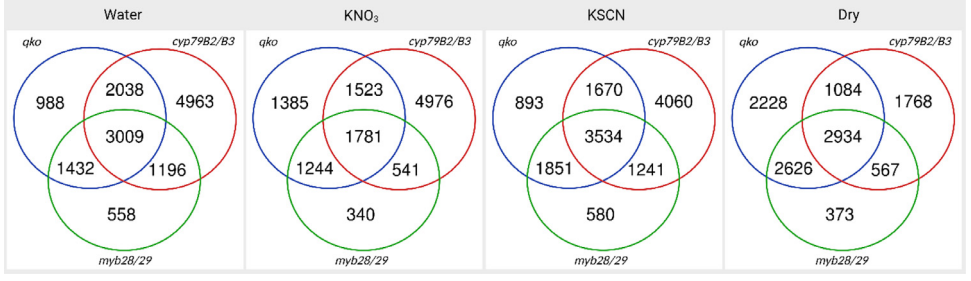

Fig. 2. Venn diagram representation of common DEGs found in GSL mutants compared to the WT reference in the different physiological conditions of the seed: imbibed seeds in water, KNO_3 (imbibed seeds in KNO_3 10 mM) or KSCN (imbibed seeds in KSCN 8µM) and dry seeds from post-harvested dry state. DEGs from WT compared to mutants, correspond to genes displaying a $\log_2$ FC > 1 or < -1 and Benjamini- Hochberg adjusted p-value < 0.05.

2. Experimental Design, Materials and Methods

This study used *Arabidopsis* seeds from Columbia (Col-0) genetic background. Seeds from WT, quadruple and double *knockout* mutants: *cyp79B2 cyp79B3 myb28 myb29* (*qko*) [11], *cyp79B2 cyp79B3* (*cyp79B2/B3*) [12], and *myb28 myb29* (*myb28/29*) [13] were surface-sterilized and dried according to Ortega-Cuadros et al. [14]. A part of seeds was collected at the initial dry state, the other part was imbibed in water, KNO_3 10mM or KSCN 8µM conditions respectively during 6 hours into a controlled incubator under 16 h photoperiod (170 µmol photons m^2 s^{-1}) at 22°C (light period)/20°C (dark period) and a constant 70% relative humidity [15].

Total RNA was extracted from 20 mg of seed tissue using NucleoSpin® RNA Plant and Fungi kit (Macherey – Nagel, Düren, Germany) according to the manufacturer's instructions. RNA quantification was obtained using a NanoDrop ND-100 (NanoDrop Technologies, DE, USA) and the RNA Integrity was measured using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). cDNA library constructions were performed and paired-end (PE100) reads were sequenced using the Sequencing Platform DNBseq^{TM} [15] by Beijing Genomics Institute (BGI, https://www.bgi.com).

A total of 217 Gb of data corresponding to clean raw sequences (Phred+33) was obtained. After quality control [2], filtered reads were mapped to *Arabidopsis* transcriptome (Araport 11) [6] using quasi-mapping alignment from Salmon, version 0.14.1 [1]. Quality control files of these analyses were merged and summarized into a single report using MultiQC tool [4] (Supplementary Table 1). Gene expression changes from comparison between mutant versus wild-type seeds were calculated using the algorithm from DESeq2 [3], and identification of differentially expressed genes (DEGs) was based on $\log_2$FC and adjusted p-value thresholds. Genes with $\log_2$FC > 1 or < -1 and Benjamini- Hochberg adjusted p-values < 0.05 were used to determine DEGs. Genes outside these parameters were defined as non-significantly differentially expressed (Fig. 1 and Table 2). Venn diagram tool (http://bioinformatics.psb.ugent.be/webtools/Venn/) was used to identify common DEGs between mutant seeds at different physiological conditions (Fig. 2). Volcano plots were made using ggplot2 (version 3.4.0) [5] package in R studio (version 4.2.0). The graphical abstract was created with BioRender.com

Ethics Statements

This work does not contain any studies with human or animal subjects.

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

Arabidopsis transcriptome dataset of the response of imbibed wild-type and glucosinolate-deficient seeds to nitrogen-containing compounds (Original data) (Mendeley Data).

CRediT Author Statement

Mailen Ortega-Cuadros: Conceptualization, Methodology, Formal analysis, Writing – review & editing; **Sophie Aligon:** Visualization, Conceptualization, Methodology, Supervision; **Nubia Velasquez:** Conceptualization, Investigation, Methodology; **Jerome Verdier:** Data curation, Formal analysis, Writing – review & editing, Supervision; **Philippe Grappin:** Funding acquisition, Visualization, Conceptualization, Investigation, Writing – review & editing, Supervision.

Acknowledgments

This work has benefited from a grant from the French State managed by the National Research Agency under the Future Investments program bearing the reference ANR-20-PCPA-0009. We would like to thank the FUNGISEM team for their support during the development of the investigation.

References

[1] R. Patro, G. Duggal, M.I. Love, R.A. Irizarry, C. Kingsford, Salmon provides fast and bias-aware quantification of transcript expression, Nat. Methods 14 (2017) 417–419, doi:10.1038/nmeth.4197.

[2] Babraham Bioinformatics, FastQC A Quality Control tool for High Throughput Sequence Data, FastQC. (2010). https://www.bioinformatics.babraham.ac.uk/projects/fastqc/(accessed March 22, 2022).

[3] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014) 550, doi:10.1186/s13059-014-0550-8.

[4] P. Ewels, M. Magnusson, S. Lundin, M. Käller, MultiQC: summarize analysis results for multiple tools and samples in a single report, Bioinformatics 32 (2016) 3047–3048, doi:10.1093/bioinformatics/btw354.

[5] H. Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag, New York, 2016 https://ggplot2.tidyverse.org accessed December 10, 2022.

[6] C. Cheng, V. Krishnakumar, A.P. Chan, F. Thibaud-Nissen, S. Schobel, C.D. Town, Araport11: a complete reannotation of the Arabidopsis thaliana reference genome, The Plant Journal 89 (2017) 789–804, doi:10.1111/tpj.13415.

[7] P. Grappin, M. Ortega-Cuadros, S. Aligon, J. Verdier, Effect of nitrogen compounds on imbibed mutant seeds impaired in glucosinolate biosynthesis. NCBI Gene Expression Omnibus, accession number GSE221567. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE221567

[8] M. Ortega-Cuadros, S. Aligon, N. Velasquez, J. Verdier, P. Grappin, Arabidopsis transcriptome dataset of the response of imbibed wild-type and glucosinolate-deficient seeds to nitrogen-containing compound, Mendeley Data (2023) V1, doi:10.17632/cggjbsmw7d.1.

[9] B.A. Halkier, J. Gershenzon, Biology and biochemistry of glucosinolates, Annu. Rev. Plant Biol. 57 (2006) 303–333, doi:10.1146/annurev.arplant.57.032905.105228.

[10] S.J. Leghari, N.A. Wahocho, G.M. Laghari, A. HafeezLaghari, G. MustafaBhabhan, K. HussainTalpur, T.A. Bhutto, S.A. Wahocho, A.A. Lashari, Role of Nitrogen for Plant Growth and Development: A review, Advances InEnvironmental Biology 10 (9) (2016) 209–218.

[11] J.Y. Sun, I.E. Sønderby, B.A. Halkier, G. Jander, M. de Vos, Non-volatile intact indole glucosinolates are host recognition cues for ovipositing plutella xylostella, J. Chem. Ecol. 35 (2009) 1427–1436, doi:10.1007/s10886-009-9723-4.

[12] Y. Zhao, A.K. Hull, N.R. Gupta, K.A. Goss, J. Alonso, J.R. Ecker, J. Normanly, J. Chory, J.L. Celenza, Trp-dependent auxin biosynthesis in arabidopsis: involvement of cytochrome P450s CYP79B2 and CYP79B3, Genes Dev. 16 (2002) 3100–3112, doi:10.1101/gad.1035402.

[13] I.E. Sønderby, B.G. Hansen, N. Bjarnholt, C. Ticconi, B.A. Halkier, D.J. Kliebenstein, A systems biology approach identifies a R2R3 MYB gene subfamily with distinct and overlapping functions in regulation of aliphatic glucosinolates, PLoS One 2 (2007) e1322, doi:10.1371/journal.pone.0001322.

[14] M. Ortega-Cuadros, T.L. de Souza, R. Berruyer, S. Aligon, S. Pelletier, J.-P. Renou, T. Arias, C. Campion, T. Guillemette, J. Verdier, P. Grappin, Seed transmission of pathogens: non-canonical immune response in arabidopsis germinating seeds compared to early seedlings against the necrotrophic fungus alternaria brassicicola, Plants 11 (2022) 1708, doi:10.3390/plants11131708.

[15] M. Ortega-Cuadros, L. Chir, S. Aligon, T. Arias, J. Verdier, P. Grappin, Dual-transcriptomic datasets evaluating the effect of the necrotrophic fungus Alternaria brassicicola on Arabidopsis germinating seeds, Data Brief 44 (2022) 108530, doi:10.1016/j.dib.2022.108530.