Tissue-specific de novo transcriptome analysis in kiwifruit [Actinidia deliciosa (A Chev) Liang et Ferguson]

CURRENT STATUS: UNDER REVIEW

BMC Plant Biology  ▶ BMC Series

JUAN SALAZAR  jasalazar@cebas.csic.es
Centro de Edafologia y Biologia Aplicada del Segura
Corresponding Author
ORCID: 0000-0003-0380-076X

Cristian Vergara
Universidad Andres Bello

Claudia Jorquera
Universidad de Chile

Patricio Zapata
Universidad Mayor

Pedro Martínez Gómez
Centro de Edafologia y Biologia Aplicada del Segura

Rodrigo Infante
Universidad de Chile

Claudio Meneses
Universidad Andres Bello

DOI: 10.21203/rs.2.17801/v1

SUBJECT AREAS  Plant Physiology and Morphology

KEYWORDS  Actinidia, kiwifruit, de novo transcriptome, leaf, flower, fruit
Abstract

Background Kiwifruit [Actinidia deliciosa (A Chev) Liang et Ferguson] is a subtropical vine from the Actinidiaceae family native from China. This specie has an allohexaploid genome (from a diploid and autotetraploid parents) contained in 174 chromosomes producing a climacteric and fleshy fruit called kiwifruit. Currently there's no too much genomic and transcriptomic information about this species. In this low molecular knowledge context, the main goal of this work is to construct a tissue-specific de novo transcriptome assembly generating a differential expression analysis among these specific tissues to obtain new useful database for a better knowledge of vegetative, floral and fruit growth in different phenological states of Actinidia deliciosa cv. ‘Hayward’.

Results In the present study we have analyzed different whole transcriptomes from shoot, leaf, flower bud, flower and fruit at 4 development stages (7, 50, 120 and 160 days after flowering; DAF) in kiwifruit by using RNA-seq. We sequenced twenty-four libraries, obtaining 604,735,364 reads which were assembled using Trinity software. The first version of Actinidia deliciosa de novo transcriptome contained 142,025 contigs ($\bar{x}=1,044$bp, N50=1,133bp). CEGMA and BUSCO were used for assembly quality assessment, obtaining close to 90.0% (35.1% partial) and over 85.0% (18.3% partial) of the ultra-conserved genes for eukaryote and plants, respectively.

Annotation was performed with BLASTx against TAIR10 protein database and we found an annotation proportion of 35.6% (50,508), leaving 64.4% (91,517) of the contigs assembly without annotation.

Conclusions These results represent a reference transcriptome for allohexaploid kiwifruit generating a database of Actinidia deliciosa genes related to leaf, flower
and fruit development. Thus, the present study provides a high valuable information, identifying over 20,000 exclusive genes including all tissue comparisons, which are associated with the proteins involved in different biological processes and molecular functions. Transcriptome assembly and refining as well as the assembly metric assessment, has implied an enough quality to be a putative database of this specie and high number of ultra-conserved proteins were found. With respect to transcriptome close to 65% of contigs did not match with any protein. Therefore, future functional annotation will be required in order to obtain a better knowledge of the tissue-specific development.

Background

*Actinidia delicosa* or *Actinidia chinensis delicosa* is an allohexaploid specie (2n = 6x) with an estimated genome size of 4.4 Gbp and 174 chromosomes [1,2,3]. It is a species originated in southwestern China and belongs to the genus *Actinidia* which covers at least 76 species [4]. It has an important economic role due to his edible fruit known as kiwifruit, being also a species of nutritional importance due to his high level of vitamin C [5]. The kiwifruit worldwide production in 2017 reached about 4.38 million tons, being China the most important producer, followed by Italy, New Zealand, Iran and Chile. However, most of the Chinese kiwifruit are locally consumed, with less than 0.2% of the total being exported (http://faostat.fao.org).

*Actinidia* genus is mainly composed of dioecious plants. The flowers are axial with five petals usually white colored and sepals that can be fused with the base or not. The fruit is a berry and has black seeds embed within the flesh where colors and morphology vary depending on the cultivar. *Actinidia delicosa* cv. ‘Hayward’ is one of the most widely distributed cultivars, bears fruits with green flesh and a fuzzy
appearance. The characteristic flavor of the kiwifruit is one of the important aspects for consumer acceptance and commercial potential. The balance of sweetness and acidity is given by the sugars such as glucose, fructose, sucrose and myo-inositol in small quantities (relative to other trees) and the acids found in mature fruit such as quinic acid, citric acid and in less proportion, malic acid [6].

As far as we know there is scarce information about genomic and transcriptomic resources in this species mainly taking in account the ploidy level. At genomic level, the whole genome sequencing of heterozygous ‘Hongyang’ cultivar [7] and the ‘Red5’ genotype [8] has been reported. In addition, transcriptomic analysis is relatively recent in kiwi species with a great potentiality regarding its physiology, fruit ripening studies or resistance to diseases. These transcriptomic studies included the sequencing of male and female transcriptomes of A. arguta flower buds [9], the study of fruit development in the cultivar “Hongyang” based on long noncoding RNA expression and alternative splicing [10], the premature budbreak dormancy analysis [11], the genome-wide gene expression profiles in male and female plants using high-throughput RNA sequencing (RNA-Seq) [12], and the study of infection by Pseudomonas syringae pv. actinidiae (Psa) [13,14]. Finally, from a metabolomics point of view, due to the economic and nutritional importance of red-fleshed kiwifruit, some studies have reported new findings on anthocyanin biosynthesis in A. chinensis [15,16] and A. arguta [17].

All the transcriptomic studies above mentioned were focused on specific areas that deal with the most important problems of kiwifruit and that suppose important agronomic and economic impact. In addition, these transcriptomic approaches gain an especial interest in referenced commercial varieties as ‘Hayward’ which is the most cultivated green fleshy kiwifruit variety in the world. Therefore, it would be
necessary to complete his genomic and transcriptomic information with new gene database available for breeders, nutritionists and plant biologists from the transcriptomic analysis of the different tissues.

The aim of this work is to construct a *de novo* transcriptome in green fleshed kiwifruit ‘Hayward’ (*Actinidia deliciosa*) using RNA-Seq analysis from leaf, flower and fruit tissues in different development phases including a differential gene expression analysis among vegetative and reproductive tissues to generate a database of *Actinidia deliciosa* genes related to leaf, flower and fruit development.

results

**Refined *de novo* transcriptome assembly**

In order to obtain a multiple-tissue *de novo* transcriptome of genes from *Actinidia deliciosa*, RNA libraries were constructed and sequenced from shoots, leaves, flower buds, flowers and fruits at 7, 50, 120 and 160 DAFB with a total of 604,735,364 million pair-end reads (150bp x 2) (Supplementary Table S1). All raw reads in FASTQ format including paired-end and replicates are available from the NCBI Short Read Archive (SRA) database under bioproject number PRJNA564374 ([https://www.ncbi.nlm.nih.gov/sra/PRJNA564374](https://www.ncbi.nlm.nih.gov/sra/PRJNA564374)). These reads were processed and trimmed to remove low quality reads and then used as the input to assemble the transcriptome with Trinity software. The assembled and refined transcriptome contains 142,025 total contigs where 106,603 contigs are trinity “genes” unigenes. Moreover, the average size for contigs was 1,044, reaching the largest one 36,186 bp while the shortest is just 501 bp long (Table 1). In general, the N50 of contigs were checked alongside the length distribution of contigs and percentage of input reads that maps back to the assembled transcriptome, but in order to conclude
about the quality of an assembled transcriptome it is needed more analyses and assessments [18].

To assess the quality for our de novo transcriptome we performed a biological based analysis approach. The biological analysis needs the use of the ultra-conserved proteins encoded in the assembled transcriptome, searching them and reporting how many are completed, fragmented and missing. Therefore, we used CEGMA and BUSCO to evaluate the completeness of ultra-conserved protein coding genes for eukaryote and plant species. CEGMA found 140 (56.8%) complete, 87 (35.1%) fragmented and 20 (8.1%) missing genes from its 248 eukaryotic ultra-conserved proteins data set. However, BUSCO found 1,007 (69.9%) complete, 263 (18.3%) fragmented and 170 (11.8%) missing genes from the ultra-conserved plant database (1,440 total plant-specific proteins) (Table 1). Finally, an additional quality control of the samples and biological replicates was performed using Trinity auxiliary scripts (Fig. 2) showing high similarities between biological replicates and variable differences among different tissues samples.

**Transcriptome annotation and differential expressed genes**

After the annotation process against the TAIR10 protein database, we obtained a total of 50,508 contigs (35.6%) annotated, leaving 91,517 contigs (64.4%) without annotation (Table 2; Supplementary Table S2). From the annotated contigs, the 99.99% corresponded to a diverse plant species. The top 3 hit species were *Vitis vinifera* (18.2%), *Juglans regia* (5.8%) and *Coffea canephora* (4.3%) (Table 2). Also, we found that 1.5% of the annotation corresponds to the *Actinidia* genus.

In order to perform a general comparison of the number of gene expression differences between tissues, a differential expression analysis was performed. When
comparing all the replicates and tissues (using a normalized factor based on fold-change between each expression and the mean expression across samples) with filters of false discovery rate (FDR) $\geq 0.005$ and FC $\geq 4$ we found around 49,000 differential expressed genes (Fig. 2; Supplementary Fig. S1). In addition, to obtain a better understanding about tissues comparation, differential gene expression was evaluated by edgeR [19], grouping different tissues as follow: shoot vs leaf, flower bud vs flower, fruit 7d vs fruit 50d, fruit 7d vs fruit 120d and fruit 7d vs fruit 160d (Fig. 3). Thus, through these comparisons we can obtain the overexpressed and under expressed genes of each tissue comparation related to leaves vegetative growth, flower and fruit development. Finally, we've obtained 22,962 exclusive genes for all tissue comparison at p-value $< 0.001$ obtaining 1,567 (shoot vs leaf), 10,634 (flower bud vs flower), 2,113 (fruit7 vs 50), 3,560 (fruit7 vs 120) and 5,088 (fruit7 vs 160) (Fig. 3; Supplementary Table S3).

**Gene annotation and enrichment analysis**

To explore the main gene functions involved in the vegetative growth leaves, floral and fruit development from the biological processes and molecular function point of view we performed a Singular Enrichment Analysis (SEA) by Panther G0 slim [20] and AgriGo v2 [21], using the top hit specie *Vitis vinifera* as reference (Fig. 4; Supplementary Figures S2 to S10; Supplementary Tables S3 and S4).

According to annotations by Panther (Fig. 4; Supplementary Table S3) in shoot vs leaf comparisons for biological processes, the top hits GO terms were metabolic process (23.1%; GO:0008152) and cellular component organization or biogenesis (23.1%; GO:0071840) while for molecular function was molecule binding (42.9%; GO:0005488). As for flower bud vs flower the most important GO term for biological processes was metabolic process (45.9%; GO:0008152) while for molecular function
was the catalytic activity (45.1%; GO:0003824) as well as for fruit development (fruit7d vs 50d–120d–160d) the top hits for biological processes and molecular function were also metabolic process (45.7%; GO:0008152) and catalytic activity (51.8%; GO:0003824) respectively.

As for GO terms annotations by AgriGO v2 (Supplementary Table S4) only eight GO terms were found for shoot vs leaf according to p-value, thus we can highlight the response to stimulus (1.3e–04; GO:0050896) and nucleic acid binding (1.8e–05; GO:0003676) for biological processes and molecular function respectively.

According to floral and fruit development we found 119 and 56 GO terms respectively. A total of 69 GO terms were exclusives for flower bud vs flower while only six were exclusives for fruit7d vs 50d–120d–160d (Supplementary Table S4).

discussion

According to the assembled and refined transcriptome, a total of 142,025 contigs (\( \bar{x} = 1,044\)bp, N50 = 1,133bp) were obtained which 106,603 contigs correspond to trinity “genes” unigenes. Within assembled contigs the size ranged from 501 bp to 36,186 bp long. In other studies, as *Actinidia eriantha* [22], the assembly of filtered reads reached a total of 69,396 unigenes obtaining sizes between 201 and 9,602 bp, which are indicating a larger number of contigs and size in this study.

As for the completeness of ultra-conserved protein evaluation by CEGMA and BUSCO, the scores suggest that the assembled transcriptome contains an important number of ultra-conserved proteins complete or fragmented, and a minor proportion of them missing, meaning that the transcriptome reached a high-quality standard in terms of completeness. However, CEGMA and BUSCO complete scores over 95% have been reported for twelve plant genomes including the model plant *Arabidopsis*
thaliana and the fruit tree *Pyrus communis* L. var ‘Bartlett’ [23].

Thus, this biological approach is indicating the highest quality score specially for BUSCO which reach almost 70% of complete genes found from the dataset. This 70% from BUSCO score seem to be lower against to quality scores reported above in different plant genomes but the construction of *de novo* transcriptome from three different plant tissues may be explain this difference. Anywise, the sum of the values of complete and fragmented genes is close to 90%, which is quite high considering the construction of a *de novo* transcriptome. In addition, despite the fact that we are talking about different species, in a specific tissue *de novo* transcriptome assembly of *Ilex paraguariensis* [24] obtained around 73% of complete genes reaching close to 85% if we include the fragmented genes being this quality score value similar to our transcriptome assembly.

At this moment, the high throughput sequencing (HTS) it is becoming more and more accessible and friendly, as this happens, more species are getting sequenced for multiple purposes including gene expression profiling, epigenomics, genomics, and transcriptomics approaches [25].

In spite of everything, there is still species with economic relevance that do not have this omics data resources yet. This is the reason why assembling a high-quality transcriptome becomes an important task when comes to studying those species. However, assembling a high quality *de novo* transcriptome depends mainly in the quality, quantity and software parameters, and even if the transcriptome was assembled, it represents a challenge in terms of determining how good it is assembled.

As a result of annotation process against the TAIR10 protein database, only 35.6 % of contigs were annotated, therefore a 64.4% remained without score. This low
match could be explained because there is a lack of information in protein databases, which implies a lower knowledge in the Actinidia genus at protein level. Other studies in kiwifruit Actinidia deliciosa var ‘Jinkui’ [26] obtained 140,187 unigenes of which 56,912 were functionally annotated while [9] obtained in Actinidia arguta 51,745 unigenes and 30,439 matches to known proteins. More recently, a RNA-seq for different fruit tissues was carried out in Actinidia eriantha reaching 69,783 non-redundant unigenes and 21,730 were annotated in different protein database [22]. As we have seen, the sequencing results of different kiwi species including A. deliciosa, A. arguta and A. eriantha show important differences at the transcriptome level.

As for gene annotation and enrichment analysis by Panther, we obtained significative functional annotations from each tissue comparisons including shoot vs leaf, flower bud vs flower and for fruit development. As for shoot vs leaf comparison the top hits for biological processes were related to metabolic process (GO:0008152) or cellular component organization (GO:0071840) while for molecular function was molecule binding (GO:0005488). Therefore, these processes may be involved in the biosynthesis of constituent macromolecules and plant cell related to leaf development and growth. As for bud vs flower and fruit development (fruit7d vs 50d-120d-160d) the top hit for biological processes was metabolic process (GO:0008152) while for molecular function the catalytic activity (GO:0003824) was the most significative for both tissue comparison which is indicating an increase of chemical reactions linked to flowering and fruit development.

In agreement with the GO terms annotations by AgriGO v2 only eight GO terms were found for shoot vs leaf may be due to the lack of protein annotations related to these tissues. However, bud vs flower and fruit development comparisons showed
119 and 56 GO terms respectively. Thus, for flower bud vs flower, some significant protein IDs were related to anatomical structure development (p-value of 6.5e-27; GO:0048856), reproductive process (3.4e-12; GO:0022414), aromatic compound biosynthesis process (7.4e-07; GO:0019438) or response to abiotic stimulus (1.00E-07; GO:0009628). Therefore, GO:0048856 is related to the progression of anatomical structures as flower bud to mature flower while proteins network involved in GO:0022414 are contributing in the reproductive process related to inheritance of genetic material from the parents which it’s happens during flower development and it’s been recently related to cytoplasmatic male sterility in soybean flower buds [27]. Moreover, aromatic compound biosynthesis process (GO:0019438) includes all of chemical reactions and pathways related to the formation of aromatic compounds which can be happening during flower development, so the aromatic composition of the kiwifruit could be forming during flower development as in *Eucalyptus grandis* floral tissues [28]. As for proteins network related to response to abiotic stimulus (GO:0009628), it is related to flower development which may be is conditioned by abiotic stresses as it was been recently reported in flower buds of transgenic blueberry [29].

As for fruit development (fruit7d vs 50d–120d–160d), some protein IDs were associated to protein metabolic process (4.1e-3; GO:0019538), DNA polymerase activity (3.9e-10; GO:0034061), transferase activity (6.1e-06; GO:0016740) or catalytic activity (1.6e-03; GO:0003824). In other plant species as cucumber [30] some of the main proteins linked to fruit development are involved in the processes of the protein metabolism (GO:0019538). In addition, some of the most important proteins involved to S2 fruit development stage in peach (cell enlargement) are involved to transferase and catalytic activities [31].
Therefore, a major catalysis reaction and an increase of enzymatic activity seem to be more related to fruit development (GO:0034061, GO:0016740 and GO:0003824). Similar approach was implemented in a de novo assembly of Persea americana cv. ‘Hass’ (avocado) transcriptome during fruit development where proteins related to fruit oily characteristics were predominant [32].

Finally, we have to take into account that the lack of information regarding protein annotations, suggest that more information would be needed to deepen in the processes related to vegetative growth of leaves or the reproductive development of the flower and fruit. However, the results obtained through the gene enrichment analysis suggest certain behaviors that can be adapted to each tissue.

conclusions

This work represents the first high quality tissue-specific de novo transcriptome of Actinidia deliciosa with a description of differentially expressed genes linked to the vegetative growth (shoot vs leaves), floral (flower bud vs flower) and fruit development. This information is of special interest in plant species, such as kiwi, with a wide ploidy level variability. The measured metrics and scores of the transcriptome assembly proved to have enough quality to be a putative database of this species transcripts. In addition, the biological approach for the transcriptome assembly ensure that our transcriptome have a high quality containing and important number of ultra-conserved proteins. As for the transcriptome annotations, around 35% of contigs were annotated, however close to 65% of contigs did not match with any protein, which is indicating a large number of sequences not associated to functional annotations. Therefore, this work provides valuable information identifying 22,962 exclusive genes for all tissue comparison allowing
the association of different protein ID with different biological process and molecular function. Finally, we have to highlight that more information would be necessary in order to deepen in the processes related to vegetative growth of leaves or the reproductive development of the flower and fruit. In addition, the information provided in this study regarding genes not associated with proteins will require the future functional annotation in new studies.

methods

Plant material and RNA sequencing

Sample tissues of *Actinidia deliciosa* cv ‘Hayward’ were collected from Germán Greve Silva experimental station in Rinconada de Maipú (University of Chile). RNA was extracted using three biological replicates including shoots, leaves, flower buds, flowers and fruits at 7, 50, 120 and 160 after full bloom (DAFB) (Fig. 1) using a RNeasy mini kit (Qiagen®). To verify sample integrity, total RNA was evaluated on Fragment Analyzer TM Automated CE System (AATI) and quantified by Qubit® 2.0 Fluorometer, using QubitTM RNA BR Assay kit (Life Technologies, Carlsbad, CA, USA). For library construction, one microgram of RNA sample was used as input for Illumina® TruSeq RNA HT Sample prep kit, according with the manufacturer’s instructions. Final libraries were analyzed on a NGS kit for Fragment Analyzer TM, and quantified by Qubit® 2.0 Fluorometer, using Qubit TM DNA BR Assay kit. Libraries were sequenced on HiSeq 2500 platform (Illumina Inc).

Transcriptome assembly and refining

Total reads were analyzed (pre and post trimming) and trimmed with FASTQC software and FLEXBAR software (https://github.com/seqan/flexbar) in order to filter low quality reads (phred score less than 25) and remove reads either short or with
Ns on the sequence. Remaining reads after trimming were used as input in the transcriptome assembly using Trinity software (https://github.com/trinityrnaseq).

The refinement of the transcriptome was made using CD-HIT (http://weizhongli-lab.org/cd-hit/) to remove duplicate contigs with a 0.9 identity setting and Corset (https://github.com/Oshlack/Corset/wiki) to filter out contigs with less than 30 reads mapping to each contig in the assembly.

**Assessment of assembly metric, quality and annotation**

Transcriptome metric measurement was performed by TrinityStats.pl software and quality assessment was performed by two ultra-conserved protein gene finder software (CEGMA and BUSCO; http://korflab.ucdavis.edu/datasets/cegma/; http://busco.ezlab.org/). Annotation was performed with DIAMOND, (https://github.com/bbuchfink/diamond) in BLASTx mode against TAIR10 protein database, reporting hits with minimum e-value of 0.001.

**Principal component analysis and sample similarity**

A principal component analysis (PCA) was conducted to evaluate in non-supervised way, the grouping and similarity of each sample and corresponding biological replicates, alongside with a sample similarity heatmap. Both graphs were constructed with the R-based PtR script ("—log2—CPM—center_rows—prin_comp 3" parameters) from Trinity analysis utilities using as input the counts matrix containing every sample and its replicates. For sample similarity spearman correlation coefficient was used.

**Differential expression analysis**

Raw counts of each contig were obtained with RSEM
(https://deweylab.github.io/RSEM/) using Bowtie2 aligner (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml) to align all the input reads to the de novo assembled transcriptome, then with all counts per tissue and replicate, it was merged into a raw counts matrix. From the raw counts matrix was calculated a TPM-TMM normalized matrix used for the study alongside the raw counts matrix.

Differential expression analysis was performed by using the R’s package edgeR from Bioconductor (http://bioconductor.org/packages/release/bioc/html/edgeR.html). Differentially expressed genes (DEGs) were used for further comparisons and downstream analyses.

abbreviations

BLAST: Basic local alignment search tool

bp: pair of basis

DAFB: Daysafter full bloom

DEGs: Differentially expressed genes

FDR: False discovery rate

Gbp: Giga base pairs

GO: Gene ontology

NGS: Next generation sequencing

Ns: Non-sequence

PCA: Principal component Analysis

RNA-seq: RNA sequencing

SRA: Short read archive database

TMM: Trimmed mean of M values

TPM: Transcripts per kilobase million
\( \bar{x} \): Mean value

declarations

Ethics declarations

Ethics approval and consent to participate

Not applicable

Consent for Publication

Not applicable

Availability of data and material

All raw reads in FASTQ format including paired-end and replicates during the current study are available in the NCBI Short Read Archive (SRA) database under bioproject number PRJNA564374 (https://www.ncbi.nlm.nih.gov/sra/PRJNA564374).

Competing interests

The authors declare that they do not have competing interests

Funding

This research was funded by the Chilean Ministry of Education through of Fondo de Fomento al Desarrollo Científico y Tecnológico (project: FONDEF D09i-1136).

Authors’ contributions

JAS designed and conducted the experiments, contributed to material collection, data analysis and drafted the manuscript. CVP contributed to RNA-seq computational analyses providing its technical expertise in bioinformatics. CJ contributed to material collection and provided its technical expertise in agronomics. PZ: contributed to RNA extractions and provided its technical expertise in molecular biology. PMG: collaborated as a supervisor contributing to the drafting and discussions. RF: collaborated as IP of the funder project supervising the data
analysis, revising the manuscript and contributing to discussions. CM: collaborated as supervisor of data analysis and revised the manuscript providing the support of the Center of Vegetal Biotechnology (CBV). All authors have read and approved the final manuscript.

Acknowledgements

This research is greatly grateful to Faculty of Agronomy (University of Chile) and the support of the Center of Vegetal Biotechnology (CBV) at Universidad Andrés Bello (Chile) for providing the computation resources needed to perform data analyses and transcriptome assembly. The authors offer grateful thanks to Seneca Foundation of the Region of Murcia for their collaboration through Juan A. Salazar in Murcia inside Saavedra Fajardo program (20397/SF/17) and the project “Breeding stone fruit species assisted by molecular tools” (19879/GERM/15).

Author information

Affiliations

Department of Plant Breeding, CEBAS-CSIC, PO Box 164, E-30100 Espinardo, Murcia, Spain.

Juan Alfonso Salazar* & Pedro Martínez-Gómez

Universidad Andrés Bello, Facultad Ciencias Biológicas, Centro de Biotecnología Vegetal and FONDAP Center for Genome Regulation, República 217, Santiago, Chile

Cristian Vergara-Pulgar* & Claudio Meneses

Universidad de Chile, Departamento de Producción Agrícola, Universidad de Chile, Santiago, Chile

Claudia Jorquera, Patricio Zapata & Rodrigo Infante

*Both authors contributed equally to this manuscript

Corresponding author
Correspondence to Juan Salazar (e-mail: jasalazar@cebas.csic.es)

references

1. Watanabe K, Takashi B, Shirato K. Chromosome Numbers in Kiwifruit (Actinidia deliciosa) and Related Species. Journal of the Japanese Society for Horticultural Science. 1990; 58:835–40.

2. Hopping ME. Flow cytometric analysis of Actinidia species. New Zealand Journal of Botany. 1994; 32:85–93.

3. Ollitrault-Sammarcelli F, Legave JM, Michaux-Ferriere N, Hirsch AM. Use of flow cytometry for rapid determination of ploidy level in the genus Actinidia. Scientia Horticulturae. 1994; 57:303–13.

4. Huang H, Ferguson AR. Actinidia in China: Natural Diversity, Phylogeographical Evolution, Interspecific Gene Flow and Kiwifruit Cultivar Improvement. Acta Horticulturae. 2007; 753:31–40

5. Zuo L, Wang Z, Fan Z, Tian S, Liu J. Evaluation of Antioxidant and Antiproliferative Properties of Three Actinidia (Actinidia kolomikta, Actinidia arguta, Actinidia chinensis) Extracts in Vitro. International Journal of Molecular Sciences. 2012; 13(5):5506–18.

6. Nishiyama I, Fukuda T, Shimohashi A, Oota T. Sugar and Organic Acid Composition in the Fruit Juice of Different Actinidia Varieties. Food Science and Technology Research. 2008; 14:67-73.

7. Huang S, Ding J, Deng D, Tang W, Sun H, Liu D, et al. Draft genome of the kiwifruit Actinidia chinensis. Nature Communications. 2013; 4: 2640.

8. Pilkington SM, Crowhurst R, Hilario E, Nardozza S, Fraser L, et al. A manually annotated Actinidia chinensis var. chinensis (kiwifruit) genome highlights the
challenges associated with draft genomes and gene prediction in plants. BMC Genomics. 2018; 19:1-19.

9. Li X, Qin H, Wang Z, Zhang Q, Liu Y, Xu P, Zhao Y, Fan S, Yang Y, Ai J. Differentially expressed genes in male and female flower buds of hardy kiwifruit (Actinidia Arguta (Sieb. et Zucc.) Planch. Ex MIQ.). Bangladesh Journal of Botany. 2015; 44(5): 909–915.

10. Tang W, Zheng Y, Dong J, Yu J, Yue J, et al. Comprehensive Transcriptome Profiling Reveals Long Noncoding RNA Expression and Alternative Splicing Regulation during Fruit Development and Ripening in Kiwifruit (Actinidia chinensis). Frontiers in Plant Science. 2016; 7:335.

11. Wu R, Wang T, Warren BAW, Allan AC, Macknight RC, Varkonyi-Gasic E. Kiwifruit SVP2 gene prevents premature budbreak during dormancy. Journal of Experimental Botany. 2017; 68:1071–1082.

12. Tang P, Zhang Q, Yao X. Comparative transcript profiling explores differentially expressed genes associated with sexual phenotype in kiwifruit. PLoS One. 2017; 12, e0180542.

13. Wang Z, Liu Y, Li D, Li L, Zhang Q, Wang S, Huang, H. Identification of Circular RNAs in Kiwifruit and Their Species-Specific Response to Bacterial Canker Pathogen Invasion. Frontiers in Plant Science. 2017; 8:413.

14. Michelotti V, Lamontanara A, Buriani G, Orrù L, Cellini A, Donati I, Vanneste JL, Cattivelli L, Tacconi G, Spinelli F. Comparative transcriptome analysis of the interaction between Actinidia chinensis var. chinensis and Pseudomonas syringae pv. actinidiae in absence and presence of acibenzolar-S-methyl. BMC Genomics. 2018; 19:585.

15. Man YP, Wang YC, Jiang ZW and Gong JJ. Transcriptomic analysis of pigmented
inner pericarp of red-fleshed kiwifruit in response to high temperature. Acta Horticulturae. 2015; 1096:215–219.

16. Li W, Liu Y, Zeng S, Xiao G, Wang G, Wang Y, Peng M, Huang H. Gene Expression Profiling of Development and Anthocyanin Accumulation in Kiwifruit (Actinidia chinensis) Based on Transcriptome Sequencing. PLoS One. 2015; 10, e0136439.

17. Li Y, Fang J, Qi X, Lin M, Zhong Y, Sun L. A key structural gene, AaLDOX, is involved in anthocyanin biosynthesis in all red-fleshed kiwifruit (Actinidia arguta) based on transcriptome analysis. Gene. 2018; 648:31–41.

18. Honaas LA, Wafula EK, Wickett NJ, Der JP, Zhang Y, Edger PP, et al. Selecting superior de novo transcriptome assemblies: Lessons learned by leveraging the best plant genome. PLoS One. 2016; 11:1-42.

19. Robinson MD, McCarthy DJ and Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26.

20. Mi H, Muruganujan A, Casagrande JT and Thomas PD. Large-scale gene function analysis with the PANTHER classification system. Nature Protocols. 2013; 8:8, 1551–66.

21. Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, et al. agriGO v2. 0: a GO analysis toolkit for the agricultural community. Nucleic acids research. 2017; 45(W1), W122-W129.

22. Guo R, Landis JB, Moore MJ, Meng A, Jian S, Yao X, Wang H. Development and Application of Transcriptome-Derived Microsatellites in Actinidia eriantha (Actinidiaceae). Frontiers in Plant Science. 2017; 8:1-13.

23. Veeckman E, Rutting T, Vandepoele K. Are We There Yet? Reliably Estimating
24. Fay JV, Watkins CJ, Shrestha RK, Litwiñiuk SL, Talavera Stefani LN, Rojas CA, Argüelles CF, Ferreras JA, Caccamo M, Miretti MM. Yerba mate (Ilex paraguariensis, A. St.-Hil.) de novo transcriptome assembly based on tissue specific genomic expression profiles. BMC Genomics. 2018; 19:891.

25. Pervaiz T, Lotfi A, Muhammad Salman H, Haifang J, Fang J. High Throughput Sequencing Advances and Future Challenges. Journal of Plant Biochemistry & Physiology. 2017;05.

26. Zhang JY, Huang SN, Mo ZH, Xuan JP, Jia XD, Wang G, Guo ZR. De novo transcriptome sequencing and comparative analysis of differentially expressed genes in kiwifruit under waterlogging stress. Molecular Breeding. 2015; 35:208.

27. Chen L, Ding X, Zhang H, He T, Li Y, Wang T, Li X, Jin L, Song Q, Yang S, Gai J. Comparative analysis of circular RNAs between soybean cytoplasmic male-sterile line NJCMS1A and its maintainer NJCMS1B by high-throughput sequencing. BMC Genomics. 2018; 19: 1-14.

28. Vining KJ, Romanel E, Jones RC, Myburg AA, Wesley-Smith J, et al. The floral transcriptome of Eucalyptus grandis. New Phytologist. 2014; 206: 1406-1422.

29. Song GQ, Gao X. Transcriptomic changes reveal gene networks responding to the overexpression of a blueberry DWARF AND DELAYED FLOWERING 1 gene in transgenic blueberry plants. BMC Plant Biol. 2017; 17.

30. Li J, Xu J, Guo QW, Wu Z, Zhang T, Zhang KJ, Cheng C. Yan, Zhu P. Yu, Lou QF, Chen JF. Proteomic insight into fruit set of cucumber (Cucumis sativus L.) suggests the cues of hormone-independent parthenocarpy. BMC Genomics.
31. Guidarelli M, Zubini P, Nanni V, Bonghi C, Rasori A, Bertolini P, Baraldi E. Gene expression analysis of peach fruit at different growth stages and with different susceptibility to Monilinia laxa. European Journal of Plant Pathology. 2014; 140: 503–513.

32. Vergara-Pulgar C, Rothkegel K, González-Agüero M, et al. De novo assembly of Persea americana cv. “Hass” transcriptome during fruit development. BMC Genomics. 2019; 20:108.

33. Salinero MC, Vela P, Sainz MJ. Phenological growth stages of kiwifruit (Actinidia deliciosa “Hayward”). Sci. Hort. 2009; 121:27–31.

tables

Table 1. De novo transcriptome metrics, quality and annotation summary.

| Metric                      | Value   |
|-----------------------------|---------|
| Total Contigs               | 142,025 |
| Trinity “Genes”             | 106,630 |
| N10 (bp)*                   | 2,641   |
| N30 (bp)*                   | 1,602   |
| N50 (bp)*                   | 1,133   |
| Largest contig (bp)         | 36,186  |
| Smallest contig (bp)        | 501     |
| Average contig (bp)         | 1,044   |
| GC (%)                      | 39.50   |
| Total nucleotides           | 148,229,664 |
| CEGMA Score (%)\(^1\)       | C: 56.8; F: 35.1; M: 8.1 |
| BUSCO Score (%)\(^2\)       | C: 69.9; F: 18.3; M: 11.8 |

\(^1\) CEGMA dataset used contains 248 eukaryotic proteins. C: Complete genes found from the dataset, F: Fragmented genes found from the dataset and M: Missing genes.

\(^2\) BUSCO dataset used contains 1440 plant specific proteins. C: Complete genes
found from the dataset, F: Fragmented genes found from the dataset and M: Missing genes.

N10, N30 and N50: minimum contig length needed to cover 10, 30 and 50% of the genoma respectively.

Table 2. Functional annotations and top hit species summary.

| Functional annotation | Total contigs annotated | Total contigs unannotated | Top hit species (%) |
|-----------------------|-------------------------|---------------------------|---------------------|
|                       |                         |                           | Vitis vinifera      |
|                       |                         |                           | 18.2%               |
|                       |                         |                           | Juglans regia      |
|                       |                         |                           | 5.8%                |
|                       |                         |                           | Coffea canephora    |
|                       |                         |                           | 4.3%                |
|                       |                         |                           | Actinidia genus     |
|                       |                         |                           | 1.5%                |

Figures

Figure 1

Tissues assayed in kiwifruit cv ‘Hayward’ including shoots leaves unfolded, leaves completely developed, flower buds, flowers and fruits.
Figure 2

(A) Principal Component Analysis (PCA) of the four development stages of study a
Figure 3

Venny diagram for differential expression genes between different plant tissues:
Figure 4

Panther GO annotations for each comparative plant tissues: shoot vs leaf, flower.

Supplementary Files

This is a list of supplementary files associated with the primary manuscript. Click to download.

Salazar et al. Figure S6.tif
Salazar et al. Supplementary Table S2.xlsx
Salazar et al. Figure S4.tif
Salazar et al. Figure S3.tif
Salazar et al. Figure S2.tif
Salazar et al. Figure S7.tif
Salazar et al. Figure S1.pdf
