Data Article

Dataset of *Arabidopsis* plants that overexpress FT driven by a meristem-specific KNAT1 promoter

L. Duplat-Bermúdez a, R. Ruiz-Medrano a, D. Landsman b, L. Mariño-Ramírez b, B. Xoconostle-Cázares a,∗

a Departamento de Biotecnología y Bioingeniería; Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional. Av. IPN 2508 San Pedro Zacatenco, 07360 Mexico City, Mexico

b Computational Biology Branch, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, United States of America, Bethesda, MD, USA

ARTICLE INFO

Article history:
Received 11 May 2016
Received in revised form
25 May 2016
Accepted 2 June 2016
Available online 8 June 2016

Keywords:
Differential expression
Bioinformatics
Flowering

ABSTRACT

In this dataset we integrated figures comparing leaf number and rosette diameter in three *Arabidopsis* FT overexpressor lines (AtF-TOE) driven by KNAT1 promoter, “A member of the KNOTTED class of homeodomain proteins encoded by the STM gene of *Arabidopsis*” [5], vs Wild Type (WT) *Arabidopsis* plants. Also, presented in the tables are some transcriptomic data obtained by RNA-seq Illumina HiSeq from rosette leaves of *Arabidopsis* plants of AtF-TOE 2.1 line vs WT with accession numbers SRR2094583 and SRR2094587 for AtFTOE replicates 1–3 and AtWT for control replicates 1–2 respectively. Raw data of paired-end sequences are located in the public repository of the National Center for Biotechnology Information of the National Library of Medicine, National Institutes of Health, United States of America, Bethesda, MD, USA as Sequence Read Archive (SRA). Performed analyses of differential expression genes are visualized by Mapman and presented in figures. “Transcriptomic analysis of *Arabidopsis* overexpressing flowering locus T driven by a meristem-specific promoter that induces early flowering” [2], described the interpretation and discussion of the obtained data.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

DOI of original article: http://dx.doi.org/10.1016/j.gene.2016.04.060

* Corresponding author.
E-mail address: bxoconos@cinvestav.mx (B. Xoconostle-Cázares).
Specifications Table

| Subject area               | More specific subject area | Type of data          | How data was acquired                                                                 |
|----------------------------|-----------------------------|-----------------------|--------------------------------------------------------------------------------------|
| Biology                    | Plant Sciences              | Figure; tables        | PCR final point and ddPCR (The QX100 Droplet Digital PCR (ddPCR™ System), direct count of leaves and measurement of rosette diameter with image analysis software (Imagej [http://rsb.info.nih.gov]), RNA-seq by Illumina HiSeq Sequencing, |

Data format: Analyzed

Experimental factors: Three lines overexpressing FT (AtFTOE) and WT Arabidopsis plants

Experimental features: Three lines of AtFTOE and WT Arabidopsis plants were grown on hydroponic under controlled conditions at 22 °C in short day (SD) photoperiod (8 h light /16 h dark) up to day 21 and then transferred to inductive conditions of long days (LD) photoperiod (16 h light /8 h dark)

Data source location: Mexico City, Mexico and at the National Center for Biotechnology Information (NCBI)

Data accessibility: Data is available with this article and at NCBI accession numbers SRR2094583 and SRR2094587

Value of the data

- The amplification of transgene by PCR and copy variation number by ddPCR is fundamental to compare independent transgenic events.
- ANOVA and T-Student test were employed to assess statistical significance of data ($\alpha=0.05$) for leaf count and rosette diameter in order to compare three AtFTOE lines and WT.
- Differentially expressed genes in the context of cellular functions are graphically presented by Mapman to understand the integrative changes in the metabolism.
- Raw data from the Illumina HiSeq sequencing are available for further analyses.

1. Data

In this article are presented the data analyses (figures) from leaf count and rosette diameter for three lines AtFTOE (2.1, 3.1 an 4.3) compared with WT Arabidopsis plants (Fig. 2A and B respectively). Data corresponding to differential expression (log2 fold change) from AtFTOE 2.1 line vs WT Arabidopsis are visualized by Mapman (Fig. 3). Some data corresponding to down-regulated genes are presented in Table 3.

2. Experimental design, materials and methods

2.1. Determination of transgene (FT) and copy variation number in three AtFTOE lines

Transgene FT (560 bp) was amplified by PCR from three AtFTOE (2.1,3.1 and 4.3) lines (Fig. 1A). Droplet digital PCR (ddPCR) was employed to determine transgene (FT) copy variation number (CVN) (Fig. 1B). As template were used 2.5 ng of genomic DNA previously digested with HindIII. Droplets were generated for PCR reaction with the specific primers AtFT-qPCR (F) (5’-TCCGTHTTAATAGTCAATC-3’), FT-qPCR (R) (5’-CCACCAATACAAAAGTAGT-3) and probe TaqMan ddPCRFT [5’FAM]
TCCTGAGGTCTTCTCCACCA [3’BHQ1]. The 152 bp of PCR-amplified product Arabidopsis HMGB1 (AT3G51880) was used as an internal, reference gene. The primers used as reference were HMGB1 probe [5’HEX]AGGACCGCTGAGAAGCCT[3’BHQ1], HMGB1F (5’-CAGAAAGGTGGGAAAGAGGA-3’), and HMGB1-R (5’-AAGGACCCAAAACAAACCCAA-3’). The HMGB1 PCR-amplified product was 96 bp. After cycling, the PCR nano droplets were counted using the droplet reader Bio-Rad QX100 system (Bio-Rad 2012).

2.2. Determination of leaf number and rosette diameter in three AtFTOE and WT plants

Three AtFTOE (2.1,3.1 and 4.3) lines and WT Arabidopsis Columbia-0 ecotype were employed. Seeds were stratified, kept at 4 °C for 3 days in the dark and then germinated and grown in hydroponic system [1] at 22 °C under controlled conditions, initially in short days (8 h light, 16 h dark) and after 21 days the seedlings were transferred to long-day (16 h light, 8 h dark) photoperiod under 100–120 mmol m⁻² s⁻¹. Plants grown under these conditions were used for determining rosette leaf
Fig. 2. Interval plot from AtFTOE lines and WT for (A) Leaf count and (B) Rosette diameter at 21 days in SD conditions.

*Individual standard deviations were used to calculate the intervals.*
number (Fig. 2A) by individual counting and rosette diameter (Fig. 2B) quantified by image analyses using the software Imagej (http://rsb.info.nih.gov). Quantitative data were statistically analyzed with ANOVA and T-Student test at ($\alpha=0.05$).

2.3. Massive mRNA-sequencing by Illumina, bioinformatics analysis and data processing

For RNA-Sequencing by illumina HiSeq was used total RNA from rosette leaves of 35-day old of AtFTOE 2.1 line and Wild Type (WT) plants. RNA was extracted with the RNeasy Plant Kit (Qiagen). The RNA-seq experiments were conducted with RNA isolated from three biological replicates per AtFTOE 2.1 line and two biological replicates for WT with accession numbers SRR2094583 and SRR2094587 respectively. Illumina sequencing was performed at Otogenetics (Georgia-USA). Illumina HiSeq Sequencing with PE50 yielded 20 million reads by triplicate (Table 2). The raw data files are available at the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA).
accession numbers SRR2094583 and SRR2094587 for AtFTOE replicates 1–3 and AtWT for control replicates 1–2 respectively. The paired-end reads were aligned to the reference *Arabidopsis* genome using Tophat (version 2.0.10) [7] and Bowtie2 (version 2.1.0) [4]. The reference *Arabidopsis* genome and gene model annotation files (TAIR10 < ftp://igenome:G3nom3s4u@ussd-ftp.illumina.com/Arabidopsis_thaliana/NCBI/TAIR10/Arabidopsis_thaliana_NCBI_TAIR10.tar.gz>) were downloaded from the Illumina iGenomes (http://support.illumina.com/sequencing/sequencing_software/igenome.html). Differential expression was determined by Cufflinks (version 2.1.1) as described by Trapnell et al. [8] and then was visualized by CummeRbound, an edgeR package [3]. To analyze the variation in expression between two replicates from WT and three replicates from AtFTOE it was calculated the absolute difference of the log2 fold change and adjusted to *P*-value ≤ 0.05.

### 2.4. Visualization of differential expression by Mapman

In order to visualize differential expression we used Mapman tool, mapping the Mapman databases (http://mapman.gabipd.org/web/guest/mapman; Ath_AGL_LOCUS_TAIR10_Aug2012-3.m02; [6]) using raw data of differential expression log2 fold change and adjusted to *P*-value ≤ 0.05 (Fig. 3).

### 2.5. Primer design

Primers design (Table 1) was performed by OligoArchitect™ Primer and Probe Design Solutions (http://www.sigmaaldrich.com/technical-documents/articles/biology/probe-design-services.html). Gene sequences were obtained from Tair database (https://www.arabidopsis.org).

### Acknowledgments

This work was supported by Departmental funds from CINVESTAV-IPN, by CONACyT grants nos. 156162 to RR-M and 105985 to BX-C, and a SENASICA-SAGARPA grant to BX-C and RR-M. LD-B was
| Genes | Locus ID | Localization | Description | FC  |
|-------|----------|--------------|-------------|-----|
| **SPL2, SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 2** | At5g43270 | Nucleus | Member of the SPL (squamosa-promoter binding protein-like) gene family, a novel gene family encoding DNA binding proteins and putative transcription factors. | 0.57 |
| **AP2, APETALA 2, ATAP2, FL1, FLO2, FLORAL MUTANT 2, FLOWER** | AT4G36920 | Nucleus | Encodes a floral homeotic gene, a member of the AP2/EREBP (ethylene responsive element binding protein) class of transcription factors and is involved in the specification of floral organ identity, establishment of floral meristem identity, suppression of floral meristem indeterminacy, and development of the ovule and seed coat. AP2 also has a role in controlling seed mass. | 0.46 |
| **BETA CA2** | At5g07690 | Nucleus | Encodes a putative transcription factor. | 0.56 |
| **CA1** | At3g01500 | Apoplast, chloroplast stroma, thylakoid membrane, plasma membrane | Encodes a putsative beta-carbonic anhydrase betaCA1. | 0.44 |
| **ATMYB29** | At3g05730 | Extracellular region | Encodes a defensin-like (DEFL) family protein. | 0.27 |
| **BETA CA2** | At5g14740 | Apoplast, chloroplast, chloroplast stroma, thylakoid membrane, cytoplasm | Encodes a beta carbonic anhydrase likely to be localized in the cytoplasm. | 0.41 |
| **BETA GLUCOSIDASE 33, BGLU33** | At3g26960 | Extracellular region | Pollen Ole e 1 allergen and extensin family protein. | 0.37 |
| **XCP2, XYLEM CYSTEINE PEPTIDASE 2** | At1g20850 | Chloroplast, extracellular space, lysosome | Cysteine-type peptidase activity, peptidase activity | 0.15 |
| **TIN1 RECRUITING MOTIF 14, TRM14** | At3g61380 | Nucleus, plasma membrane | Phosphatidylinositol N-acetylglucosaminyltransferase subunit P-related. | 0.39 |
| **ATIMD3, IMD3, IPMDH1** | At1g31180 | Chloroplast stroma, mitochondrion, plastid, thylakoid | The AtIMD3 is one out of 3 genes encoding the enzyme 3-isopropylmalate dehydrogenase involved in leucine biosynthesis in Arabidopsis. Its subcellular localization has been targeted to plastids. | 0.61 |
| **At3g63160** | Chloroplast, chloroplast envelope, chloroplast outer membrane, | Molecular_function unknown | | 0.48 |
supported by a doctoral fellowship from CONACyT no 388937. This research was partially supported by the Intramural Research Program of the NIH, NLM, NCBI.

### Transparency document. Supporting information

Transparency data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2016.06.002.

### Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2016.06.002.

---

| Genes | Locus ID | Localization | Description | FC |
|-------|----------|--------------|-------------|----|
| OEP6, OUTER ENVELOPE PROTEIN 6 | At1g35190 | Cytoplasm | 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein | 0.55 |
| AGAMOUS-LIKE 19, AGL19, GL19 | At4g22950 | Nucleus | MADS-box protein AGL19. | 0.23 |
| | At3g27200 | Extracellular region | Cupredoxin superfamily protein, electron carrier activity, copper ion binding. | 0.11 |
| | At2g39850 | Extracellular region | Subtilisin-like serine endopeptidase family protein; metabolic process, proteolysis. | 0.56 |
| | At3g53100 | Extracellular region | GD$SL$-motif esterase/acyltransferase/lipase, hydrolase activity, acting on ester bonds. | 0.33 |
| | At5g44400 | Cell wall, cytoplasm, plasmodesma | FAD-binding Berberine family protein. | 0.11 |
| MORPHOGENESIS OF ROOT HAIR 1, MRH1 | At4g18640 | Chloroplast | Protein serine/threonine kinase activity. | 0.11 |
| ATPA$P_3$, PAP$3$, PURPLE ACID PHOSPHATASE 3 | At1g14700 | Extracellular region, vacuole | Acid phosphatase activity, metal ion binding, protein serine/threonine phosphatase activity. | 0.25 |
| | At3g29030 | Cell wall, extracellular region, membrane | Encodes an expansin. Naming convention from the Expansin Working Group. | 0.17 |
| | At1g04040 | Cell wall, plant-type cell wall, plasmodesma, vacuolar membrane, vacuole | HAD superfamily, subfamily IIIB acid phosphatase | 0.27 |
| TPPH, Trehalose-6-phosphate phosphatase H | At4g39770 | Cytoplasm, cytosol, nucleus | Trehalose biosynthetic process. | 0.61 |
| | At5g63180 | Pectin lyase fold/virulence factor | Pectin lyase-like superfamily protein | 0.35 |
References

[1] S.J. Conn, B. Hocking, M. Dayod, B. Xu, A. Athman, S. Henderson, M. Gilliham, Protocol: optimising hydroponic growth systems for nutritional and physiological analysis of Arabidopsis thaliana and other plants, Plant Methods 9 (1) (2013) 4. http://dx.doi.org/10.1186/1746-4811-9-4.

[2] L. Duplat-Bermúdez, R. Ruiz-Medrano, D. Landsman, L. Mariño-Ramírez, Xoconostle-Cázares, Transcriptomic analysis of Arabidopsis overexpressing flowering locus T driven by a meristem-specific promoter that induces early flowering, Gene (2016). http://dx.doi.org/10.1016/j.gene.2016.04.060.

[3] L. Goff, C. Trapnell, D. Kelley, Cummerbund: analysis, exploration, manipulation, and visualization of Cufflinks high-throughput sequencing data. R package version 2.8.2., 2013.

[4] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods 9 (4) (2012) 357–359. http://dx.doi.org/10.1038/nmeth.1923.

[5] J. Long, E.E. Moan, J. Medford, K. Barton, A member of the KNOTTED class of homeodomain proteins encoded by the STM gene of Arabidopsis, Lett. Nat. 379 (1996) 66–69. http://dx.doi.org/10.1038/379066a0.

[6] O. Thimm, O. Bläsing, Y. Gibon, A. Nagel, S. Meyer, P. Krüger, J. Selbig, L.A. Müller, S.Y. Rhee, M. Stitt, mapman: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes, Plant J. 37 (2004) 914–939. http://dx.doi.org/10.1111/j.1365-313X.2004.02016.x.

[7] C. Trapnell, L. Pachter, S.L. Salzberg, TopHat: discovering splice junctions with RNA-Seq, Bioinformatics 25 (9) (2009) 1105–1111. http://dx.doi.org/10.1093/bioinformatics/btp120.

[8] C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D.R. Kelley, L. Pachter, Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks, Nat. Protoc. 7 (3) (2012) 562–578. http://dx.doi.org/10.1038/nprot.2012.016.