Data in Brief 51 (2023) 109784

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Unraveling the dataset transcriptomic response of *Hydrangea macrophylla* stem to mechanical stimulation: *De novo* assembly and functional annotation

Béra Ley-Ngardigal[a,b], Sandra Pelletier[a], Vincent Guérin[a], Lydie Huché-Thélier[a], Nathalie Brouard[a], Hanaé Roman[b], Nathalie Leduc[a,*]

[a] *Institut Agro, INRAE, IRHS, SFR QUASAV, University Angers, Angers F-49000, France*
[b] *Hortensia France, La Bodinière, Rives-du-Loir-en-Anjou 49140, France*

ARTICLE INFO

Article history:
Received 1 August 2023
Revised 2 November 2023
Accepted 3 November 2023
Available online 7 November 2023

Dataset link: Unraveling the dataset transcriptomic response of Hydrangea macrophylla stem to mechanical stimulation: de novo assembly and functional annotation (Reference data)

Keywords:
Transcriptional analysis
RNA sequencing
Alternative to PGR
Ornamental species
Plant compactness
Mechanical stress
Stem bending
Bi-lateral stem flexures

ABSTRACT

A crucial attribute of potted ornamental plants is compactness characterized by well branched plants with rather short stems bearing numerous flowers. To gain plant compactness, producers use plant growth regulators (PGRs), in particular growth retardants during culture. However, due to their negative environmental impacts, growth retardants are progressively withdrawn from the market. As a response, eco-friendly alternative methods to chemicals need to be developed. One method consists in mimicking mechanical stimulation (MS) imposed by wind on plants which causes reduction in stem elongation, an increase in stem diameter and an increase in branching, all contributing to plant compactness. So far, few plant species were studied under MS and little is known on molecular response mechanisms to MS. This first transcriptomic data after MS in *Hydrangea macrophylla* will contribute unravelling how plants respond to mechanical stimuli. RNAseq data were obtained from total mRNA of stems collected 15 min before MS and 1, 3, 24 and 72 h after MS treatment. RNA from non-MS treated plants were used as control. MS treatment consisted in 12 consecutive bend-

* Corresponding author.
E-mail address: nathalie.leduc@univ-angers.fr (N. Leduc).

https://doi.org/10.1016/j.dib.2023.109784

ings (i.e. 6 forth and 6 back) applied at 9 a.m. during 1 h and for a single day. From RNAseq data a *de novo* assembly of the transcriptome was produced and 78,398 transcripts functionally annotated. These transcriptomic data also contribute to a better knowledge of how outdoor crop respond to the increasing frequency of strong harmful winds under climate change.

Specifications Table

Subject	Agricultural and Biological Sciences
Specific subject area	Omics: Transcriptomics Plant Science: Plant Physiology
Data format	RNA-Seq raw data (FASTQ format) Analysed RNA-seq data files: list of annotated transcripts (.fasta format) and read counts (.txt format)
Type of data	Tables Figures
Data collection	Mechanical stimulation (MS) of *Hydrangea macrophylla* stems consisted in 12 consecutive bendings applied at 9 a.m. during 1 h and for a single day. Stem samples were collected 15 min before MS and 1 h, 3 h, 24 h and 72 h after MS. Non-MS plants were used as control and collected at same time points. Total RNA sequencing was performed by Novogene using an Illumina NovaSeq 6000 platform. Transcriptome assembly was performed using Trinity v2.6.6, CORSET v4.6 and EvidentialGene pipeline. A quantitative assessment of completeness in terms of the expected gene content of the transcriptome was evaluated using BUSCO v3.0.2. Gene functional annotation was performed using 8 databases: NCBI non-redundant protein/nucleotide sequences, PFAM, PROSITE, UniProtKB/Swiss-Prot, KOG/COG, KEGG and GO.
Data source location	Institution: Growth chambers from the Institut de Recherche en Horticulture et Semences, INRAE City: Beaucouzé 49020 Country: France GPS coordinates: 47°28′37.7"N 0°36′42.1"W
Data accessibility	Raw, processed and analyzed data are submitted in Gene Expression Omnibus (GEO, accession number: GSE238196) Direct URL to data: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?&acc=GSE238196

1. Value of the Data

- The present study displays the first transcriptomic dataset from an ornamental species (*Hydrangea macrophylla* cv. 'Wudu'®) in response to mechanical stimulation.
- This dataset considerably increases the amount of information available for *Hydrangea* species, and is proving useful as a reference for other plant species subjected to abiotic (e.g. wind, rain, touch) or biotic (e.g. wounding) mechanical stress.
- Further analysis of the transcriptomic dataset described here will also provide useful information for developing mechanical stimulation markers to use in engineering and varietal selection.

2. Data Description

This study presents a *de novo* transcriptome sequencing and assembly of *Hydrangea macrophylla* cv. 'Wudu'® plants subjected to mechanical stimulation (MS). Transcriptomic data were

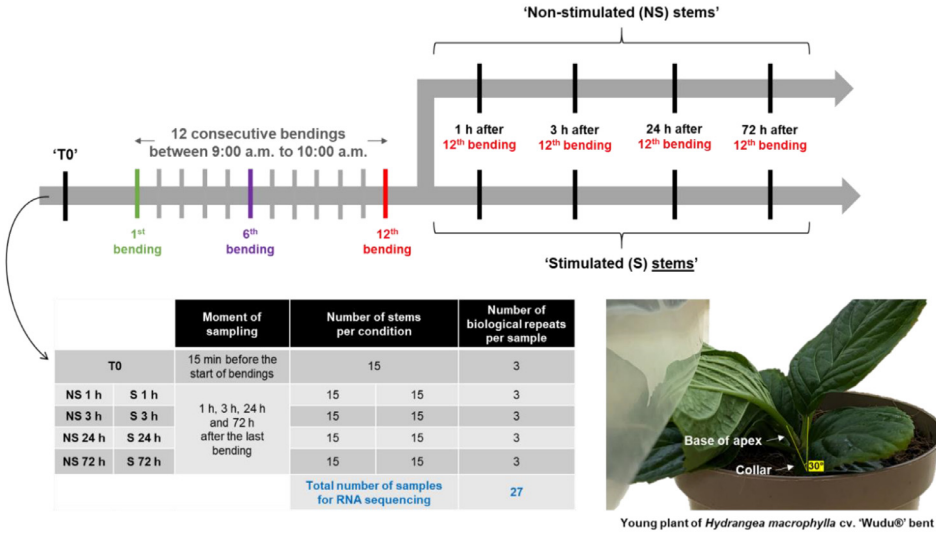

Fig. 1. Schematic experimental design.

Table 1
Descriptive summary of dataset after quality control.

Samples	Library	Raw reads	Raw bases	Clean reads	Clean bases	Error rate	Q20	Q30	GC %
T0R1	ERRA220002973-1a	20257075	6.08G	20046940	6.01G	0.03	96.73	91.59	45.04
T0R2	ERRA220002974-1a	21223987	6.37G	21005112	6.3G	0.03	96.88	91.95	45.32
T0R3	ERRA220002975-1a	24932564	7.48G	24677218	7.4G	0.03	96.70	91.54	44.70
S1HR1	ERRA220002976-1a	22871797	6.86G	22648948	6.79G	0.03	96.85	91.83	44.75
S1HR2	ERRA220002977-1a	23163991	6.95G	22921145	6.88G	0.03	96.59	91.32	44.85
S1HR3	ERRA220002978-1a	27097344	8.13G	26808375	8.04G	0.03	96.02	90.04	44.54
NS1HR1	ERRA220002979-1a	19653906	5.9G	19453297	5.84G	0.03	95.46	88.91	44.46
NS1HR2	ERRA220002980-1a	21204925	6.36G	20993690	6.3G	0.03	95.78	89.58	44.73
NS1HR3	ERRA220002981-1a	24423428	7.33G	24156831	7.25G	0.03	95.98	89.99	44.57
S3HR1	ERRA220002982-1a	24381328	7.31G	24135552	7.24G	0.03	95.73	89.47	44.67
S3HR2	ERRA220002983-1a	25385251	7.62G	25129181	7.54G	0.03	95.81	89.61	44.60
S3HR3	ERRA220002984-1a	23745142	7.12G	23444458	7.03G	0.03	95.85	89.71	44.44
NS3HR1	ERRA220002985-1a	24624545	7.39G	24391479	7.32G	0.03	95.92	89.86	44.53
NS3HR2	ERRA220002986-1a	20911557	6.27G	20684652	6.21G	0.03	95.77	89.57	44.89
NS3HR3	ERRA220002987-1a	26931716	8.08G	26627121	7.99G	0.03	96.68	91.51	44.38
S24HR1	ERRA220002988-1a	28896246	8.67G	28569379	8.57G	0.03	96.60	91.32	44.65
S24HR2	ERRA220002989-1a	22321251	6.7G	22093999	6.63G	0.03	96.56	91.25	44.38
S24HR3	ERRA220002990-1a	21839641	6.55G	21608623	6.48G	0.03	96.83	91.81	44.46
NS24HR1	ERRA220002991-1a	26959563	8.09G	26648131	7.99G	0.03	96.69	91.54	44.28
NS24HR2	ERRA220002992-1a	28425970	8.53G	28137092	8.44G	0.03	96.71	91.59	44.61
NS24HR3	ERRA220002993-1a	22261608	6.68G	22034695	6.61G	0.03	96.64	91.42	44.52
S72HR1	ERRA220002994-1a	23623775	7.09G	23372707	7.01G	0.03	96.66	91.45	44.12
S72HR2	ERRA220002995-1a	25917809	7.78G	25645385	7.69G	0.03	96.84	91.83	44.49
S72HR3	ERRA220002996-1a	21321198	6.4G	21063403	6.32G	0.03	95.88	89.81	44.17
NS72HR1	ERRA220002997-1a	31481028	9.44G	31141929	9.34G	0.03	96.74	91.64	44.49
NS72HR2	ERRA220002998-1a	28752692	8.63G	28462807	8.54G	0.03	96.82	91.81	44.55
NS72HR3	ERRA220002999-1a	23227177	6.97G	22984286	6.9G	0.03	96.70	91.55	44.84

obtained from whole stem collected 15 min before MS (T0) and 1, 3, 24 and 72 h after 12 consecutive bendings (S group) or without bendings (NS group) (Fig. 1). Novogene (UK) Company Limited conducted RNA sequencing using the Illumina NovaSeq 6000 platform, resulting in an average of 24M paired-end reads with a length of 150 bp per sample. The short reads produced a total output of 195 Gb (Table 1), the assembly of which generated 142,123 transcripts. Finally, a new transcriptome assembly reduced the list of generated transcripts to 78,398. The raw paired-

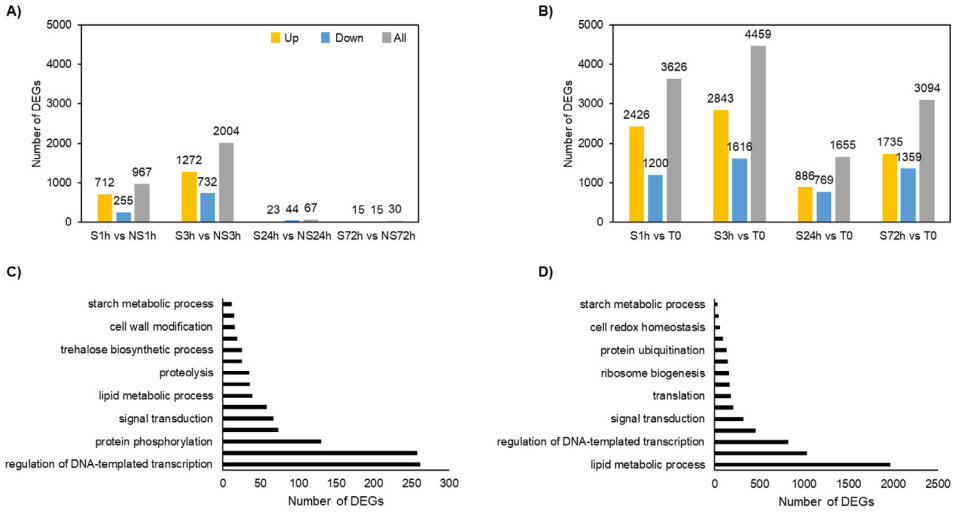

Fig. 2. Differentially expressed genes (DEGs) and GO terms in response to mechanical stimulation (**A, C**) in stimulated (S) plants compared to non-stimulated (NS) or (**B, D**) to plants sampled at T0 before mechanical stimulation.

end reads from 27 samples, the raw gene counts, the normalized counts and both assemblies are publicly available on the GEO database (GEO accession ID: GSE238196). Differentially expressed genes (DEGs) and the top of gene ontology (GO) categories were analyzed by pairwise comparison: S versus NS groups at different points in time, and S versus T0 groups (Fig. 2).

3. Experimental Design, Materials and Methods

3.1. Plant Material and Growth Conditions

Micro-cuttings of *Hydrangea macrophylla* cv. 'Wudu'® were produced by Hortensia France (Rives-du-Loir-en-Anjou, France) for the purposes of the experiment. After approximately 5 weeks, when the root system of the micro-cuttings had completely filled the micro-plugs (Green Products, Jongkind substrate: 35% Swedish peat/27.5% brown peat/37.5% perlite; v/v/v), the plants were transferred into 10 cm diameter pots containing a mixed substrate of peat and fertilized coconut (75%/25%; v/v) fertilized by 0.70 kg Pg-Mix Haifa 12-14-25, 0.3 kg microelements, 0.70 kg limestone, 1 kg dolomite lime and 5 kg clay per m^3. At the beginning of the experiment, young plants bore 3 pairs of well-developed leaves from a single unbranched stem (1.9 cm length on average) composed of 3 internodes, the first of which (at the collar) had an average diameter of around 2.7 mm. MS experiments were carried in a growth chamber (T° day/night: 23/21 °C; RH: 80%) where plants were sub-irrigated 2 days before the start of the experiment with a nutrient solution (Angibaud-Soluveg® ALC 47; EC: 1.8 mS/cm; pH: 6). Plants were grown under a 16 h/day photoperiod from artificial light using light-emitting diodes (Topband) with a photon flux density [380–780 nm] of 100 ± 2 µmol/m^2/s. Not mechanically stimulated plants (T0 and NS groups) were grown under same environmental conditions as mechanically stimulated plants (S group).

3.2. Mechanical Stimulation Treatment

MS device as described by Ley-Ngardigal et al. is composed of 2 arms allowing 2 simultaneous mechanical treatments, each arm moving back and forth at a speed of 12 m/h over a 2.1 m^2

surface area (2.1 m length x 1.0 m width) [1]. MS consisted in 12 consecutive bendings (i.e. 6 forth and 6 back) applied in the morning between 9:00 and 10:00 a.m. for a single day. The stimulation material was an unfringed 2 mm thick plastic curtain (2.3 kg/m^2) with the lower part of the curtain placed 0.5 cm below the plant apex. Stems were bent by the curtain at an angle of 30° with reference to verticality for about 30 s (Fig. 1).

3.3. Stem Sampling

Samples consisted in the whole stem (from the base of the apex to the collar, without leaves) of the plants and were collected 15 minutes before mechanical stimulation treatment (T0) then, following 12 consecutive bendings, at 1, 3, 24 and 72 h after the last bending (stimulated batches (S), Fig. 1). Four other batches collected on non-stimulated plants were also harvested at the same times (NS 1 h to NS 72 h). A total of 45 plants were collected per batch forming 3 biological repeats of 15 plants.

3.4. RNA Isolation, Library Preparation and Sequencing

Approximately 70 mg of stems were crushed in a mortar using liquid nitrogen. RNAs were extracted using the RNA plant extraction kit (Macherey-Nagel). Quality and quantity of RNAs were checked using the Nanodrop One (Thermo Scientific, Waltham, MA, USA), a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and their integrity was analyzed on agarose gel. All samples were of good quality (mean values for OD 260/280 ratio = 2.2 ± 0.01, OD 260/230 ratio = 2.3 ± 0.3; RNA Integrity Number, RIN = 8.6 ± 0.8; 28S/18S ratio = 1.7 ± 0.1) and were sent to the Novogene (UK) Company Limited (United Kingdom; https://www.novogene.com). Libraries of the 27 samples were sequenced using the Illumina NovaSeq 6000 PE150 (Novogene, United Kingdom) and yielding an average 24M of 150 bp reads per sample.

3.5. RNA seq Data Analyses

Reads containing adapters, low-quality sequences (≤ Q20) and with more than 10% poly-N sequences were removed using a Novogene in-house workflow before *de novo* transcriptome assembly. Quality assessment of the reads is shown in Table 1. Transcriptome assembly was performed by Trinity v2.6.6 software, followed by CORSET v4.6 software to remove redundancy from Trinity results [2,3]. BUSCO v3.0.2 was then used to assess the quality of structural annotation of the transcriptome [4]. Finally, the EvidentialGenes pipeline [5] was executed on the Novogene assembly, reducing the number of transcripts from 142,123 to 78,398. Gene functional annotation was performed using InterProScan-5.60-92.0 [6,7], Diamond v0.8.22 [8,9], NCBI blast v2.9.0, Hmmscan HMMER 3.1 [10], blast2go (b2g4pipe_v2.5) [11] with 8 databases: NR (NCBI non-redundant protein sequences), NT (NCBI nucleotide sequences), PFAM (Protein family), PROSITE (Protein families and domains), UniProtKB/Swiss-Prot (UniProt Knowledgebase), KOG/COG (eu-Karyotic Orthologous Groups/Cluster of Orthologous Groups of proteins), KEGG (Kyoto Encyclopedia of Genes and Genome) and GO (Gene Ontology). Gene expression levels were estimated for each sample using RSEM v1.2.28 software [12]. Differential expression analysis was performed using DESeq2 [13] package from AnaDiff v4.3 [14]. Genes with log2 Fold Change ≥ 1 or ≤ −1 and Benjamini-Hochberg adjusted p-values ≤ 0.05 were considered as differentially expressed. An overview of the bioinformatic tools and parameters used to analyze the *Hydrangea macrophylla* transcriptome is presented in Table 2.

Table 2
Bioinformatic tools used to assemble and analyze the *Hydrangea macrophylla* transcriptome.

Analysis	Software	Refs.	Parameters	Databases	Websites
Assembly	Trinity v2.6.6	[2]	minKmerCov= 3 min_glue=4	-	https://github.com/trinityrnaseq/trinityrnaseq/wiki
	CORSET v4.6	[3]	-f ture, Default, -m 10	-	https://github.com/Oshlack/Corset/wiki
	BUSCO v3.0.2	[4]	-m tran	-	http://busco.ezlab.org/
	EvidentialGene	[5]	tr2aacds2.pl-MINCDS=90	-	http://eugenes.org/EvidentialGene/evigene/
Gene Functional Annotation	InterPro Scan-5.60-92.0	[6,7]	-	PROSITE, PFAM annotation	https://prosite.expasy.org/ http://pfam.xfam.org
	Diamond v0.8.22	[8]	e-value = 1e-5	NR, KOG/COG, UniProtKB/Swiss-Prot	http://www.ncbi.nlm.nih.gov/COG/ http://www.ebi.ac.uk/uniprot/
	Diamond v0.8.22, KAAS	[9]	e-value = 1e-5	KEGG annotation	http://www.genome.jp/kegg/
	NCBI blast v2.9.0	-	e-value = 1e-5	NT annotation	http://www.ncbi.nlm.nih.gov/
	Hmmscan (HMMER 3.1 package)	[10]	e-value = 0.01	PFAM annotation	http://pfam.xfam.org
	blast2go (b2g4pipe_v2.5)	[11]	e-value = 1e-6	GO annotation	http://www.geneontology.org/
Quantification	RSEM v1.2.28	[12]	–estimate-rspd -mismatch-rate 0.3	-	https://github.com/deweylab/RSEM
Differential Expression Analysis	AnaDiff v4.3 (DESeq2 package)	[13,14]	-	-	https://zenodo.org/record/6477918

Limitations

None.

Ethics Statement

This work does not contain any studies with human or animal subjects.

Data Availability

Unraveling the dataset transcriptomic response of Hydrangea macrophylla stem to mechanical stimulation: de novo assembly and functional annotation (Reference data) (NCBI GEO).

CRediT Author Statement

Béra Ley-Ngardigal: Formal analysis, Investigation, Data curation, Methodology, Project administration, Validation, Visualization, Writing – original draft, Writing – review & editing; **Sandra Pelletier:** Formal analysis, Software, Data curation, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing; **Vincent Guérin:** Conceptualization, Supervision, Validation, Writing – review & editing; **Lydie Huché-Thélier:** Conceptualization, Supervision, Validation, Writing – review & editing; **Nathalie Brouard:** Investigation; **Hanaé Roman:** Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Validation, Writing – review & editing; **Nathalie Leduc:** Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Acknowledgments

This work is produced by the ANR (French National Research Agency) Common Laboratory MATCH associating the UMR1345 IRHS (Research Institute of Horticultural and Seeds) and Hortensia France. Béra Ley-Ngardigal received her PhD grant from Angers University and Hortensia France Company. The authors thank ANAN technical platforms of SFR Quasav and especially Muriel Bahut for her advices in molecular analysis. We are also grateful to Jean-Pierre Renou (retired researcher of UMR1345 IRHS), Nathalie Leblanc-Fournier (UMR547 PIAF, University of Clermont), Stéphane Maury (LBLGC, INRA, University of Orléans, EA1207 USC 1328) and Philippe Morel (retired researcher of UMR1345 IRHS) for fruitful scientific discussions.

Declaration of Competing Interests

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] B. Ley-Ngardigal, V. Guérin, L. Huché-Thélier, N. Brouard, T. Eveleens, H. Roman, N. Leduc, Impact of mechanical stimulation on *Hydrangea macrophylla*, Acta Hortic. (2023) 275–282.
[2] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome, Nat. Biotechnol. 29 (2011) 644–652.

[3] N.M. Davidson, A. Oshlack, Corset: enabling differential gene expression analysis for *de novo* assembled transcriptomes, Genome Biol. 15 (2014) 410.
[4] M. Manni, M.R. Berkeley, M. Seppey, E.M. Zdobnov, BUSCO: assessing genomic data quality and beyond, Curr. Protoc. 1 (2021).
[5] D. Gilbert, Geneomes built from mRNA seq not genome DNA, in: Proceedings of the 7th Annual Arthropod Genomics Symposium, Notre Dame, 2013, doi:10.7490/f1000research.1112594.1. http://arthropods.eugenes.org/EvidentialGene/about/EvigeneRNA2013poster.pdf and http://globalhealth.nd.edu/7th-annual-arthropod-genomics-symposium/.
[6] M. Blum, H.Y. Chang, S. Chuguransky, T. Grego, S. Kandasaamy, A. Mitchell, G. Nuka, T. Paysan-Lafosse, M. Qureshi, S. Raj, L. Richardson, G.A. Salazar, L. Williams, P. Bork, A. Bridge, J. Gough, D.H. Haft, I. Letunic, A. Marchler-Bauer, H. Mi, D.A. Natale, M. Necci, C.A. Orengo, A.P. Pandurangan, C. Rivoire, C.J.A. Sigrist, I. Sillitoe, N. Thanki, P.D. Thomas, S.C.E. Tosatto, C.H. Wu, A. Bateman, R.D. Finn, The InterPro protein families and domains database: 20 years on, Nucleic Acids Res. 49 (2021) D344–D354.
[7] P. Jones, D. Binns, H.Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A.F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S.Y. Yong, R. Lopez, S. Hunter, InterProScan 5: genome-scale protein function classification, Bioinformatics 30 (2014) 1236–1240.
[8] B. Buchfink, C. Xie, D.H. Huson, Fast and sensitive protein alignment using DIAMOND, Nat. Methods 12 (2015) 59–60.
[9] Y. Moriya, M. Itoh, S. Okuda, A.C. Yoshizawa, M. Kanehisa, KAAS: an automatic genome annotation and pathway reconstruction server, Nucleic Acids Res. 35 (2007) W182–W185.
[10] R.D. Finn, J. Tate, J. Mistry, P.C. Coggill, S.J. Sammut, H.R. Hotz, G. Ceric, K. Forslund, S.R. Eddy, E.L.L. Sonnhammer, A. Bateman, The Pfam protein families database, Nucleic Acids Res. 36 (2007) D281–D288.
[11] S. Gotz, J.M. Garcia-Gomez, J. Terol, T.D. Williams, S.H. Nagaraj, M.J. Nueda, M. Robles, M. Talon, J. Dopazo, A. Conesa, High-throughput functional annotation and data mining with the Blast2GO suite, Nucleic Acids Res. 36 (2008) 3420–3435.
[12] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, (2011).
[13] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014) 550.
[14] S. Pelletier, AnaDiff: a tool for differential analysis of microarrays and RNAseq (v4.3), (2022), doi:10.5281/zenodo.6477918.