Leaf metabolomic data of eight sunflower lines and their sixteen hybrids under water deficit

Thierry Berton1,2,a, Stéphane Bernillon1,2,a, Olivier Fernandez1,2, Harold Durufle3,e, Amélie Flandin1,2, Cédric Cassan1,2, Daniel Jacob1,2, Nicolas B. Langlade3, Yves Gibon1,2 and Annick Moing1,2,*

1 INRAE, Univ. Bordeaux, Biologie du fruit et pathologie, UMR 1332, Centre INRAE de Nouvelle Aquitaine–Bordeaux, 33140 Villenave d’Ornon, France
2 Bordeaux Metabolome, MetaboHUB**, PHENOME, IBVM, Centre INRAE de Nouvelle Aquitaine–Bordeaux, 33140 Villenave-d’Ormon, France
3 LIPME, Université de Toulouse, INRAE, CNRS, Castanet-Tolosan, France

Received 20 May 2021 – Accepted 13 July 2021

Abstract – This article describes how metabolomic data were produced on sunflower plants subjected to water deficit. Twenty-four sunflower (Helianthus annuus L.) genotypes were selected to represent genetic diversity within cultivated sunflower and included both inbred lines and their hybrids. Drought stress was applied at the vegetative stage to plants cultivated in pots using the high-throughput phenotyping facility Heliaphen. Here, we provide untargeted and targeted metabolomic data of sunflower leaves. These compositional data differentiate both plant water status and different genotype groups. They constitute a valuable resource for the community to study the adaptation of crops to drought and the metabolic bases of heterosis.

Keywords: Helianthus / abiotic stress / drought stress / LC-MS / metabolomic profiling

Résumé – Données métabolomiques foliaires de huit lignées de tournesol et de leurs seize hybrides sous déficit hydrique. Cet article décrit comment les données métabolomiques ont été produites sur des plants de tournesol soumis à un déficit hydrique. Vingt-quatre génotypes de tournesol (Helianthus annuus L.) ont été sélectionnés pour représenter la diversité génétique du tournesol cultivé et comprennent à la fois des lignées consanguines et leurs hybrides. Une limitation hydrique a été appliquée au stade végétatif aux plantes cultivées en pots à l’aide de la plateforme de phénométrie à haut débit Heliaphen. Ici, nous mettons à disposition des données métabolomiques non ciblées et ciblées de feuilles de tournesol. Ces données de composition permettent de différencier l’état hydrique des plantes et différents groupes de génotypes. Elles constituent une ressource précieuse pour la communauté afin d’étudier l’adaptation des cultures à la sécheresse et les bases métaboliques de l’hétérosis.

Mots clés : Helianthus / stress abiotique / stress hydrique / LC-MS / profils métabolomiques

Highlights

Leaf metabolomic data were produced on sunflower plants of inbred lines and their hybrids subjected to water deficit at the vegetative stage.

They differentiate both plant water status and different genotype groups.

They constitute a valuable resource to be combined with other omics data and study the adaptation to drought and the bases of heterosis.

© T. Berton et al., Published by EDP Sciences, 2021

Available online at: www.ocl-journal.org

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
1 Specifications table

| Subject area          | Biology                      |
|-----------------------|------------------------------|
| More specific subject area | Metabolomc data             |
| Type of data          | LC-MS: LC-MS acquisition files, R command text file for spectra processing, LC-MS/MS acquisition files, Word file for LC-MS annotation table, tab file for calculated data table |
| How data was acquired | The Heliaphen robot and targeted robotized analyses of major compounds or LC-MS analyses of polar extracts |
| Data format           | Targeted-analyses processed data: txt |
| Experimental factors  | 24 genotypes of Helianthus annuus in two environmental conditions (irrigated or not) with three replicates |
| Experimental features | Absolute contents of major compounds of sunflower leaf |
| Data source location  | The outdoor Heliaphen phenotyping platform at INRAE station, Auzeville-Tolosane, France (43°31'41.8″N, 1°29'58.6″E) |
| Data accessibility    | The LC-MS data are publicly available in Data INRAE repository (https://data.inrae.fr/dataverse/sunflower, https://doi.org/10.15454/2KOXOH) under license etalab-2.0 |
|                      | The targeted analyses data are publicly available in Data INRAE repository (https://data.inrae.fr/dataverse/sunflower, https://doi.org/10.15454/STJH47) under license etalab-2.0 |
| Related research article | (Blanchet et al., 2018; Gody et al., 2020; Balliau et al., 2021) |

2 Value of the data

Drought stress is a crucial issue for crop adaptation to climate change and sunflower is particularly impacted as it is mostly cultivated in marginal lands (Debaeke et al., 2017). In the present experiment, plants were subjected to two treatments (Well-Watered or Water-Deficit) during the vegetative stage. This experiment was performed in the outdoor high-throughput, semi-automated phenotyping facility Heliaphen (https://www6.inrae.fr/phenotoul_eng/WHO-we-are/PhenoToul/HeliaPhen).

Heterosis is an outstanding phenomenon involved in natural selection and used in crop breeding to adapt plants to environmental constraints. Twenty-four genotypes of cultivated sunflower consisting in four maintainer lines, four restorer lines and their 16 corresponding hybrids are included in this experiment which allows studying heterosis effect on metabolism.

This dataset provides metabolomic data of sunflower leaves of lines and hybrids under control and water deficit conditions. These data consist in unique untargeted and targeted metabolomic profiles of sunflower responses to drought based on a large genetic variability.

3 Data

Climate change is affecting plant biodiversity, and crop choice and yields. A better knowledge of plant adaptation mechanisms to this recent phenomenon is, therefore, of major interest for crop science, agriculture and for feed and food security (Porter et al., 2019). Helianthus annuus L., the domesticated sunflower, is the fourth most important oilseed crop in the world (USDA, 2019). It seems promising for agriculture adaptation to global change because it can maintain stable yields across a range of environmental conditions, especially during stress induced by water limitation (Debaeke et al., 2017). It can be considered as an archetypical systems biology model with large drought stress response which involves many molecular pathways (Moschen et al., 2017) and subsequent metabolic and physiological processes.

In this data article, we are sharing the metabolomic data of 24 sunflower genotypes grown in two environmental conditions in an outdoor phenotyping facility. This dataset is part of a larger project that integrates other omics data (Blanchet et al., 2018; Gody et al., 2020; Balliau et al., 2021).

The LC-MS data and metadata associated with this article were deposited in the Data INRAE repository. The targeted analyses data were deposited in the Data INRAE repository.

4 Experimental design, plant material and growth conditions

The experiment was performed from May to July 2013 on the outdoor Heliaphen phenotyping facility at the Institut National de Recherche pour l’Agriculture, l’Alimentation et l’Environnement (INRAE) station, Auzeville, France (43°31’41.8″N, 1°29’58.6″E) as previously described (Blanchet et al., 2018; Gosseau et al., 2019). Briefly, germinated plantlets were transplanted into individual pots filled with 15-l potting soil and covered with a 3-mm-thick polystyrene sheet to prevent soil water evaporation. Plants were fertilized with Peters Professional fertilizer (17-07-27; 500 mL, 0.6 g/L) and an oligo-element mixture soluton (Hortilion, 0.46 g/L) at 17 days after germination (DAG), and treated with Polyaxe (5 mg/L applied on foliage) against thrips at 21 DAG.
In total, 144 plants, corresponding to 24 genotypes, four maintainer (SF009, SF092, SF109 and SF193) and four restorer (SF279, SF317, SF326 and SF342) lines and their corresponding hybrids obtained by crossing, were grown in two conditions: well-watered (WW) and water-deficit (WD) with three biological replicates (Blanchet et al., 2018; Gody et al., 2020). Before the beginning of the water deficit application at 35 DAG, pots were saturated with water and excessive water was drained. Pots were weighed to obtain the full soil water retention mass. At 38 DAG, irrigation was stopped (approximately 20-leaf stage) for WD plants as described previously (Gosseau et al., 2019). Plants were weighed by the Heliaphen robot to estimate transpiration (Gosseau et al., 2019). WW plants were re-watered at each weighing to reach soil water full retention capacity. Pairs of WD and WW plants were harvested when the fraction of transpirable soil water of the stressed plant reached 0.1 (occurring between 42 and 47 DAG). Two out of the three SF342 line plants died under the WW condition. The corresponding plant samples could not be harvested and data could not be obtained.

At harvest, leaves for metabolome analyses were cut without their petiole and immediately frozen in liquid nitrogen from 11 a.m. to 1 p.m. The harvested leaf was the leaf above the leaf that had reached its maximum size the most recently, as determined using a robotic Star/Starlet platform (Hamilton, Villebon-sur-Yvette, France) and spectrophotometers.

5.2 Targeted analyses of major compounds

The targeted analyses of major compound in all samples were performed as done previously for the parents only (Fernandez et al., 2019) and as previously described (Biais et al., 2014) using enzymatic analyses and colorimetric assays performed using a robotic Star/Starlet platform (Hamilton, Villebon-sur-Yvette, France) and spectrophotometers. Glucose, fructose and sucrose were determined in the ethanolic supernatant obtained as described above (Stitt et al., 1989) and expressed in μmol per g dry weight (DW). Total free amino acids were determined in the supernatant with a fluorescamine-based assay (Bantan-Polak et al., 2001) and expressed as glutamate equivalents. Protein content was determined (Bradford, 1976) on the pellet re-suspended in 100 mM NaOH and heated at 95 °C for 20 min and expressed as mg bovine serum albumine equivalents per g DW. After neutralisation of the suspended pellet, starch was determined and expressed in glucose equivalents per g DW (Hendriks et al., 2003). Absorbencies were read at 340 or 595 nm using an MP96 microplate reader (SAFAS, Monaco). For fluorescence, 405 nm excitation and 485 nm emission were used with a Xenius multifunction microplate reader (SAFAS, Monaco).

All chemicals and substrates for targeted analyses were purchased from Sigma-Aldrich Ltd. (Gillingham, United Kingdom). All enzymes were purchased from Roche Applied Science (Meylan, France).

5.3 LC-MS based metabolomic profiling

LC-MS-based metabolomic profiling of extracts was performed using the same extracts as for targeted analyses. The sample injection order was randomized. The QC sample was injected every 12 samples to correct for mass spectrometer signal drift. The extracts were analysed using LC-MS (Ultimate 3000 – LTQ-Orbitrap Elite, ThermoScientific, Bremen, Germany), using a C18 chromatographic column (C18-Gemini 2.0 × 150 mm, 3 μm, 110 Å, Phenomenex, Torrance, CA, USA), a 18 min acetonitrile gradient in acidified water (solvent A: ultrapure water + 0.1% formic acid, solvent B: LC-MS grade acetonitrile) with a 300 μL min⁻¹ flow rate and the following elution gradient: 0-0.5 min, 3% B; 0.5-1 min, 3-10% B; 1-9 min, 10-50% B; 9-13 min, 50-100% B; 13-14 min, 100% B; 14-14.5 min 100-3% B; 14.5-18 min, 3% B. The column temperature was 30°C. The injection volume was 5 μL. The LC-MS instrument was equipped with an HESI source operated in the positive-ion mode. Source parameters were the following: source voltage, 3.2 kV; sheath gas, 45 arbitrary units (a.u.); auxiliary gas, 15 a.u.; sweep gas, 0 a.u.; capillary temperature, 350°C, heater temperature, 350°C. Full Scan MS spectra were acquired at 240k resolution power with a 50-1000 mass range. Data dependent MS/MS spectra were acquired at 60k resolution power. The selected ions were fragmented in CID mode at a 35% normalized collision energy. The MS data were processed using R (R Core Team, 2018) with XCMS (Smith et al., 2006) and MetNormalizer (Shen et al., 2016) packages. Briefly, the corresponding MS-based variables were named using their nominal masses in Da and retention time in s (MxxxyTyyt). Variables detected in blank extracts were filtered out. Variables with m/z values varying by more than 0.005 Da or with retention time varying by more than 20 s between different samples were also filtered out. Variables with intensity coefficients of variation in QC greater than 20% were also removed. This resulted in a data matrix of 4843 variables. Intensity drift was corrected using support vector regression. Finally, intensities were normalized according to the sample powder mass used for extraction. Annotation of intense ions
Table 1. Annotation of LC-MS signatures of sunflower leaf ethanolic extracts with LC-MS and LC-MS/MS data in positive ionization mode.

| Variable | Metabolite [M + H]+ a | Metabolite [M + H]+ a | LC-MS/MS fragments m/z a,b | Putative name | Metabolite ID | Class | Neutral molecular formula | Calculated molecular formula [M + H]+ m/z | MSI level c |
|----------|------------------------|------------------------|-----------------------------|---------------|--------------|-------|--------------------------|----------------------------------------|------------|
| M284T195 | 3.26 284.0997 152.0568 | Guanosine | CHEBI:16750 | Nucleosides | C10H13N5O5 | 284.09894 | 2 |
| M166T209 | 3.30 166.0864 120.0806 | Phenylalanine | CHEBI:28044 | Amino acids | C9H11NO2 | 166.0862 | 2 |
| M382T210 | 3.50 382.1730 220.1195; 202.1091; 136.0618 | Pantothenic acid-hexose | – | Amino compounds | C15H27NO10 | 382.17077 | 3 |
| M285T209 | 3.50 285.0837 ND | Xanthosine | CHEBI:18107 | Nucleosides | C10H12N4O6 | 285.08296 | 3 |
| M336T275 d | 4.60 355.1024 ND | 5-O-Caffeoylquinic acid | CHEBI:16384 | Cinnamic acids | C16H18O9 | 355.10236 | 2 |
| M342T281 d | 4.68 341.0869 179.0340 | Esculin | CHEBI:4853 | Coumarins | C13H10O6 | 341.0871 | 2 |
| M336T315 d | 5.25 355.1017 163.0392; 145.0286 | 3-O-Caffeoylquinic acid | CHEBI:16112 | Cinnamic acids | C16H18O9 | 355.10236 | 2 |
| M336T341 d | 5.68 355.1023 ND | 4-O-Caffeoylquinic acid | CHEBI:75491 | Cinnamic acids | C16H18O9 | 355.10236 | 2 |
| M340T362 d | 6.04 339.1179 321.1128; 147.0442 | Coumaroylquinic acid | CHEBI:1945 | Cinnamic acids | C16H18O9 | 339.10744 | 2 |
| M611T367 | 6.10 611.1630 465.1028; 303.0502 | Rutin | CHEBI:28527 | Flavonoids | C21H20O12 | 611.16066 | 2 |
| M465T374 | 6.21 465.1034 303.0503 | Quercetin hexoside | – | Flavonoids | C21H20O12 | 465.10257 | 3 |
| M370T374 d | 6.22 369.1179 207.0995; 177.0550; 145.0286 | Feruloylquinic acid | CHEBI:86388 | Cinnamic acid | C17H20O8 | 369.11800 | 2 |
| M642T389 d | 6.47 641.1736 495.1133; 333.0603 | Penta hydroxyl, methoxyflavone hexoside-deoxyhexoside | – | Flavonoids | C28H30O17 | 641.17122 | 3 |
| M518T468 d | 7.74 517.1334 ND | 3,4-Dicaffeoylquinic acid | CID: 5281780 | Cinnamic acids | C25H24O12 | 517.13405 | 3 |
| M549T502_2 | 8.36 549.1256 ND | Trihydroxy, methoxyflavone malonylhexoside | – | Flavonoids | C25H24O14 | 549.12388 | 3 |
| M379T524 | 8.71 379.1747 ND | Niveusin C or Hydroxyleptocarpin | – | Sesquiterpenoids | C21H26O7 | 379.17513 | 3 |
| M274T582 | 9.68 273.0759 255.0655; 163.0392; 137.0234 | Butein | CHEBI:3237 | Flavonoids | C15H12O5 | 273.07574 | 2 |
| M409T617 | 10.28 409.1862 ND | 3-O-MethylNiveusin A | CID: 131752540 | Sesquiterpenoids | C21H28O8 | 409.18569 | 3 |

a Rt and m/z measured with the raw data files. Only the [M + H]+ ion was considered as it was the most abundant ion in each acquired MS spectrum.

b MS/MS fragmentation data from data dependent scan in CID mode at a 35% normalized collision energy value; ND: spectrum was not acquired.

c Level of metabolite identification (Sumner et al., 2007).

d This variable corresponds to M1 of the [M + H]+ ion.
(Fernandez et al., 2019; Stelzner et al., 2019) was performed using RT, accurate m/z and fragment ions from an MS/MS acquisition of an aliquot of the QC sample. This resulted in the annotation of 18 compounds belonging to eight compound families (Tab. 1). All chemicals for LC-MS analyses were purchased from Sigma Aldrich (Saint-Quentin Fallavier, France) and Extrasynthèse (Genay, France).

Finally, due to plant death or the lack of leaf material for several plants, 121 and 125 samples out of the 144 initial ones were analysed by the targeted (6 variables) and LC-MS based metabolomic (4843 variables) approaches, respectively. To get an overview of each data set, a principal component analysis (PCA) was performed using BioStatFlow web tool (Jacob et al., 2020) on data mean-centred and scaled to unit variance. The two treatments tended to separate along PC2 explaining about 28% of total variability for the targeted analyses (Fig. 1A) and about 9% for the LC-MS based data (Fig. 1B). The lines and hybrids tended to separate along PC1 explaining about 10% of total variability for the LC-MS based data (Fig. 1B). These metabolome data can be combined with other omic and phenotypic data of the same samples (Blanchet et al., 2018; Gody et al., 2020; Balliau et al., 2021) to get deeper insights into drought effects and heterosis.

**Supplementary material**

DATA-TargetedAnalyses-SunflowerLeaf.txt: This file contains targeted measurements of major compounds for each genotype and their three biological replicates (in columns) for WW and WD conditions.

DATA-LCMS-SunflowerLeaf.txt: This file contains the intensities of LC-MS-Orbitrap metabolite signatures for each genotype and their three biological replicates (in columns) for WW and WD conditions.

The Supplementary Material is available at http://www.ocl-journal.org/10.1051/ocl/2021029/olm.

**Acknowledgments.** We thank the Heliaphen team (especially Nicolas Blanchet) for plant culture. These data were produced with the funding of the French National Research Agency (SUNRISE ANR-11-BTBR-0005, MetaboHUB ANR-11-INBS-0010, PHENOME ANR-11-INBS-0012). This work was part of the “Laboratoire d’Excellence (LABEX)” TULIP (ANR-10-LABX-41).

**Conflicts of interest.** The authors declare that they have no conflicts of interest in relation to this article.

**References**

Balliau T, Duruflé H, Blanchet N, et al. 2021. Proteomic data from leaves of twenty-four sunflower genotypes underwater deficit. *OCL* 28: 12. https://doi.org/10.1051/ocl/2020074.

Bantan-Polak T, Kassai M, Grant KB. 2001. A comparison of fluorescamine and naphthalene-2, 3-dicarboxaldehyde fluorogenic reagents for microplate-based detection of amino acids. *Anal Biochem* 297(2): 128–136. https://doi.org/10.1006/abio.2001.5338.

Biais B, Bénard C, Beauvoit B, et al. 2014. Remarkable reproducibility of enzyme activity profiles in tomato fruits grown under contrasting environments provides a roadmap for studies of fruit metabolism. *Plant Physiol* 164(3): 1204–1221. https://doi.org/10.1104/pp.113.231241.

Blanchet N, Casadebaig P, Debaeke P, et al. 2018. Data describing the eco-physiological responses of twenty-four sunflower genotypes
Debaeke P, Casadebaig P, Flénet F, et al. 2017. Sunflower crop and climate change: vulnerability, adaptation, and mitigation potential from case-studies in Europe. *OCL* 24(1): 15. https://doi.org/10.1051/ocl/2016052.

Fernandez O, Urrutia M, Berton T, et al. 2019. Metabolomic characterization of sunflower leaf allows discriminating genotype groups or stress levels with a minimal set of metabolic markers. *Metabolomics* 15(4): 56. https://doi.org/10.1007/s11306-019-1515-4.

Hendriks JHM, Kolbe A, Gibon Y, et al. 2003. ADP-glucose pyrophosphorylase is activated by posttranslational redox-modification in response to light and to sugars in leaves of Arabidopsis and other plant species. *Plant Physiol* 133(2): 838–849. https://doi.org/10.1104/pp.103.024513.

Sumner LW, Amberg A, Barrett D, et al. 2007. Proposed minimum reporting standards for chemical analysis. *Metabolomics* 3(3): 211–221. https://doi.org/10.1007/s11306-007-0082-2.