OCL 2021, 28, 12
© T. Balliau et al., Hosted by EDP Sciences, 2021
https://doi.org/10.1051/ocl/2020074

OCL Oilseeds & fats Crops and Lipids

Available online at:
www.ocl-journal.org

DATA PAPER — OPEN ACCESS

Proteomic data from leaves of twenty-four sunflower genotypes under water deficit

Thierry Balliau[2,a], Harold Duruflé[1,a], Nicolas Blanchet[1], Mélisande Blein-Nicolas[2], Nicolas B. Langlade[1] and Michel Zivy[2,*]

[1] LIPME, Université de Toulouse, INRAE, CNRS, Castanet-Tolosan, France
[2] Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE–Le Moulon, PAPPSO, 91190 Gif-sur-Yvette, France

Received 11 May 2020 – Accepted 17 December 2020

Abstract – This article describes a proteomic data set produced from sunflower plants subjected to water deficit. Twenty-four sunflower genotypes were selected to represent genetic diversity within cultivated sunflower. They included both inbred lines and their hybrids. Water deficit was applied to plants in pots at the vegetative stage using the high-throughput phenotyping platform Heliaphen. We present here the identification of 3062 proteins and the quantification of 1211 of them in the leaves of the 24 genotypes grown under two watering conditions. These data allow the study of both the effects of genetic variations and watering conditions. They constitute a valuable resource for the community to study adaptation of crops to drought and the molecular basis of heterosis.

Keywords: Helianthus / abiotic stress / proteomics / drought / heterosis

Résumé – Données protéomiques produites à partir de vingt-quatre génotypes de tournesol soumis à un déficit hydrique. Cet article décrit un jeu de données protéomiques produites à partir de plantes de tournesol soumises ou non à un déficit hydrique. Vingt-quatre génotypes incluant des lignées pures et leurs hybrides ont été sélectionnés pour représenter la diversité génétique des tournesols cultivés. Les plantes ont été cultivées en pots sur la plateforme Heliaphen de phénotypage à haut débit et le déficit hydrique a été appliqué à un stade végétatif. Nous présentons l'identification de 3062 protéines et la quantification de 1211 d'entre elles dans les feuilles des vingt-quatre génotypes cultivés dans les deux conditions d'arrosage. Ces données permettent d'étudier l'effet des variations génétiques et du déficit hydrique sur le protéome. Elles sont une ressource intéressante pour la communauté, permettant d'étudier l'adaptation des plantes cultivées à la sécheresse et les bases moléculaires de l'hétérosis.

Mots clés : Helianthus / stress abiotique / protéomique / sécheresse / hétérosis

Subject area	Biology
More specific subject area	Proteomics
Type of data	Peptide and protein identification; protein quantification
How data was acquired	Mass spectrometry (LC-MS)
Data format	mzXML open format for raw mass spectrometry data; opendocument format for protein identification (.ods file); R data file for protein quantification (.RData file)
Experimental factors	24 genotypes of *Helianthus annuus* in two environmental conditions (irrigated or not) with 3 replicates
Experimental features	Identification and quantification of sunflower leaf proteins
Data source location	The outdoor Heliaphen phenotyping platform at the Institut national de la recherche agronomique (INRA) station, Auzeville, France (43°31'41.8"N, 1°29'58.6"E)
Data accessibility	These data are publicly available in ProticDB with following DOI: https://doi.org/10.15454/TW59-P718
Related research article	Badouin *et al.*, 2017; Blanchet *et al.*, 2018

*Correspondence: michel.zivy@inrae.fr
[a] Co-first authors

Table 1. Parameters used for mass spectrometry analyses.

Identification software	X!Tandem Piledriver (2015.04.01.1) http://www.thegpm.org/
Filtering and inference software	X!TandemPipeline 3.4.3 "Elastine Durcie"
Filters	Proteins: log (e-value) < -5 Proteins: minimum 2 peptides Peptides: e-value <0.01
Digestion	Trypsine No semi-tryptic peptide allowed 1 misscleavage allowed in first pass and 5 in refine pass
Modifications	**Fixed** Carbamidomethylation of Cys residues = +57.04 **Possible** Oxidation of Met and Trp residues = +15.99 N-ter acetylation = +42.01 N-ter deamidation of Gln residues = −17.02 Nter deamidation on carbamidomethylated Cys = −17,02 Loss of H_2O on N-ter Glu residues = −18.01 Deamidation of Gln and Asn residues = +0.98 N-ter Met excision

1 Specifications table

2 Value of the data

Climate change is a current issue of major concern because of its actual effects on biodiversity and the agricultural sector. A better understanding of the adaptation of plants to this recent phenomenon is therefore of major interest to crop science and society. *Helianthus annuus* L., the domesticated sunflower, is the fourth most important oilseed crop in the world (USDA, 2020). It is promising for agriculture adaptation to climate change as it can maintain yield better than most other crops in a wide range of environments, especially during drought stress (Debaeke and Bertrand, 2008; Debaeke *et al.*, 2017). It constitutes an archetypical systems biology model, as response to drought stress involves many molecular pathways and subsequent physiological processes. In addition, heterosis is a phenomenon commonly used to improve yield in allogamous crops such as sunflower. To study its impact on the response to drought at the molecular level, we analysed the leaf proteome of twenty-four genotypes of cultivated sunflower comprising four maintainer lines, four restorer lines and their 16 hybrids, that were either well watered or subjected to water deficit. Different ecophysiological traits were measured on these plants (Blanchet *et al.*, 2018), making it possible to study the relationships between protein expression and plant traits as a function of water stress and heterozygosity.

3 Data

This dataset provides identification and quantification data for proteins of sunflower leaves from 142 plants distributed in 24 genotypes grown in water deficit (WD) or well watered (WW) conditions with 3 replicates (two missing data).

The raw data associated with this article (data from the mass spectrometer in mzXML format) as well as annotated spectra, identification and quantitative data can be found at the following link https://doi.org/10.15454/TW59-P718 or directly at http://moulon.inra.fr/protic/sunrise. Parameters used for mass spectrometry analyses are shown in Table 1.

4 Experimental design, plant material and growth conditions

One hundred and forty-four plants corresponding to 24 genotypes (four female lines [SF009, SF092, SF109 and SF109], four male lines [SF279, SF317, SF326, SF342]) and their 16 hybrids were grown under two treatment conditions (WW and WD), with three plants per genotype and per treatment.

The experiment design and plant material are fully described in Blanchet *et al.* (2018). Briefly, the experiment was performed in the outdoor Heliaphen phenotyping platform at the Institut national de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE) station, Auzeville, France. The growth conditions were as previously described in Gosseau *et al.* (2019). In summary, seeds were germinated in Petri dishes for 78 h at 28 °C, then germinated plantlets were transplanted in 15 L pots in the Heliaphen platform. Pots were covered with a polystyrene sheet to prevent evaporation from the soil.

Pots were normally irrigated, according to Rengel *et al.* (2012), up to 38 days after germination (DAG; ~20-leaf stage corresponding to bud formation phase (Schneiter and Miller, 1981). At this stage, irrigation of WD plants was stopped. Soil evaporation was estimated according to Marchand *et al.* (2013). Both WW and WD plants were weighed three or four times per day by a robot to estimate transpiration (Gosseau *et al.*, 2019). After weighting, the robot watered WW plants to maintain soil water at retention capacity, while WD plants were not watered. When the fraction of transpirable soil water reached 0.1 for a WD plant, a leaf sample was harvested both from this plant and from a WW plant. The harvested leaf was the leaf above the leaf that had reached its maximum size the most recently. Two plants (SF342 inbred line) died during the experiment and could not be harvested.

5 Proteomics

5.1 Protein extraction

Leaf proteins were extracted using the TCA-acetone protocol described in Méchin *et al.* (2007). Protein digestion was performed according to the liquid digestion protocol described in Hervé *et al.* (2016). Briefly, proteins of the TCA-acetone pellet were solubilized in a buffer containing urea, thiourea, dithiothreitol, Tris-Hcl pH 8.8 and a zwitterionic acid labile surfactant (ZALSI). Proteins were alkylated by using iodoacetamide and digestion by trypsine was performed after dilution in an ammonium bicarbonate solution. Digestion was stopped by adding trifluoroacetic acid that also allowed ZALS cleavage. The resulting peptide mix was desalted by using C18 solid solid phase extraction. Eluted peptides were then speedVac dried.

LC-MS/MS was performed as described in Duruflé *et al.* (2017). Peptides (400 ng) were solubilized in a solution containing 2% acetonitrile and 0.1% formic acid. LC-MS/MS was performed by using an Eksigent nanoLC Ultra 2D nanoHPLC (SCIEX) coupled to a Qexactive mass spectrometer (Thermo, Waltham, MA, USA). After desalting on a trapcolumn, peptides were submitted to a gradient of 5 to 35% ACN that was carried out in 40 min. Data-dependent MS analysis was performed with full scans at a 75,000 resolution and MS/MS scans at a 17,500 resolution. The isolation window was set to 3 m/z. MS/MS was repeated for the eight most intense ions detected in full scan and dynamic exclusion was set to 40 s.

5.2 Identification of proteins by LC-MS/MS

Protein identification was made by searching the Heliagen database using genome HanXRQv1 (Badouin *et al.*, 2017) with the X!Tandem search engine (Craig *et al.*, 2004). Data filtering and protein inference were performed by using X!TandemPipeline 3.3.4 (Langella *et al.*, 2017). Trypsin digestion was set with one and five possible miss cleavages in the first and refine pass, respectively. Only proteins identified with at least two different peptides in the same sample were considered (Valot *et al.*, 2011).

5.3 Bioinformatics annotation of proteins and quantification

Quantification by integration of the extracted ion current (XIC) was operated using the MassChroQ software (Valot *et al.*, 2011). Only proteins quantified with at least 2 specific peptides that were present in at least 90% of the samples were selected for analysis. A total of 1211 of the 3062 identified proteins met this criterion. Functional annotation of proteins is given according to the INRAE Sunflower Bioinformatics Resources (www.heliagene.org/HanXRQ-SUNRISE/).

Acknowledgments. These data were produced with the funding of the French National Research Agency (ANR SUNRISE ANR-11-BTBR-0005). This work was part of the "Laboratoire d'Excellence (LABEX)" TULIP (ANR-10-LABX-41).

References

Badouin H, Gouzy J, Grassa CJ, *et al*. 2017. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature* 546: 148–152.

Blanchet N, Casadebaig P, Debaeke P, *et al*. 2018. Data describing the eco-physiological responses of twenty-four sunflower genotypes to water deficit. *Data in Brief* 21: 1296–1301.

Craig R, Cortens JP, Beavis RC. 2004. Open source system for analyzing, validating, and storing protein identification data. *Journal of Proteome Research* 3: 1234–1242.

Debaeke P, Bertrand M. 2008. Évaluation des impacts de la sécheresse sur le rendement des grandes cultures en France. *Cahiers Agricultures* 17: 437–443.

Debaeke P, Casadebaig P, Flenet F, Langlade N. 2017. Sunflower crop and climate change: vulnerability, adaptation, and mitigation potential from case-studies in Europe. *OCL* 24: D102.

Duruflé H, Clemente HS, Balliau T, Zivy M, Dunand C, Jamet E. 2017. Cell wall proteome analysis of Arabidopsis thaliana mature stems. *Proteomics* 17: 1600449.

Gosseau F, Blanchet N, Varès D, *et al*. 2019. Heliaphen, an outdoor high-throughput phenotyping platform for genetic studies and crop modeling. *Frontiers in Plant Science* 9: 1908.

Hervé V, Duruflé H, San Clemente H, *et al*. 2016. An enlarged cell wall proteome of Arabidopsis thaliana rosettes. *Proteomics* 16: 3183–3187.

Langella O, Valot B, Balliau T, Blein-Nicolas M, Bonhomme L, Zivy M. 2017. X!TandemPipeline: a tool to manage sequence redundancy for protein inference and phosphosite identification. *J Proteome Res* 16: 494–503.

Marchand G, Mayjonade B, Varès D, *et al*. 2013. A biomarker based on gene expression indicates plant water status in controlled and natural environments. *Plant, Cell & Environment* 36: 2175–2189.

Méchin V, Damerval C, Zivy M. 2007. Total protein extraction with TCA-acetone. *Methods in Molecular Biology (Clifton, N.J.)* 355: 1–8.

Rengel D, Arribat S, Maury P, *et al*. 2012. A gene-phenotype network based on genetic variability for drought responses reveals key physiological processes in controlled and natural environments (M. Bendahmane, Ed.). *PLoS ONE* 7: e45249.

Schneiter AA, Miller JF. 1981. Description of sunflower growth stages. *Crop Sciences* 21: 901–903.

USDA. 2020. Oilseeds: world markets and trade. [WWW document]. Available from https://apps.fas.usda.gov/psdonline/circulars/oilseeds.pdf [accessed 3 July 2020].

Valot B, Langella O, Nano E, Zivy M. 2011. MassChroQ: a versatile tool for mass spectrometry quantification. *Proteomics* 11: 3572–3577.

Cite this article as: Balliau T, Duruflé H, Blanchet N, Blein-Nicolas M, Langlade NB, Zivy M. 2021. Proteomic data from leaves of twenty-four sunflower genotypes under water deficit. *OCL* 28: 12.