Data in Brief 38 (2021) 107392

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

RNA sequencing data for responses to drought stress and/or clubroot infection in developing seeds of *Brassica napus*

Grégoire Bianchetti[a], Vanessa Clouet[a], Fabrice Legeai[a], Cécile Baron[a], Kévin Gazengel[a], Aurélien Carrillo[a], Maria J. Manzanares-Dauleux[a], Julia Buitink[b,1], Nathalie Nesi[a,1,*]

[a] IGEPP, INRAE, Institut Agro, Université de Rennes 1, BP35327, Le Rheu 35650, France
[b] IRHS, INRAE, Institut Agro, Université d'Angers, SFR4207 QuaSaV, Beaucouzé 49070, France

ARTICLE INFO

Article history:
Received 6 August 2021
Revised 14 September 2021
Accepted 16 September 2021
Available online 20 September 2021

Keywords:
Oilseed rape
Seed development
Drought stress
Clubroot infection
Plasmodiophora brassicae
RNA-seq

ABSTRACT

Oilseed rape (*Brassica napus* L.) is the third largest oil crop worldwide. Like other crops, oilseed rape faces unfavorable environmental conditions resulting from multiple and combined actions of abiotic and biotic constraints that occur throughout the growing season. In particular drought severely reduces seed yield but also impacts seed quality in oilseed rape. In addition, clubroot disease, caused by the pathogen *Plasmodiophora brassicae*, limits the yield of the oilseed rape crops grown in infected areas. Clubroot induces swellings or galls on the roots that decrease the flow of water and nutrients within the plant. Furthermore, combinations of different stresses lead to complex plant responses that can not be predicted by the simple addition of individual stress responses. Indeed, an abiotic constraint can either reduce or stimulate the plant response to a pathogen or pest. Transcriptome datasets from different conditions are key resources to improve our knowledge of environmental stress-resistance mechanisms in plant organs. Here, we describe a RNA-seq dataset consisting of 72 samples of immature *B. napus* seeds from plants grown either under drought, infected

* Corresponding address at: IGEPP, INRAE, Institut Agro, Université de Rennes 1, BP35327, 35650 Le Rheu, France
E-mail address: nathalie.nesi@inrae.fr (N. Nesi).
[1] Co-last authors

https://doi.org/10.1016/j.dib.2021.107392

with *P. brassicae*, or a combination of both stresses. A total of 67.6 Gb of transcriptome paired-end reads were filtered, mapped onto the *B. napus* reference genome Darmor-*bzh* and used for identification of differentially expressed genes and gene ontology enrichment. The raw reads are available under accession PRJNA738318 at NCBI Sequence Read Archive (SRA) repository. The dataset is a resource for the scientific community exploring seed plasticity.

Specifications Table

Subject	Agricultural and Biological Sciences
Specific subject area	Omics: Transcriptomics Plant Science: Plant Physiology
Type of data	Tables Figures
How data were acquired	RNA samples were sent to the BGI, Hong Kong (https://www.bgi.com), for library construction using the QuantSeq 3′ mRNA-Seq Library Prep Kit REV (Lexogen) and sequencing using an Illumina HiSeq4000 platform that generated an average 9.38 M of 100 bp paired-end reads per sample.
Data format	Filtered raw reads (FASTQ) Analysed RNA-seq data files (counts and DEG lists)
Parameters for data collection	Total RNAs were extracted from seeds of *Brassica napus* (cv. "Express"). Seed-bearing mother plants were grown under (1) standard conditions, (2) water shortage, (3) clubroot infection after inoculation at 7 days after germination with *Plasmodiophora brassicae* (isolate eH), or (4) combination of both clubroot infection and water shortage. Seeds were collected from plants at 8 timepoints during seed development that corresponded to the times T1-T6 and T8-T9 described by Bianchetti et al. [1].
Description of data collection	RNA-seq dataset was obtained from paired-end sequencing of cDNA libraries with 100 bp reads. Processing of RNA-seq data included (1) raw reads filtering, (2) paired-reads mapping onto the *B. napus* reference genome Darmor-*bzh* [2], (3) bioinformatic analysis for differential gene expression, and (4) gene set enrichment analyses.
Data source location	Institution: Institute of Genetics, Environment and Plant Protection (IGEPP), INRAE, Institut Agro, Univ. Rennes 1 City/Town/Region: 35650 Le Rheu Country: France Latitude and longitude for collected samples/data: 48°06'37.1"N, 1°47'46.3"W
Data accessibility	Repository name: NCBI Sequence Read Archive (SRA) Data identification number: BioProject ID PRJNA738318 Direct URL to data: http://www.ncbi.nlm.nih.gov/bioproject/738318 https://dataview.ncbi.nlm.nih.gov/object/PRJNA738318?reviewer=tdei8sk5o4a78ihcm3t0r1dq92 https://data.inrae.fr/dataset.xhtml?persistentId=doi:10.15454/9OXYM5

Value of the Data

- The present study reports a transcriptomic dataset from developing seeds of *B. napus* harvested from plants grown under four environmental conditions (standard watering/drought with or without inoculation by *P. brassicae*, and inoculation by *P. brassicae* alone).

- These data are useful for the scientific community working on the impact of single or combined abiotic and biotic constraints on seed quality.
- Presented RNA-seq data provides a key resource to unravel the transcriptomic programs that support the seed development under standard conditions or environmental stresses. The data will provide candidate genes including regulatory factors to predict seed quality under various conditions.
- Further analyses of the transcriptome dataset described here will provide useful information to accelerate the development of stress-resistant oilseed rape cultivars.

1. Data Description

This paper describes a transcriptomic dataset of developing seeds of *Brassica napus* (cv. "Express") that were produced under drought stress and/or *Plasmodiophora brassicae* infection. Among the abiotic stress, drought is one of the main threats to the crop productivity [3] and genetic improvement of oilseed rape with respect to drought tolerance requires the exploration of drought-response genes during seed development in particular. On the other hand, clubroot is a major soil-borne disease of Brassicaceae species [4] that is caused by the telluric biotroph protist *P. brassicae*. It induces the formation of swellings or galls on the root, ultimately leading to the premature death of the plant. In addition, combinations of multiple stresses have been shown to trigger complex plant responses at the transcriptome, cellular and physiological levels that significantly differ from the responses to individual stresses [5]. This prompted us to design an experiment to generate RNA-seq data from seeds of oilseed rape plants grown under combinations of abiotic and biotic constraints, resulting in the four treatments presented in Fig. 1 and hereafter: (1) standard watering and no *P. brassicae* inoculation (C, control); (2) water shortage (WS); (3) clubroot infection by inoculating 7-day old seedlings with resting spores of *P. brassicae* (Pb); and (4) a combination of clubroot and water shortage (Pb+WS). Immature seeds were collected at different timepoints during seed development resulting in a total of 72 samples for paired-end RNA-sequencing. Table 1 displays the quality of the transcriptome dataset. An average of 9.39 M paired reads (2 $\times$ 100 bp each) was obtained per sample with good quality scores (Phred scores $\approx$ 33 or 26 for reads 1 or reads 2, respectively) of which 82.6% in average were single mapped onto the reference genome of *B. napus* (Darmor-*bzh*-v4; [2]) and 64.3% were assigned to an annotated gene (Darmor-*bzh*-v5; [2]). After mapping, the quantification led to a set of 42,467 genes for which the transcript levels were above the minimum transcript level threshold of 1 CPM (count per million) (**Table S1**). From this count table, the 72 samples were hierarchically clustered based on their Euclidean distance using the Ward's criterion to validate reproducibility of replicates (Fig. 2) and 6564 differentially expressed genes (DEGs; FDR $<$ 0.05 – **Table S2**) were identified between control and stress conditions at the timepoint T5 of seed development (Fig. 3). The Venn diagram presented in Fig. 3**A** points out a set of 870 genes that are commonly deregulated in the three treatments (WS, Pb, Pb+WS) compared to the control samples, of which 655 are down- and 215 are up-regulated in the stressed conditions (Fig. 3**B**). Fig. 4 depicts the GO terms enriched within the set of 870 common DEGs.

2. Experimental Design, Materials and Methods

2.1. Plant material and growth conditions

Oilseed rape plants (cv. "Express") were cultivated in 2017-2018 at INRAE Le Rheu, France (48°06'37.1"N, 1°47'46.3"W) as described by Bianchetti et al. [1]. This cultivar shown a moderate resistance to *P. brassicae* (M. Manzanares-Dauleux, personal communication). Briefly, plants were grown using a semi-controlled system that mimics field conditions and allows working with a reconstructed canopy while fine monitoring applied stress or constraints during the plant cycle.

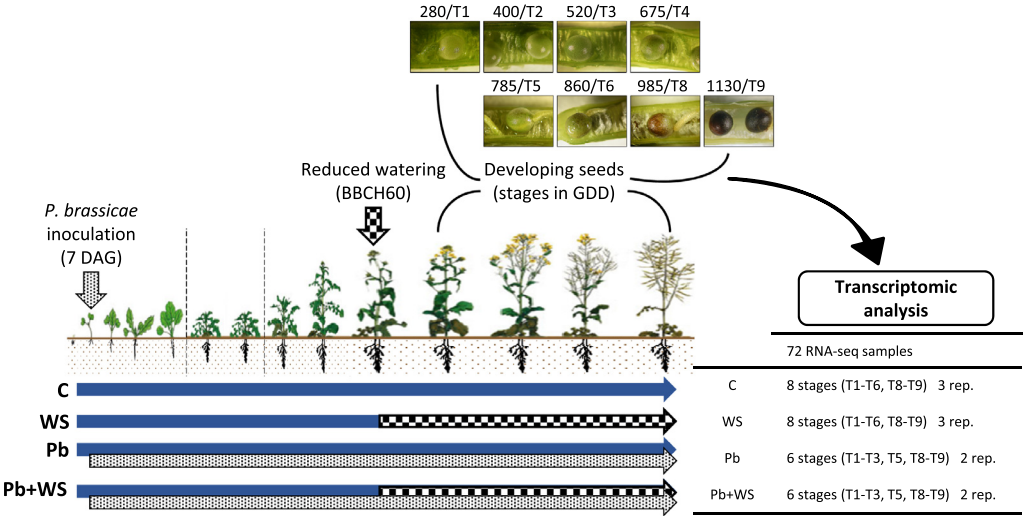

Fig. 1. Experimental design. Seeds were collected from oilseed rape plants (cv. "Express") grown under a sheltered tunnel and submitted to four different treatments. Developing seeds were harvested at 8 different timepoints (T1-T6; T8-T9 as described by Bianchetti et al. [1] and used for the RNA sequencing assays. C, control; WS, water shortage (soil water potential was maintained at -600 mbar); Pb, *P. brassicae* inoculation; Pb+WS, *P. brassicae* inoculation and water shortage; DAG, days after germination; BBCH60, start of flowering; rep, biological replicate; GDD, growing degree-days calculated in base 0 °C from BBCH60.

Table 1
Summary of RNA-seq samples with numbers of clean reads provided by the BGI sequencing platform and used for RNA-seq mapping and percentages of read pairs that were single mapped and assigned to the annotated genes of the *Brassica napus* reference genome Darmor-*bzh* [2]. Samples are identified as 'Treatment_Seed stage_Replicate' with C, control; WS, water shortage; Pb, *P. brassicae* inoculation; Pb+WS, *P. brassicae* inoculation and water shortage; T1-T6, T8-T9, seed stages (refer to Fig. 1); 1, 2 or 3, replicate.

Sample ID	Number of clean reads	% single mapped read pairs	% assigned to annotated genes
C_T1_1	20,830,752	80.6%	62.4%
C_T1_2	20,458,468	81.7%	63.5%
C_T1_3	21,697,122	81.4%	62.6%
WS_T1_1	14,665,982	80.8%	62.5%
WS_T1_2	19,673,244	82.0%	62.9%
WS_T1_3	22,512,296	82.9%	63.8%
Pb_T1_1	20,962,980	81.4%	62.8%
Pb_T1_2	20,511,396	82.7%	63.8%
Pb+WS_T1_1	22,445,538	82.1%	65.1%
Pb+WS_T1_2	24,682,642	82.7%	64.6%
C_T2_1	26,878,272	82.9%	63.8%
C_T2_2	25,694,410	84.1%	63.8%
C_T2_3	12,333,574	83.4%	64.8%
WS_T2_1	10,458,254	83.8%	64.3%
WS_T2_2	13,251,126	85.1%	64.9%
WS_T2_3	16,573,066	84.8%	64.8%
Pb_T2_1	14,900,892	82.9%	63.7%
Pb_T2_2	14,468,604	83.9%	63.2%
Pb+WS_T2_1	13,701,444	85.3%	65.4%
Pb+WS_T2_2	15,404,956	84.2%	63.6%
C_T3_1	16,218,352	86.3%	62.6%
C_T3_2	17,893,296	86.5%	61.0%
C_T3_3	14,300,972	85.3%	61.7%
WS_T3_1	23,926,946	84.8%	60.4%
WS_T3_2	19,331,180	83.4%	57.4%
WS_T3_3	22,604,484	84.6%	59.9%
Pb_T3_1	20,437,896	84.5%	61.1%
Pb_T3_2	22,261,432	84.0%	59.7%
Pb+WS_T3_1	23,000,604	84.1%	59.9%
Pb+WS_T3_2	24,110,262	86.0%	62.1%
C_T4_1	21,108,776	83.9%	62.9%
C_T4_2	21,564,872	83.7%	63.9%
C_T4_3	23,941,612	82.8%	61.4%
WS_T4_1	25,262,418	83.4%	63.5%
WS_T4_2	20,978,824	83.0%	62.8%
WS_T4_3	22,467,070	84.2%	63.3%
C_T5_1	22,503,844	83.3%	64.5%
C_T5_2	19,816,636	83.4%	64.9%
C_T5_3	21,293,156	82.5%	63.5%
WS_T5_1	20,078,156	84.6%	64.5%
WS_T5_2	16,067,004	84.2%	64.1%
WS_T5_3	25,120,386	83.8%	64.4%
Pb_T5_1	21,753,704	83.1%	65.2%
Pb_T5_2	19,597,922	85.5%	67.7%
Pb+WS_T5_1	18,695,306	84.1%	65.4%
Pb+WS_T5_2	21,800,314	83.1%	64.8%
C_T6_1	24,817,606	84.0%	66.6%
C_T6_2	13,214,278	84.6%	66.9%
C_T6_3	13,784,832	83.3%	65.5%
WS_T6_1	14,479,538	83.9%	65.9%
WS_T6_2	14,926,262	83.3%	63.8%
WS_T6_3	16,172,958	81.3%	63.3%
C_T8_1	13,598,468	83.7%	65.6%
C_T8_2	13,120,628	84.5%	65.7%

(*continued on next page*)

Table 1 (*continued*)

Sample ID	Number of clean reads	% single mapped read pairs	% assigned to annotated genes
C_T8_3	11,069,704	83.0%	65.0%
WS_T8_1	13,372,530	82.9%	64.4%
WS_T8_2	16,283,610	84.2%	65.9%
WS_T8_3	16,244,272	81.5%	63.7%
Pb_T8_1	15,266,498	84.2%	66.3%
Pb_T8_2	16,643,082	84.9%	67.1%
Pb+WS_T8_1	15,596,326	83.2%	65.5%
Pb+WS_T8_2	17,836,864	84.1%	66.3%
C_T9_1	23,211,022	82.0%	64.0%
C_T9_2	21,258,822	83.2%	64.4%
C_T9_3	18,846,012	83.2%	65.2%
WS_T9_1	22,684,746	84.0%	65.5%
WS_T9_2	21,236,922	85.1%	66.9%
WS_T9_3	18,662,922	84.1%	66.0%
Pb_T9_1	19,317,378	83.9%	66.4%
Pb_T9_2	18,819,888	83.2%	65.7%
Pb+WS_T9_1	15,904,268	84.4%	66.4%
Pb+WS_T9_2	13,200,142	83.5%	66.7%

The current experiment included 10 tanks of ~1 m^3 filled with a mixture of sand, topsoil and Irish peat (5/3.4/1.6:v/v/v) to facilitate water drain where pre-germinated seeds were regularly spaced at a density of 42 plants/m^2. Plants were grown under four different treatments: (1) the control (C) condition, where standard watering was applied to maintain the water potential of the substrate above -200 mbar; (2) the water shortage (WS) treatment, where watering was monitored to maintain the water potential of the substrate around -600 mbar from the onset of flowering onwards; (3) the clubroot inoculation (Pb) treatment, for which 7-day old seedlings were inoculated with ml of a resting spore suspension (10^6 spores.ml^{-1}) of *P. brassicae* (eH isolate) as described by Manzanares-Dauleux et al. [6]; and (4) the Pb+WS that combined both clubroot infection and water shortage. The experimental design included three replicated tanks for the C and WS treatments and two for the Pb and Pb+WS treatments. Clubroot symptoms were checked just before the application of the WS treatment and at the final harvest as described previously [1] (**Fig. S1**).

2.2. Seed sampling

For each sample, seeds were harvested on the main inflorescences only and were collected as a mix from 3 to 5 of the 20 central plants of each tank to increase the plot representativeness and limit the border effects. Developing seeds were collected from pods that were tagged on the day of flowering (*i.e.*, BBCH60 developmental stage according to Lancashire et al. [7]). A total of 8 timepoints were considered in the current study. They corresponded to timepoints T1-T6 and T8-T9 described previously [1] and were harvested at thermal times such as following: T1, 280 GDD (growing degree-days calculated from the start of flowering with a base temperature of 0°C); T2, 400; T3, 520; T4, 675; T5, 785; T6, 860; T8, 985; T9, 1130. Immature seeds were carefully removed from siliques, instantly frozen into liquid nitrogen and stored at -80°C before RNA extraction.

2.3. RNA isolation and sequencing

Total RNAs were extracted from immature seeds that were collected at 8 stages during seed development under C and WS conditions and 6 stages during seed development under Pb and

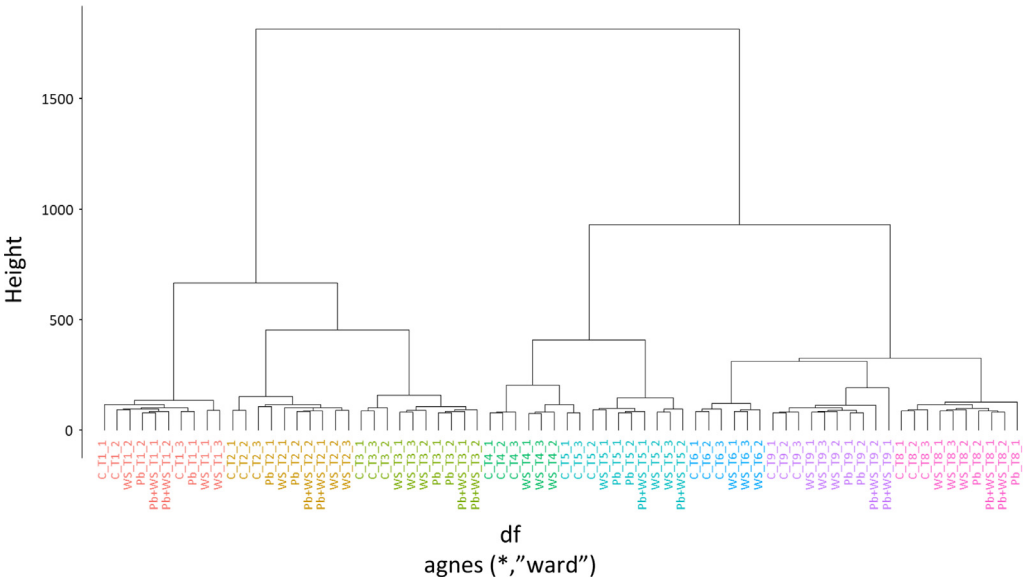

Fig. 2. Cluster dendogram of the RNA-seq dataset to validate reproducibility of replicates. An agglomerative nesting hierarchical clustering (agnes) of the 72 samples was produced using the transcript levels of the 42,467 genes that were above the minimum transcript level threshold of 1 CPM (count per million). Samples are identified as 'Treatment_Seed stage_Replicate' and colored according to the corresponding seed developmental stage. Treatment: C, control; WS, water shortage; Pb, *P. brassicae* inoculation; Pb+WS, *P. brassicae* inoculation and water shortage. Seed stage: T1-T6; T8-T9 (refer to Fig. 1). Replicate: 1, 2 or 3.

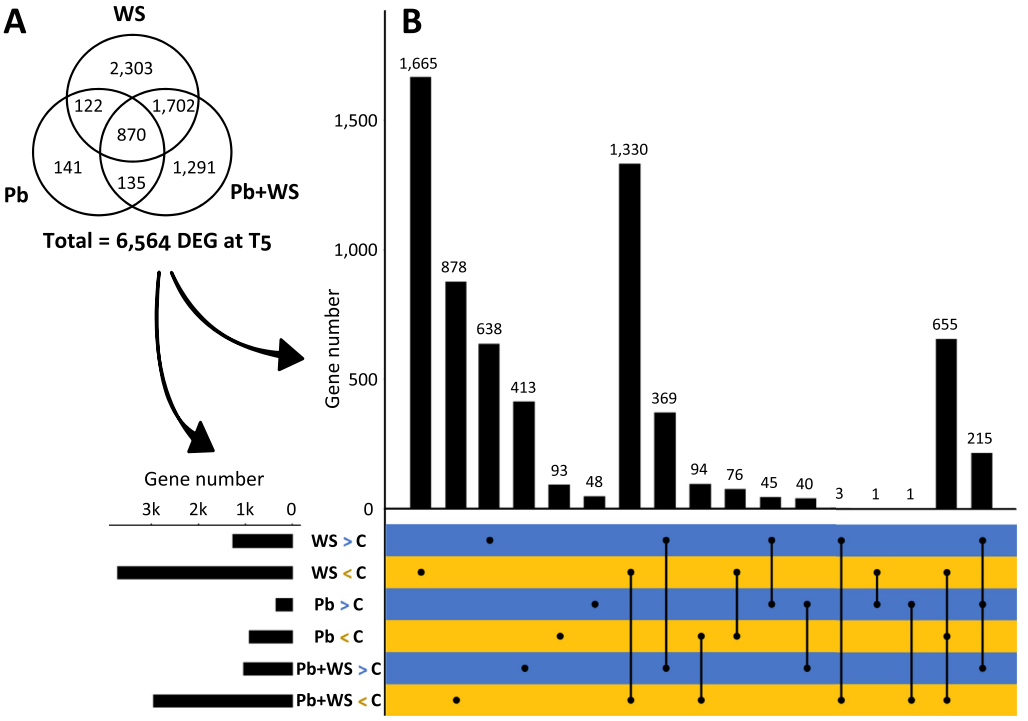

Fig. 3. Distribution of the differentially expressed genes (DEGs) in seeds of *B. napus* according to the treatment at timepoint T5 (785 GDD) of seed development. **A.** Venn diagram with the 6,564 DEGs (FDR < 0.05) between control and stress conditions. The DEGs were selected using the AskoR R library [21] starting with the transcript levels of the 42,467 genes that were above the minimum transcript level threshold of 1 CPM (count per million). Significant differences (likelihood ratio test) were determined using the log2 (CPM+1) values with a minimum fold change of 1. **B.** Up (blue) or down (yellow) DEGs in stressed treatments compared to standard conditions. C, control; WS, water shortage; Pb, *P. brassicae* inoculation; Pb+WS, *P. brassicae* inoculation and water shortage. GDD, growing degree-days (see Fig. 1). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

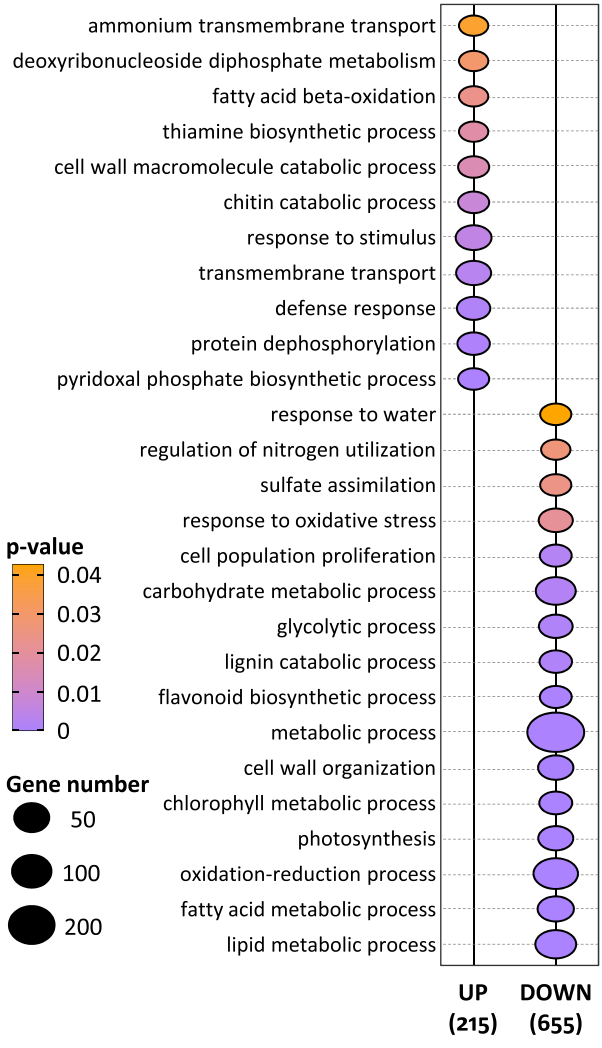

Fig. 4. Dot plot of overrepresented biological functions in the set of the 870 DEGs identified for all stresses in comparison to control samples at timepoint T5 of seed development. Enriched GO terms were identified using the topGO R package (v2.40.0; [22]) with the GO available for the reference genome Darmor-*bzh*-v5 of *B. napus*. Dot size stands for the number of genes and colors for the *p*-value.

Pb+WS conditions. Samples were ground into liquid nitrogen and RNA extractions were performed using the NucleoSpin® RNA Plus Kit according to the supplier's instructions (Macherey-Nagel GmbH & Co., Düren, Germany). To remove any genomic DNA contamination, 4 µg of RNA were treated with 4 U of DNAse I RNase free (Thermo Fisher Scientific, Waltham, MA, USA) in the presence of 80 U of RiboLock RNase inhibitor (Thermo Fisher Scientific). Finally, the RNA preparations were purified using the NucleoSpin® RNA Clean-up kit (Macherey-Nagel). RNA quality was checked using a NanoDrop® ND-1000 UV-Vis Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) and a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) (mean values for OD260/280 = 2.2, OD260/230 = 1.9, RIN = 8.7, 28S/18S = 2.1, baseline smooth). All samples were sent to the Beijing Genomics Institute (BGI, Hong Kong; https://www.bgi.com). The QuantSeq 3′ mRNA-Seq Library Prep Kit REV (Lexogen GmbH, Vi-

enna, Austria ; [8]) was used according to the manufacturer's instruction to generate compatible libraries for Illumina sequencing with a resulting average size of 250 bp each. Libraries of the 72 samples were sequenced using the Illumina HiSeq4000 sequencing platform of the BGI, yielding an average 18.8 M of 100 bp reads per sample.

2.4. RNA seq data analyses

Processing of the RNA-seq data was carried out using the adapted nf-core RNA-seq pipeline version v1.4.2 ([9]; https://github.com/nf-core/rnaseq). Briefly, the pipeline was based on Nextflow v19.07.0 [10] and processed data using the default options.. Read mapping onto the reference genome Darmor-*bzh*-v4 [2] was performed with Hisat2 v2.1.0 [11], and transcript quantification with featureCounts v1.6.4 [12] according to the Darmor-*bzh*-v5 gene annotation [2]. The pipeline outpout files were plotted and summarized with multiQC tools v1.7 [13]. For gene expression analyses, we used the AskoR R library [14] with the following steps: (1) data filtering, (2) data normalization, (3) visualization of correlations, and (4) differential expression analysis. GO enrichment analyses were run with topGO R package (v2.40.0; [15]) using the default parameters and Darmor-*bzh*-v5 GO annotation [2].

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

CRediT Author Statement

Grégoire Bianchetti: Investigation, Data curation, Visualization, Writing – original draft; **Vanessa Clouet:** Investigation, Data curation, Visualization, Writing – original draft; **Fabrice Legeai:** Investigation, Data curation, Visualization; **Cécile Baron:** Investigation; **Kévin Gazengel:** Methodology; **Aurélien Carrillo:** Investigation; **Maria J. Manzanares-Dauleux:** Methodology; **Julia Buitink:** Supervision, Conceptualization, Writing – review & editing; **Nathalie Nesi:** Supervision, Conceptualization, Writing – review & editing.

Acknowledgments

We thank all our colleagues from IGEPP, INRAE Le Rheu (France), especially J.P. Constantin, L. Charlon, F. Le Tertre and P. Rolland for supplying excellent greenhouse services, P. Glory and C. Lariagon for assistance with *P. brassicae* infection, and I. Glais, A-S. Grenier and M. Boudet for help with RNA-seq data management. We are grateful to the BrACySol biological resource center (INRAE Ploudaniel, France) for kindly providing us with seeds of Express. The BGI sequencing platform (https://www.bgi.com) and the GenOuest bioinformatics platform (https://www.genouest.org/) are acknowledged for access to their facilities. This research was funded through an ANR grant (ANR-14-JFAC-007-01) in the framework of the SYBRACLIM project (FACCE-JPI-ERA-NET+ CLIMATE SMART AGRICULTURE project call) and a grant from INRAE to J. Buitink and N. Nesi (INRAE-BAP-IB2017-SQUAL). G. Bianchetti was supported by INRAE and the Région de Bretagne through a 3 years PhD grant.

Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.dib.2021.107392.

References

[1] G. Bianchetti, C. Baron, A. Carrillo, S. Berardocco, N. Marnet, M.-H. Wagner, D. Demilly, S. Ducournau, M.J. Manzanares-Dauleux, F.Le Cahérec, J. Buitink, N. Nesi, Dataset for the metabolic and physiological characterization of seeds from oilseed rape (*Brassica napus L.*) plants grown under single or combined effects of drought and clubroot pathogen *Plasmodiophora brassicae*, Data Br. (2016) 10724, doi:10.1016/j.dib.2021.107247.

[2] B. Chalhoub, F. Denoeud, S. Liu, I.A.P. Parkin, H. Tang, et al., Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome, Science 345 (2014) 950–953, doi:10.1126/science.1253435.

[3] A.E.L. Sabagh, A. Hossain, C. Barutçular, M.S. Islam, D. Ratnasekera, N. Kumar, Drought and salinity stress management for higher and sustainable canola (Brassica napus L.) production: a critical review, Aust. J. Crop Sci. 13 (2019) 88–97, doi:10.21475/ajcs.19.13.01.p1284.

[4] G.R. Dixon, The occurrence and economic impact of *Plasmodiophora brassicae* and clubroot disease, J. Plant Growth Regul. 28 (2009) 194–202, doi:10.1007/s00344-009-9090-y.

[5] S. Rasmussen, P. Barah, M.C. Suarez-Rodriguez, S. Bressendorff, P. Friis, P. Costantino, A.M. Bones, H.B. Nielsen, J. Mundy, Transcriptome responses to combinations of stresses in Arabidopsis, Plant Physiol. 161 (2013) 1783–1794, doi:10.1104/pp.112.210773.

[6] M.J. Manzanares-Dauleux, I. Divaret, F. Baron, G. Thomas, Evaluation of French *Brassica oleracea* landraces for resistance to *Plasmodiophora brassicae*, Euphytica 113 (2000) 211–218.

[7] P.D. Lancashire, H. Bleiholder, T.V.D. Boom, P. Langelüddeke, R. Stauss, E. Weber, A. Witzenberger, A uniform decimal code for growth stages of crops and weeds, Ann. Appl. Biol. 119 (1991) 561–601, doi:10.1111/j.1744-7348.1991.tb04895.x.

[8] P. Moll, M. Ante, A. Seitz, T. Reda, QuantSeq 3′ mRNA sequencing for RNA quantification, Nat. Methods 11 (2014), doi:10.1038/nmeth.f.376.

[9] P.A. Ewels, A. Peltzer, S. Fillinger, H. Patel, J. Alneberg, A. Wilm, M.U. Garcia, P. Di Tommaso, S. Nahnsen, The nf-core framework for community-curated bioinformatics pipelines, Nat. Biotechnol. 38 (2020) 276–278, doi:10.1038/s41587-020-0439-x.

[10] P. Di Tommaso, M. Chatzou, E.W. Floden, P.P. Barja, E. Palumbo, C. Notredame, Nextflow enables reproducible computational workflows, Nat. Biotechnol 11 (2017) 316–319, doi:10.1038/nbt.3820.

[11] D. Kim, J.M. Paggi, C. Park, C. Bennett, S.L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype, Nat. Biotechnol. 37 (2019) 907–915, doi:10.1038/s41587-019-0201-4.

[12] Y. Liao, G.K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, Bioinformatics 30 (2014) 923–930, doi:10.1093/bioinformatics/btt656.

[13] 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.

[14] S. Alves Carvalho, K. Gazengel, S. Robin, S. Masanelli, S. Daval, F. Legeai, AskoR: a pipeline for the analysis of gene expression data using edgeR, https://github.com/askomics/askoR. Accessed June 1, 2021.

[15] A. Alexa, J. Rahnenführer, T. Lengauer, Improved scoring of functional groups from gene expression data by decorrelating GO graph structure, Bioinformatics 22 (2006) 1600–1607, doi:10.1093/bioinformatics/btl140.