



ELSEVIER

Contents lists available at ScienceDirect

## Data in brief

journal homepage: [www.elsevier.com/locate/dib](http://www.elsevier.com/locate/dib)

## Data Article

# Transcriptomic and cell wall proteomic datasets of rosettes and floral stems from five *Arabidopsis thaliana* ecotypes grown at optimal or sub-optimal temperature



Harold Duruflé <sup>a</sup>, Philippe Ranocha <sup>a</sup>, Thierry Balliau <sup>b</sup>,  
Christophe Dunand <sup>a</sup>, Elisabeth Jamet <sup>a,\*</sup>

<sup>a</sup> Laboratoire de Recherche en Sciences Végétales, Université de Toulouse, CNRS, UPS, 24 chemin de Borde Rouge, Auzeville, BP42617, 31326 Castanet Tolosan, France

<sup>b</sup> PAPPSo, GQE - Le Moulon, INRA, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, 91190 Gif-sur-Yvette, France

## ARTICLE INFO

## Article history:

Received 17 July 2019

Received in revised form 11 September 2019

Accepted 23 September 2019

Available online 28 September 2019

## Keywords:

*Arabidopsis thaliana*

Cell wall

Proteomics

Pyrenees ecotype

Transcriptomics

## ABSTRACT

This article provides experimental data describing the RNA and the cell wall protein profiles of rosettes and flower stems of five *Arabidopsis thaliana* ecotypes. Four newly-described Pyrenees ecotypes [1] are analyzed in addition to the well-described and sequenced Columbia (Col) ecotype of *A. thaliana*. All five ecotypes have been grown at two different temperatures, 22 °C and 15 °C. We provide transcriptomics and cell wall proteomics data regarding (i) rosettes at the bolting stage, and (ii) floral stems at the first flower stage. These data are a valuable resource to study the adaptation of *A. thaliana* ecotypes to sub-optimal temperature growth conditions.

© 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

\* Corresponding author.

E-mail addresses: [dunand@lrsv.ups-tlse.fr](mailto:dunand@lrsv.ups-tlse.fr) (C. Dunand), [jamet@lrsv.ups-tlse.fr](mailto:jamet@lrsv.ups-tlse.fr) (E. Jamet).

## Specifications Table

Subject area	Biology
More specific subject area	Transcriptomics, Cell wall proteomics
Type of data	Raw and processed data
How data was acquired	Illumina HiSeq 3000, LC-MS/MS
Data format	Tables
Experimental factors	5 ecotypes of <i>Arabidopsis thaliana</i> at 2 growth temperatures (15 and 22 °C) 2 organs (rosettes and floral stems) 3 biological replicates
Experimental features	Transcriptomics and cell wall proteomics of rosettes and floral stems
Data source location	Laboratoire de Recherche en Sciences Végétales, Université de Toulouse
Data accessibility	Resulting sequences are available at NCBI short read archive (SRA, BioProject PRJNA344545). LC-MS/MS data are deposited at PROTIcDb ( <a href="http://moulon.inra.fr/protic/walloomics">http://moulon.inra.fr/protic/walloomics</a> ) Cell wall proteomics data are available in this article as 2 <a href="#">Supplementary files</a> .

**Value of the Data**

- The data provide a first level of molecular description at the transcriptomics and the cell wall proteomics levels of four newly-identified *A. thaliana* altitudinal ecotypes from the Pyrenees mountains.
- The datasets include RNAseq and cell wall proteomics quantitative analyses of two organs (rosettes and stems) of five *A. thaliana* ecotypes at two growth temperatures.
- The data can be used to study the impact of sub-optimal temperature growth conditions on *A. thaliana* ecotypes responses.

**1. Data**

The data provided here two different omics approaches to contribute to the understanding of the impact of sub-optimal temperature growth conditions on *A. thaliana* responses. The raw data include information about differential gene expression and cell wall protein abundance in five wild *A. thaliana* ecotypes. All the LC-MS/MS proteomics data have been deposited at PROTIcDb (<http://moulon.inra.fr/protic/walloomics>). The RNAseq data are available at NCBI short read archive (<https://www.ncbi.nlm.nih.gov/sra>, BioProject PRJNA344545).

The transcriptomic and proteomic quantification datasets in rosettes and floral stems are provided in two Supplementary files including:

- [Supplementary file 1](#): Quantitative cell wall proteomics data: proteins were extracted from purified cell walls of *A. thaliana* rosettes or floral stems after growth at 22 °C or 15 °C by saline solutions (CaCl<sub>2</sub> 0.2 M and LiCl 2 M) and identified by LC-MS/MS and bioinformatics
- [Supplementary file 2](#): Transcript levels (RNAseq data): RNA were extracted from *A. thaliana* rosettes or floral stems after growth at 22 °C or 15 °C

**2. Experimental design, materials, and methods****2.1. Plant material**

The experimental design and plant materials have been previously detailed [1]. In summary, five altitudinal ecotypes of *A. thaliana* were used: Grip, Hosp, Hern, Roch collected at different altitudes in the Pyrenees mountains [2], and the well-described Columbia ecotype (Col) ([https://www.ebi.ac.uk/ols/ontologies/efo/terms?short\\_form=EFO\\_0005147](https://www.ebi.ac.uk/ols/ontologies/efo/terms?short_form=EFO_0005147)). After being sowed in Jiffy-7® peat pellets (Jiffy International, Kristiansand, Norway), seeds were stratified for 48 h at 4 °C in darkness. Plants were grown at two different temperatures (22 °C or 15 °C), under 90 μmol photons.m<sup>-2</sup>.s<sup>-1</sup> of light intensity under a long-day condition (16 h light/8 h dark) with 70% humidity. In total, about 20 plants from three independent batches have been used and pooled for the triplicate molecular

analyses. The rosettes and floral stems were collected just before bolting (stage 5.10 [3]) and at the first flower stage of development (stage 6 [3]), respectively. The same biological triplicates were used to prepare samples for both transcriptomics and cell wall proteomics analyses.

## 2.2. Sequential extraction of proteins from purified cell wall

Cell wall purification was performed as described [4]. The sequential extraction of proteins from purified cell walls was done as described [5]. The final protein extract was lyophilized. Proteins were quantified with the CooAssay Protein Assay kit (Interchim, Montluçon, France). Typically, 0.2 g of lyophilized cell walls was used for one extraction and about 500 µg proteins were obtained.

## 2.3. Identification of proteins by LC-MS/MS

The identification of proteins extracted from cell walls was performed by LC-MS/MS at the PAPPSo proteomics platform ([pappso.inra.fr/](http://pappso.inra.fr/)) after tryptic digestion in solution as described [6]. Parameters for MS data processing in the X!Tandem software (JACKHAMMER, 2013.6.15, [www.thegpm.org/tandem/](http://www.thegpm.org/tandem/)) and the X!Tandem Pipeline 3.3.4 [7] are detailed in Ref. [8]. Trypsin digestion was declared with no possible miscleavage. Only proteins identified with at least two different specific peptides in the same sample and found in at least two biological replicates were validated. Furthermore, quantification was performed on peptides with standard deviation retention times lower than 20 s.

## 2.4. Bioinformatics annotation of proteins and quantification

The prediction of sub-cellular localization of proteins was performed with the *ProtAnnDB* tool (<http://www.polebio.lrsv.ups-tlse.fr/ProtAnnDB/>, [9]). A protein was considered as a cell wall protein (CWP) if two bioinformatics programs predicted it as secreted, no intracellular retention signal was found and no more than one trans-membrane domain was predicted as described in Ref. [10]. Quantification was only operated for CWPs using the MassChroQ 2.2.12 software (<http://pappso.inra.fr/bioinfo/masschroq/>, [11]) and it was done as in Ref. [12]. Briefly, a background noise was used to replace missing data in order to facilitate the statistical analysis. This step was done differently to distinguish between validated (identification with at least two specific peptides in at least two of the three biological replicates of the ecotype/temperature combination), non-validated proteins (identification of only one specific peptide and/or in only one biological replicate) and undetectable proteins (no peptide identified in this combination). If the protein was validated, a background noise corresponding to the mean of the minimum and the first statistical quartile of the biological replicate was applied. If the protein was undetectable, a background noise of 6 (value lower than the minimum value found in the whole experiment) was applied [12].

## 2.5. RNA sequencing (RNAseq)

Protocols used for the transcriptomic analysis have been detailed in Ref. [12]. The RNAseq data have been obtained at the Get-PlaGe platform (<https://get.genotoul.fr/>). Short pair-end sequencing reads generated from the Illumina platform (<https://ng6.toulouse.inra.fr/>) were trimmed based on the quality scores (limit: 0.05), end ambiguity (maximum allowed number of ambiguities: 2) and adaptor sequences, using the commercial CLC Genomic Workbench 8.0 software (CLC bio, Aarhus, Denmark). The online CLC protocol for trimming and assembly has been followed (<https://www.qiagenbioinformatics.com/>). The reads shorter than 50 bp after trimming were discarded. The TAIR 10 database has been used for the assembly process (<https://www.arabidopsis.org/>). The calculation of gene expression level (RPKM) has been obtained with the CLC software (settings: minimum mapped read length fraction=0.95; minimum similarity=0.98). Finally, the expression levels lower than one RPKM per gene for the total of the conditions were considered as “not expressed”.

Note that only two biological replicates could be considered for the Grip and Hosp ecotypes grown at 15 °C due to insufficient quality of the reads of the third replicates.

## Acknowledgments

The authors are thankful to the Paul Sabatier Toulouse 3 University and to the *Centre National de la Recherche Scientifique* (CNRS) for granting their work. HD was supported by a grant from the Midi-Pyrénées Region and the Federal University of Toulouse.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.dib.2019.104581>.

## Conflict of 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] H. Duruflé, C. Albenne, E. Jamet, C. Dunand, Phenotyping and cell wall polysaccharide composition of five *Arabidopsis* ecotypes grown at optimal or sub-optimal temperatures, *Data Brief* 25 (2019) 104318, <https://doi.org/10.1016/j.dib.2019.104318>.
- [2] H. Duruflé, P. Ranocha, D.L. Mbadanga Mbadanga, S. Déjean, M. Bonhomme, H. San Clemente, S. Viudes, A. Eljebbawi, V. Delorme-Hinoux, J. Sáez-Vásquez, J.-P. Reichheld, N. Escaravage, M. Burrus, C. Dunand, Phenotypic trait variation as a response to altitude-related constraints in *Arabidopsis* populations, *Front. Plant Sci.* 10 (2019) 430, <https://doi.org/10.3389/fpls.2019.00430>.
- [3] D.C. Boyes, A.M. Zayed, R. Ascenzi, A.J. McCaskill, N.E. Hoffman, K.R. Davis, J. Görlach, Growth stage-based phenotypic analysis of *Arabidopsis*: a model for high throughput functional genomics in plants, *Plant Cell* 13 (2001) 1499–1510, <https://doi.org/10.1105/TPC.010011>.
- [4] L. Feiz, M. Irshad, R.F. Pont-Lezica, H. Canut, E. Jamet, Evaluation of cell wall preparations for proteomics: a new procedure for purifying cell walls from *Arabidopsis* hypocotyls, *Plant Methods* 2 (2006) 10, <https://doi.org/10.1186/1746-4811-2-10>.
- [5] M. Irshad, H. Canut, G. Borderies, R. Pont-Lezica, E. Jamet, A new picture of cell wall protein dynamics in elongating cells of *Arabidopsis thaliana*: confirmed actors and newcomers, *BMC Plant Biol.* 8 (2008) 94, <https://doi.org/10.1186/1471-2229-8-94>.
- [6] H. Duruflé, H. San Clemente, T. Balliau, M. Zivy, C. Dunand, E. Jamet, Cell wall proteome analysis of *Arabidopsis thaliana* mature stems, *Proteomics* 17 (2017) 8, <https://doi.org/10.1002/pmic.201600449>, 1600449.
- [7] O. Langella, B. Valot, T. Balliau, M. Blein-Nicolas, L. Bonhomme, M. Zivy, X! TandemPipeline: a tool to manage sequence redundancy for protein inference and phosphosite identification X!, *J. Proteome Res.* 16 (2017) 494–503, <https://doi.org/10.1021/acs.jproteome.6b00632>.
- [8] V. Hervé, H. Duruflé, H. San Clemente, C. Albenne, T. Balliau, M. Zivy, C. Dunand, E. Jamet, An enlarged cell wall proteome of *Arabidopsis thaliana* rosettes, *Proteomics* 16 (2016) 3183–3187, <https://doi.org/10.1002/pmic.201600290>.
- [9] H. San Clemente, R. Pont-Lezica, E. Jamet, Bioinformatics as a tool for assessing the quality of sub-cellular proteomic strategies and inferring functions of proteins: plant cell wall proteomics as a test case, *Bioinf. Biol. Insights* 3 (2009) 15–28, <https://doi.org/10.4137/BBI.S2065>.
- [10] C. Albenne, H. Canut, E. Jamet, Plant cell wall proteomics: the leadership of *Arabidopsis thaliana*, *Front. Plant Sci.* 4 (2013) 111, <https://doi.org/10.3389/fpls.2013.00111>.
- [11] B. Valot, O. Langella, E. Nano, M. Zivy, MassChroQ: a versatile tool for mass spectrometry quantification, *Proteomics* 11 (2011) 3572–3577, <https://doi.org/10.1002/pmic.201100120>.
- [12] H. Duruflé, V. Hervé, P. Ranocha, T. Balliau, M. Zivy, J. Chourré, H. San Clemente, V. Burlat, C. Albenne, S. Déjean, E. Jamet, C. Dunand, Cell wall modifications of two *Arabidopsis thaliana* ecotypes, Col and Sha, in response to sub-optimal growth conditions: an integrative study, *Plant Sci.* 263 (2017) 183–193, <https://doi.org/10.1016/j.plantsci.2017.07.015>.