



Data Article

Transcriptome datasets from leaves and fruits of the sweet cherry cultivars 'Bing', 'Lapins' and 'Rainier'



Jonathan Maldonado^a, Amit Dhingra^b, Basilio Carrasco^c,
Lee Meisel^d, Herman Silva^{a,*}

^a Universidad de Chile, Facultad de Ciencias Agrónomas, Departamento de Producción Agrícola, Laboratorio de Genómica Funcional, & Bioinformática, 820808 La Pintana, Santiago, Chile

^b Department of Horticulture, Washington State University, Pullman, WA, USA

^c Facultad de Agronomía e Ingeniería Forestal, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Casilla 306, Macul, Chile

^d Universidad de Chile, Instituto de Nutrición y Tecnología de los Alimentos (INTA), El Líbano 5524, 7830490 Macul, Santiago, Chile

ARTICLE INFO

Article history:

Received 11 December 2018

Received in revised form

8 January 2019

Accepted 18 January 2019

ABSTRACT

Sweet cherry fruits from different cultivars have different pre- and post-harvest qualities. Here we present the transcriptome profile datasets of leaves and mature fruits of three sweet cherry cultivars ('Bing', 'Lapin' and 'Rainier'). Using 454 GS-FLX technology (454 Life Sciences, Roche), transcriptomes of leaves and mature fruits were obtained from these cultivars. These transcriptome data sets are reported here.

© 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

Specifications table

Subject area	Biology
More specific subject area	Plant biology; Bioinformatics; Agronomy
Type of data	Table, text file, graph, figure

* Corresponding author.

E-mail addresses: jomaldon@uchile.cl (J. Maldonado), adhingra@wsu.edu (A. Dhingra), bcarrasco@uc.cl (B. Carrasco), lmeisel@inta.uchile.cl (L. Meisel), hesilva@uchile.cl (H. Silva).

<https://doi.org/10.1016/j.dib.2019.01.044>

2352-3409/© 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

How data was acquired	RNA deep sequencing, 454 FLX
Data format	Raw
Experimental factors	Extraction of total RNA from leaves and fruit of Bing, Lapins and Rainier sweet cherry cultivars.
Experimental features	Shotgun sequencing of total RNA followed by bioinformatics analysis for differential gene expression
Data source location	La Palma Experimental Station, Faculty of Agronomy PUCV, Quillota (latitude 32° 54' S and longitude 71° 12' W); Los Andes (latitude 32° 49' S and longitude 70° 35' W) and San Francisco de Mostazal (latitude 33° 59' S and longitude 70° 41' W).
Data accessibility	The nucleotide sequences of raw reads were submitted to NCBI's Sequence Read Archive through the BioProject PRJNA497458 (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA497458) and RNAseq data were submitted to NCBI's Gene Expression Omnibus database with the accession number GSE121997 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE121997).
Related research article	J.C. Rios, F. Robledo, L. Schreiber, V. Zeisler, E. Lang, B. Carrasco, H. Silva, Association between the concentration of n-alkanes and tolerance to cracking in commercial varieties of sweet cherry fruits, <i>Scientia Horticulturae</i> , 197 (2015) 57-65. http://dx.doi.org/10.1016/j.scienta.2015.10.037

Value of the data

- These sweet cherry transcriptome datasets may be used to improve the genome assembly and annotation of *Prunus avium* [1].
 - Transcriptome datasets may be used to mine for Single Nucleotide Polymorphisms (SNPs) that are polymorphic between these sweet cherry cultivars.
 - Transcriptome datasets may be used to identify differentially spliced transcripts in sweet cherry leaves and fruits.
 - These datasets may be used to identify differentially expressed genes that may correlate with the phenotypic variabilities among these and other sweet cherry cultivars.
 - Profiling of leaf and fruit transcriptome datasets may be used to identify genetic targets for molecular markers assisted breeding programs [2–4].
-

1. Data

Here we report the transcriptome datasets of leaves and fruits of three sweet cherry cultivars ('Bing', 'Lapin' and 'Rainier') with contrasting pre- and post-harvest fruit qualities (Table S1). Total RNA sequencing resulted in 956,609 total raw reads, comprising 362 Mb of data, with an average read length of 378 bp. After trimming the number of reads was 938,279 (98% of raw reads) and the average read length was found to be 380 bp. Reads mapped against *Prunus avium* genome predicted genes ($n = 43,673$ genes) [1] correspond to 55% of filtered reads and covered on average 24% of all genes in all samples. The summary of the data is listed in Table 1. Predicted genes were also functionally annotated using Blast2GO (Table S2).

A Principal Component Analysis of the predicted genes from all the tissue samples in the sweet cherry cultivar transcriptome datasets was performed in order to determine the similarities between the whole RNA profiles of these samples (Fig. 1).

We analyzed the Differentially Expressed Genes (DEGs) of these transcriptome datasets using three criteria for their expression profiles: (i) the number of reads that maps to each contig was more than or equal to 10 when compared to another cultivar (dual comparison); (ii) the average fold change in transcript accumulation was more than or equal to 2-fold between cultivars and (iii) the Z-Test

Table 1
Statistics of the sweet cherry transcriptome datasets.

Feature	Leaf			Fruit		
	'Bing'	'Lapins'	'Rainier'	'Bing'	'Lapins'	Rainier
Cultivar	'Bing'	'Lapins'	'Rainier'	'Bing'	'Lapins'	Rainier
NCBI Bioproject ID	PRJNA497458					
NCBI Biosample ID	SAMN	SAMN	SAMN	SAMN	SAMN	SAMN
	10258413	10258416	10258418	10258415	10258417	10258419
NCBI SRA ID	SRX4958445	SRX4958446	SRX4958447	SRX4958448	SRX4958449	SRX4958450
NCBI GEO ID	GSE121997					
Sequence type	454 FLX (454 Life Sciences, Roche)					
Total number of reads (raw)	165,657	143,076	143,988	208,854	144,893	50,141
Read length (raw) in base pairs	383.3	361.4	378.0	411.3	369.9	365.9
Total number of reads (filtered)	162,773	140,205	141,408	205,440	141,677	146,776
Read length (filtered) in base pairs	384.0	362.7	378.5	412.5	371.8	367.5
Total number of mapped reads	90,761	71,694	71,706	118,019	75,862	85,481
% of mapped reads	55.8%	51.1%	50.7%	57.4%	53.5%	58.2%
Number of genes with mapped reads	11,231	10,365	10,117	11,086	9276	9600
% of genes with mapped reads	25.7%	23.7%	23.2%	25.4%	21.2%	22.0%

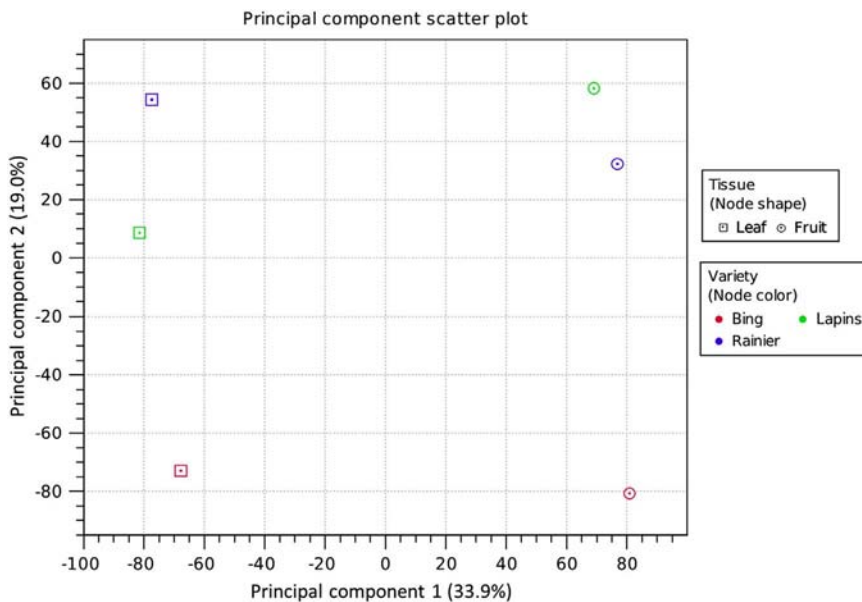


Fig. 1. Principal component scatter plot comparing the predicted genes in the transcriptome datasets from leaf and tissue samples of 'Bing', 'Lapins' and 'Rainier' sweet cherry cultivars.

p-value was less than or equal to 0.05. Under these criteria, 138 and 52 genes were identified to be differentially represented in the pairwise comparison of the leaf or fruit transcriptome datasets, respectively (Table 2 and Table 3). The transcriptomic analyses for all samples are available in Table S3 and Table S4. Table S5 and Table S6 contain the complete list of genes that are differentially represented in these transcriptome datasets using the pair-wise comparison of the cultivar specific transcriptome datasets of fruit and leaf tissues, respectively.

Table 2

Number of genes differentially represented in the pairwise comparison of the fruit transcriptome datasets of three sweet cherry cultivars. The cutoff was: fold of change $\geq 2X$ and p-value < 0.05 .

FRUIT	Number of over-represent genes respect to			Total
	'Rainier'	'Rainier' and 'Bing'	'Bing'	
'Lapins'	219	11	11	241
	'Lapins'	'Lapins' and 'Rainier'	'Rainier'	
'Bing'	64	65	55	184
	'Bing'	'Bing' and 'Lapins'	'Lapins'	
'Rainier'	168	5	2	175

Table 3

Number of genes with differentially represented in the pairwise comparison of the leaf transcriptome datasets of three sweet cherry cultivars. The cutoff was: fold of change $\geq 2X$ and p-value < 0.05 .

LEAF	Number of over-represented genes respect to			Total
	'Rainier'	'Rainier' and 'Bing'	'Bing'	
'Lapins'	90	6	3	99
	'Lapins'	'Lapins' and 'Rainier'	'Rainier'	
'Bing'	31	19	10	60
	'Bing'	'Bing' and 'Lapins'	'Lapins'	
'Rainier'	39	9	10	58

2. Experimental design, materials, and methods

2.1. Plant material

Sweet cherry leaf and fruit samples from the 'Bing' cultivar were collected from La Palma Experimental Station, Faculty of Agronomy PUCV, Quillota (latitude 32° 54' S and longitude 71°12' W). Samples from the 'Lapins' cultivar were collected from Los Andes (latitude 32°49' S and longitude 70°35' W). Samples from the 'Rainier' cultivar were collected from San Francisco de Mostazal (latitude 33°59' S and longitude 70°41' W). The fruits from these three cultivars were harvested at commercial maturity. All the fresh samples were frozen in liquid nitrogen and stored at -80°C until used for total RNA isolation. These cultivars were selected because they are commercially important cultivars that have contrasting pre- and post-harvest fruit quality (Table S1).

2.2. Library construction and deep sequencing

The total RNA was isolated using the protocol of Meisel et al. [5]. The quality and quantity of RNA was determined spectrophotometrically ($A_{260}/A_{280} = 1.8$ and $A_{260}/A_{230} = 2.0$) and electrophoretically using denaturing formaldehyde agarose gel.

Library construction and 454 FLX deep sequencing (454 Life Sciences, Roche) was performed by the Center for Integrated Biotechnology, Washington State University, using 1/8 plate. Trimming and quality filters were applied to the sequences using the CLC Genome Workbench software, version 11.0.1 (CLC Bio [<http://www.clcbio.com>]) [6].

2.3. RNA sequence analysis

The predicted coding sequences within the *Prunus avium* genome (43,673 predicted genes [1]) were used as reference sets to map the transcripts of these transcriptome datasets. The sequences from each cultivar and tissue were separately mapped against the corresponding reference transcriptome using RNA-seq function of the CLC Genome Workbench version 11.0.1 under the following parameters, similarity: 0.9; length fraction: 0.6; insertion/deletion cost: 3; mismatch cost: 3 and unspecific match limit: 10.

The relative transcript abundance in these datasets were obtained as the unique number of reads mapped to each gene. The transcript abundance in these datasets were compared using a Z-Test [7]. This test compared read counts by considering the proportions in which they make up the total sum of counts in each dataset, correcting for the size of the dataset. For visual inspection, the relative transcript abundance values were transformed using the Log10 method and then normalized by the Quantile method that was the best to fit the result [8,9].

2.4. Functional annotation

Functional annotation was performed on the genes differentially represented in the pairwise comparison of the leaf and fruit transcriptome datasets, using the coding regions predicted in the *Prunus avium* genome [1] and the Blast2GO CLC plugin version 1.11.9 [10] as described in [11]. A best-hit annotation was determined for the differentially represented gene sets by using these genes in a BLASTX (version 2.6.0) analysis against the nr NCBI database with an e-value cutoff of $1e^{-6}$. INTERPROSCAN analysis (version 5.31-70) with Blast2GO default parameters were also performed [12]. Blast2GO was also used for gene ontology mapping, with the program defaults being applied for all annotation steps and a False Discovery Rate (FDR) cut-off at the 0.05% probability level. The data from the INTERPRO terms, enzyme classification codes (EC), and metabolic pathways (KEGG, Kyoto Encyclopedia of Genes and Genomes) were merged with GO terms to provide a larger accumulation of evidence to support the annotations represented in the supplemental tables (Table S2).

Acknowledgments

Herman Silva and Basilio Carrasco are supported by CONICYT/FONDECYT Regular N°1160600 Chile. Lee Meisel is supported by CONICYT, FONDECYT/Regular N°1171016, Chile.

Transparency document. Supporting information

Transparency data associated with this article can be found in the online version at <https://doi.org/10.1016/j.dib.2019.01.044>.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at <https://doi.org/10.1016/j.dib.2019.01.044>.

Appendix B. Supporting information

Supplementary data associated with this article can be found in the online version at <https://doi.org/10.1016/j.dib.2019.01.044>. These data include Google maps of the most important areas described in this article.

References

- [1] K. Shirasawa, K. Isuzugawa, M. Ikenaga, Y. Saito, T. Yamamoto, H. Hirakawa, S. Isobe, The genome sequence of sweet cherry (*Prunus avium*) for use in genomics-assisted breeding, *DNA Res.: Int. J. Rapid Publ. Rep. Genes Genomes* 24 (2017) 499–508. <https://doi.org/10.1093/dnares/dsx020>.
- [2] B. Carrasco, L. Meisel, M. Gebauer, R.G. Gonzalez, H. Silva, Breeding in peach, cherry and plum: from a tissue culture, genetic, transcriptomic and genomic perspective, *Biol. Res.* 46 (2013) 219–230. <https://doi.org/10.4067/S0716-97602013000300001>.
- [3] C. Balbontin, H. Ayala, R.M. Bastias, G. Tapia, M. Ellena, C. Torres, J.A. Yuri, J. Quero-Garcia, J.C. Rios, H. Silva, Cracking in sweet cherries: a comprehensive review from a physiological, molecular, and genomic perspective, *Chil. J. Agric. Res.* 73 (2013) 66–72. <https://doi.org/10.4067/S0718-58392013000100010>.
- [4] J.C. Rios, F. Robledo, L. Schreiber, V. Zeisler, E. Lang, B. Carrasco, H. Silva, Association between the concentration of n-alkanes and tolerance to cracking in commercial varieties of sweet cherry fruits, *Sci. Hortic.* 197 (2015) 57–65. <https://doi.org/10.1016/j.scienta.2015.10.037>.
- [5] L. Meisel, B. Fonseca, S. Gonzalez, R. Baeza-Yates, V. Cambiazo, R. Campos, M. Gonzalez, A. Orellana, J. Retamales, H. Silva, A rapid and efficient method for purifying high quality total RNA from peaches (*Prunus persica*) for functional genomics analyses, *Biol. Res.* 38 (2005) 83–88. <https://doi.org/10.4067/S0716-97602005000100010>.
- [6] A. Morales, A. Zurita-Silva, J. Maldonado, H. Silva, Transcriptional responses of Chilean Quinoa (*Chenopodium quinoa* Willd.) Under water deficit conditions uncovers ABA-independent expression patterns, *Front. Plant Sci.* (2017), <https://doi.org/10.3389/fpls.2017.00216>.
- [7] A.J. Kal, A.J. van Zonneveld, V. Benes, M. van den Berg, M.G. Koerkamp, K. Albermann, N. Strack, J.M. Ruijter, A. Richter, B. Dujon, W. Ansoorge, H.F. Tabak, Dynamics of gene expression revealed by comparison of serial analysis of gene expression transcript profiles from yeast grown on two different carbon sources, *Mol. Biol. Cell* 10 (1999) 1859–1872. <https://doi.org/10.1091/mbc.10.6.1859>.
- [8] M.-A. Dillies, A. Rau, J. Aubert, C. Hennequet-Antier, M. Jeanmougin, N. Servant, C. Keime, G. Marot, D. Castel, J. Estelle, G. Guernec, B. Jagla, L. Jouneau, D. Laloë, C. Le Gall, B. Schaëffer, S. Le Crom, M. Guedj, F. Jaffrézic, A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis, *Briefings in Bioinformatics*, [http://dx.doi.org/10.1093/bib/bbs046\(2012\)](http://dx.doi.org/10.1093/bib/bbs046(2012)). (<http://dx.doi.org/10.1093/bib/bbs046>).
- [9] K.D. Hansen, R.A. Irizarry, Z. Wu, Removing technical variability in RNA-seq data using conditional quantile normalization, *Biostatistics* 13 (2012) 204–216. <https://doi.org/10.1093/biostatistics/kxr054>.
- [10] A. Conesa, S. Gotz, Blast2GO: a comprehensive suite for functional analysis in plant genomics, *Int. J. Plant Genom.* (2008) 619832. <https://doi.org/10.1155/2008/619832>.
- [11] J. Immanen, K. Nieminen, H. Duchens Silva, F. Rodríguez Rojas, L.A. Meisel, H. Silva, V.A. Albert, T.R. Hvidsten, Y. Helariutta, Characterization of cytokinin signaling and homeostasis gene families in two hardwood tree species: *populus trichocarpa* and *Prunus persica*, *BMC Genom.* 14 (2013) 885. <https://doi.org/10.1186/1471-2164-14-885>.
- [12] S. Hunter, R. Apweiler, T.K. Attwood, A. Bairoch, A. Bateman, D. Binns, P. Bork, U. Das, L. Daugherty, L. Duquenne, R.D. Finn, J. Gough, D. Haft, N. Hulo, D. Kahn, E. Kelly, A. Laugraud, I. Letunic, D. Lonsdale, R. Lopez, M. Madera, J. Maslen, C. McAnulla, J. McDowall, J. Mistry, A. Mitchell, N. Mulder, D. Natale, C. Orengo, A.F. Quinn, J.D. Selengut, C.J. Sigrist, M. Thimmia, P. D. Thomas, F. Valentin, D. Wilson, C.H. Wu, C. Yeats, InterPro: the integrative protein signature database, *Nucleic Acids Res.* 37 (2009) D211–D215. <https://doi.org/10.1093/nar/gkn785>.