Données transcriptomiques de feuilles de huit lignées de tournesol et de leurs 16 hybrides soumis à un stress hydrique

Louise Gody, Harold Duruflé, Nicolas Blanchet, Clément Carré, Ludovic Legrand, Baptiste Mayjonade, Stéphane Muños, Lise Pomies, Simon De Givry, Nicolas Langlade, et al.

HAL Id: hal-02974900

https://hal.inrae.fr/hal-02974900v1

Submitted on 22 Oct 2020

OCL
Oilseeds & fats Crops and Lipids

DATA PAPER

OPEN ⧉ ACCESS

Transcriptomic data of leaves from eight sunflower lines and their sixteen hybrids under water deficit

Louise Gody[1,a], Harold Duruflé[1,a], Nicolas Blanchet[1], Clément Carré[2], Ludovic Legrand[1], Baptiste Mayjonade[1], Stéphane Muños[1], Lise Pomiès[2], Simon de Givry[2], Nicolas B. Langlade[1,*] and Brigitte Mangin[1]

[1] LIPM, Université de Toulouse, INRAE, CNRS, Castanet-Tolosan, France
[2] MIAT, Université de Toulouse, INRAE, Castanet-Tolosan, France

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

Keywords: *Helianthus* / abiotic stress / phenotyping platform / drought stress

Résumé – Données transcriptomiques de feuilles de huit lignées de tournesol et de leurs 16 hybrides soumis à un stress hydrique. Cet article décrit la production de données transcriptomiques sur des plantes de tournesol soumises à un stress hydrique. Vingt-quatre génotypes de tournesol (*Helianthus annuus*) ont été sélectionnés pour représenter la diversité génétique parmi le tournesol cultivé et comprennent à la fois des lignées et leurs hybrides. Le stress hydrique a été appliqué sur des plantes en pots au stade végétatif grâce à la plateforme de phénotypage haut-débit Heliaphen. Ici, nous mettons à disposition les données transcritomiques de feuilles de ces plantes. Les données permettent de différencier à la fois le statut hydrique et les différents génotypes. Elles constituent une ressource importante pour la communauté pour étudier l'adaptation des plantes à la sécheresse et les bases de l'hétérosis au niveau transcriptomique.

Mots clés : *Helianthus* / stress abiotique / plateforme de phénotypage / stress hydrique

Specifications table

Subject area	Biology
More specific subject area	Transcriptomic data
Type of data	Table, .csv files
How data was acquired	The Heliaphen robot and Illumina HiSeq 2000
Data format	Raw, processed and filtered data
Experimental factors	24 genotypes of Helianthus annuus in two environmental conditions (irrigated or not) with 3 replicates
Experimental features	RNAseq 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	These data are publicly available in GEO depository with following GSE145709 code accession Resulting sequences are available at NCBI short read archive (SRA): BioProject PRJNA345532Note to referees regarding public databases in which experimental data have been deposited: URL: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE145709token:yzwrmyworvclrql
Related research article	Badouin *et al.*, 2017; Blanchet *et al.*, 2018

*Correspondence: `nicolas.langlade@inrae.fr`
[a] Co-first authors.

1 Value of the data

Drought stress is an important issue related to crop adaptation to climate change and sunflower is particularly impacted as it is grown mainly in marginal lands (Debaeke *et al.*, 2017). In this experiment, plants were subjected during the vegetative stage to two treatments (Well-Watered or Water-Deficit) managed on the outdoor high-throughput phenotyping platform Heliaphen.

Heterosis is the most outstanding phenomenon used by natural selection and mankind to adapt plants to environmental constraints. Twenty-four genotypes of cultivated sunflower comprising four maintainer lines, four restorer lines and their 16 corresponding hybrids are included in this experiment and allow the study of heterosis.

This dataset provides transcriptomic data of sunflower leaves under water deficit.

These data represent a unique transcriptomic profiling of sunflower responses to drought including a large genetic variability.

2 Data

Climate change is a current issue of major concern because of its potential effects on biodiversity and the agricultural sector. Better understanding of adaptation of plants to this recent phenomenon is, therefore, a major interest for crop science and society. *Helianthus annuus* L., the domesticated sunflower, is the fourth most important oilseed crop in the world (USDA, 2019) and is promising for agriculture adaptation because it can maintain stable yields across a wide variety of environmental conditions, especially during drought stress (Badouin *et al.*, 2017). It constitutes an archetypical systems biology model with large drought stress response, which involves many molecular pathways and subsequent physiological processes.

In this data article, we are sharing the transcriptomic data of 24 sunflower genotypes grown in two environmental conditions in the outdoor Heliaphen platform. This dataset is part of a larger project that integrates other omics data at different biological levels (Blanchet *et al.*, 2018).

The raw data associated with this article can be found at NCBI SRA BioProject PRJNA345532 and the table of counts is available at the GEO depository with GSE145709 code accession.

3 Experimental design, plant material and growth conditions

The experiment was performed from May to July 2013 on the outdoor Heliaphen phenotyping platform 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 in Gosseau *et al.* (2019). Bleach-sterilized seeds were germinated on Petri dishes with Apron XL and Celeste solutions (Syngenta, Basel, Switzerland) for 78 hours at 28 °C. Germinated plantlets were transplanted in individual pots

filled with 15 L of P.A.M.2 potting soil (Proveen distributed by Soprimex, Chateaurenard, Bouches-du-Rhône, France) and covered with a 3-mm thick polystyrene sheet to prevent soil water evaporation. Seventeen days after germination (DAG), plants were fertilized with 500 mL of Peter's Professional 17-07-27 (0.6 g/L) and extra mix composed of oligo-element Hortilon (0.46 g/L) solution. Twenty-one DAG, Polyaxe at 5 mg/L was applied on foliage against thrips.

In total, 144 plants, corresponding to 24 genotypes (four maintainers and four restorer 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). Each pot was adequately fertilized and irrigated as in Rengel *et al.* (2012) before the beginning of the water deficit application at 35 DAG, pots were saturated with water and excessive water was drained (∼ for two hours), pots were weighed to obtain the full soil water retention mass. Thirty-eight DAG, irrigation was stopped (∼20-leaf stage corresponding to stage R1, R2 or R3 according to genotypes; Schneiter and Miller, 1981) for WD plants as described in Gosseau *et al.* (2019). Soil water evaporation was estimated according to Marchand *et al.* (2013). Both WW and WD plants were weighed three or four times per day by the Heliaphen robot to estimate transpiration (Gosseau *et al.*, 2019). WW plants were re-watered at each weighing by the robot to reach soil water full retention capacity. Pairs of WD and WW plants were harvested when the Fraction of Transpirable Soil Water (FTSW) of the stressed plant reached 0.1 (occurring between the 42 and the 47 DAG). Two out of three SF342 plants died under control condition. Plant samples could not be harvested and data could not be obtained.

At harvest, leaves for molecular analysis were cut without their petiole and immediately frozen in liquid nitrogen from 11 a.m. to 1 p.m. On sunflower, the mature leaf developmental stage corresponds to a dark green leaf, assumed to be experiencing its highest photosynthetic rate and having recently reached its maximum size (Andrianasolo *et al.*, 2016). More precisely, the mature leaf is positioned at three-fifths of the plant (leaf rank $n = 16.4 \pm 1.9$ SD) (Blanchet *et al.*, 2018). The selected leaf to harvest for the molecular analysis was the leaf above the mature one (leaf rank $n + 1$).

4 Transcriptome analysis

4.1 RNA extraction and sequencing

Protocols used for the transcriptomic analysis have been detailed in Badouin *et al.* (2017). Briefly, grinding was performed using a ZM200 grinder (Retsch, Haan, Germany) with a 0.5-mm sieve. Total RNA was extracted using QIAzol Lysis Reagent following the manufacturer's instructions (Qiagen, Dusseldorf, Germany). The RNA quality was checked by electrophoresis on an agarose gel and quality and quantity were assessed using the Agilent RNA 6000 nano kit (Agilent, Santa Clara, CA, USA). Sequencing was performed on the Illumina HiSeq 2000 by DNAVision (Charleroi, Belgium) as paired-end libraries (2×100 bp, oriented) using the TruSeq sample preparation kit (Illumina San Diego, CA, USA) according to manufacturer's instructions.
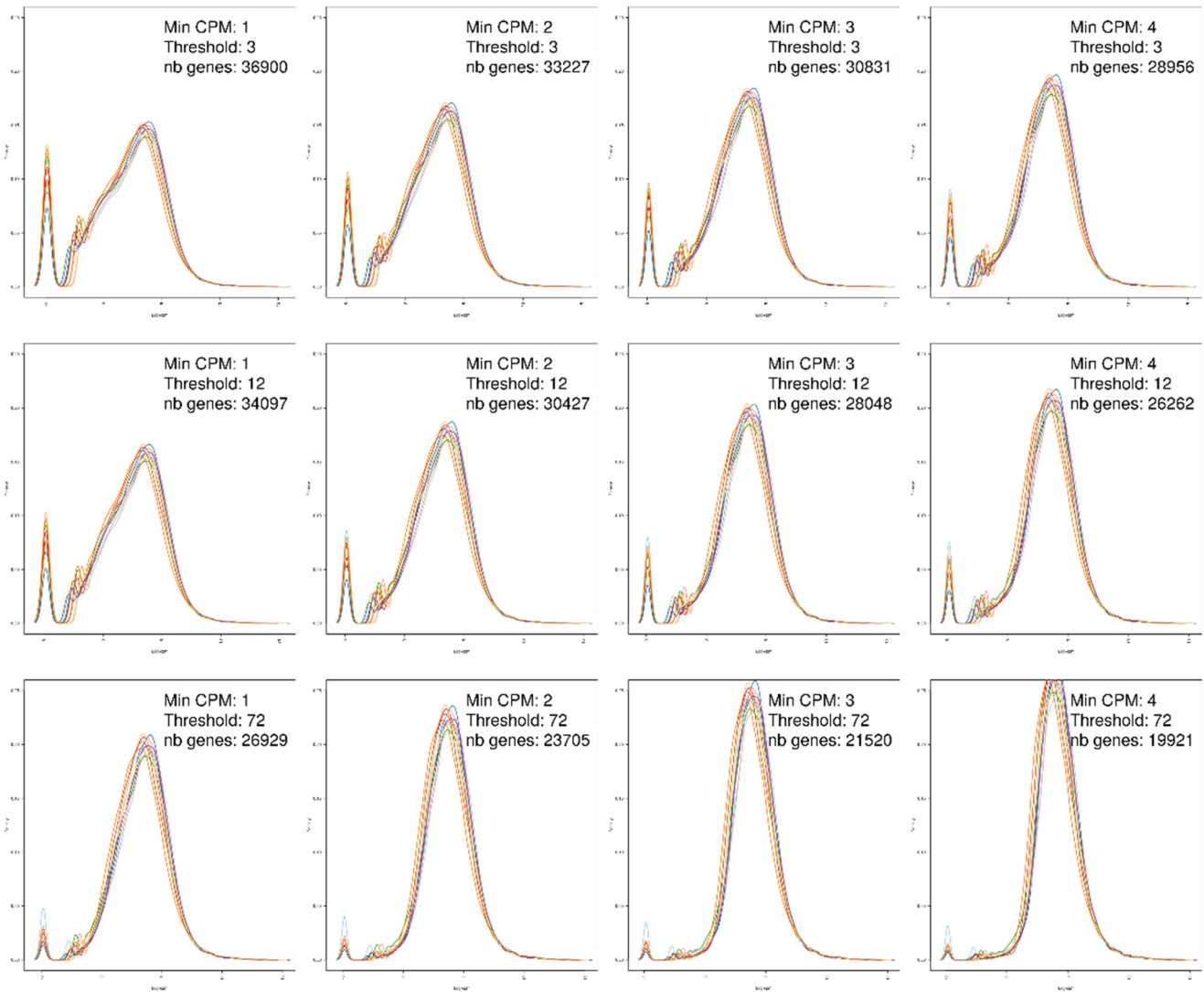
Fig. 1. Raw counts distributions in log counts per million in all samples after each filter parameter combination. Threshold = number of samples.

4.2 Reads mapping and expression measurements

RNAseq read pairs were mapped on the sunflower genome HanXRQv1.0 (Badouin *et al.*, 2017) using the glint software with parameters set as follows: matches ≥30 nucleotides, with ≤4 mismatches, no gap allowed, only best-scoring hits taken into account (glint mappe – mmis 4 – lmin 30 – mate-dist 10000 – best-score – no-lc-filtering). Ambiguous matches (same best score) were removed. Pair counts were performed at the exon level (taking into account the strand for stranded libraries), and counts were then propagated at the level of corresponding transcripts.

Given the two missing plants that died during the experiment, we finally were able to analyse 142 samples. The transcriptomic study was performed with the EdgeR package version 3.16.5 on R version 3.3.3 (Another Canoe) with the Counts Per Millions (CPM) function.

4.3 Filtering lowly expressed genes and normalization

Hierarchical cluster analysis revealed that the "SF326 ctrl R3" and "SF009 stress R1" samples belonged to different clusters and were removed from the analysis, reducing the sample number to 140.

To determine which genes have sufficiently large counts to be retained in the statistical analysis, usual practice in edgeR package is to use the *filterByExpr* function. However, given the specific design of our dataset, we wanted to be able to identify treatment:genotype specific expressions that would be eliminated by *filterByExpr* (three samples with expression among 140). For this, we replaced this step with an *ad hoc* method. This consisted in keeping genes with a minimum of CPM in at least a fixed number of samples. Several values for these two parameters were tested: 1, 2, 3 or 4 minimum of CPM and in at least 3, 12 or 72 samples detailed in Figure 1.

Table 1. Number of genes studied and retained after filtering and descriptive statistics of the 142 library sizes.

	Number of genes		Library size (mapped counts)				
	Total	With counts	Min	Median	Max	Mean	SD
Before filtering	58 050	39 319	2 796 719	6 922 010	13 765 614	7 397 044	2 549 417
After filtering	30 831	30 831	2 786 928	6 894 423	13 702 329	7 370 699	2 540 113

SD: standard deviation.

The log counts per million distribution should tend toward normality after filtering. Given the high amount of samples (140) and the high amount of lowly expressed genes, normality could not be achieved. The set of parameters that were deemed to have the best balance between normality was where genes were considered expressed if there were at least three libraries with at least three CPM.

Thirty thousand eight hundred and thirty-one genes were found with at least three CPM in three libraries. Normalization by the method of trimmed mean of M-values (TMM) was performed using the *calcNormFactors* function of edgeR package as in the user guide. Table 1 describes the library sizes and number of genes studied before and after filtering.

5 Data records

5.1 13HP02_count.csv

This file contains raw-counts for each genotype and their three biological replicates (in columns).

5.2 13HP02_count_after_filtering.csv

This file contains filtered and normalized counts for each genotype and their three biological replicates (in columns).

References

Andrianasolo FN, Casadebaig P, Langlade N, Debaeke P, Maury P. 2016. Effects of plant growth stage and leaf ageing on the response of transpiration and photosynthesis to water deficit in sunflower. *Funct Plant Biol* 43: 797–805.

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.

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.

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

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 Environ* 36: 2175–2189.

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. *PLoS One* 7: e45249.

Schneiter AA, Miller JF. 1981. Description of sunflower growth stages. *Crop Sci*. Available from https://doi.org/10.2135/cropsci1981.0011183x002100060024x.

USDA F. 2019. Oilseeds: world markets and trade. USDA F.