Dynamic Changes of Ascorbic Acid Occurring During Fruit Development and Ripening of Actinida Latifolia and Their Associated Molecular Mechanisms Deciphered by Weighted Gene Co-Expression Network Analysis

Hui Xia  
Sichuan Agricultural University

Honghong Deng  
Sichuan Agricultural University

Rongping Hu  
Sichuan Agricultural University

Lijin Lin  
Sichuan Agricultural University

Jin Wang  
Sichuan Agricultural University

Qunxian Deng  
Sichuan Agricultural University

Kunfu Xu  
Sichuan Agricultural University

Zhihui Wang  
Sichuan Agricultural University

Xiulan Lv  
Sichuan Agricultural University

Dong Liang (✉ liangeast@sina.com)  
Sichuan Agricultural University

---

Research Article

**Keywords:** HPLC, PacBio SMRT sequencing, full-length transcriptome, modules, metabolic pathways

**DOI:** [https://doi.org/10.21203/rs.3.rs-133017/v1](https://doi.org/10.21203/rs.3.rs-133017/v1)

**License:** ☭️ ☪️ This work is licensed under a Creative Commons Attribution 4.0 International License.  
[Read Full License](https://creativecommons.org/licenses/by/4.0/)
Abstract

Background: Actinidia latifolia is an exceptional source with extremely high ascorbic acid (AsA) content. However, its transcriptome atlas is lacking and how AsA accumulates during fruit development and ripening of this special kiwifruit and its associated molecular mechanisms are still poorly understood.

Results: Herein, the dynamic changes of AsA content of six stages of A. latifolia fruit development and ripening determined by HPLC demonstrated a rapid increasing profile during the initial expansion stage with a peak around 60 days after flowering (1194.21±69.25 mg 100g⁻¹ FW), followed by a progressive, albeit not significant decrease tendency and reached the minimum levels (1028.76±31.19 mg 100g⁻¹ FW) at maturity. A high-quality full-length (FL) transcriptome of A. latifolia was successfully constructed by third-generation sequencing for the first time, comprising 326,926 FL non-chimeric reads, 15,505 coding sequences, 2882 transcription factors, 18,797 simple sequence repeats, 3328 long noncoding RNAs, and 231 alternative splicing events. Illumina RNA-seq in combination with weighted gene co-expression network analysis revealed a network module highly correlated (r=0.5, p=0.03) with AsA content. Gene co-expression in this network module was explained by its roles in protein processing in endoplasmic reticulum (ko04141), glycolysis/gluconeogenesis (ko00010), and carbon metabolism (ko01200). Moreover, the expression patterns of genes involved in AsA biosynthesis and metabolism validated by qRT-PCR exhibited a similar trend with AsA accumulation.

Conclusions: Overall, the dynamic changes of AsA content and associated key genes and enriched metabolic pathways were deciphered, which paves the way for genetic improvement that aims for development of kiwifruit with super-high AsA content.

Background

Ascorbic acid (AsA) or vitamin c (Vc) is a vital water-soluble low molecular weight antioxidant [1] exerting a central role in supporting plant growth and development [2] and responding to environmental stress [3]. In addition to plants, AsA is a multifunctional molecule maintaining human health such as preventing some chronic and complex disease (stroke, cancer, and scurvy) [4], and even has possible physiological effects on the severe coronavirus disease 2019 (COVID 2019) [5]. Human and other primates have lost the capability to synthesize AsA by themselves due to a lack of L-gulonolactone oxidase (GLO) gene, and are thus entirely dependent upon food supplies such as consuming plant products [6]. Horticulture plants, in particular fresh fruits and vegetables, are affluent sources of AsA for human nutrition satisfying daily requirements [7].

Kiwifruit, a member of the genus Actinidia Lindl. of the family Actinidiaceae, is an important fruit crop widely distributed throughout most of east Asia, with the center of evolution being Yangzi and Pearl rivers and ranges of China [8]. Kiwifruit has long been considered as “the king of fruits” because it has a remarkably high AsA content [8], and other multiple nutritional phytochemicals [9, 10]. When compared with other commercially important fruit crops, the AsA content in equal fresh weight (FW) of kiwifruit averaged approximately 1–2 times higher than that of orange, whilst it was 10 times higher than that of banana, and up to 15 times higher than that of apple [11, 12]. Within the genus, there is a great diversity in AsA content, but without doubt, A. latifolia is one of very few genotypes that are regarded as an exceptional source of extremely high AsA [13–16]. Accordingly, A. latifolia is a valuable source to improve the AsA content in other commercially important kiwifruit cultivars.

During the past decades, De novo assembly of transcriptome by the secondary-generation sequencing (NGS), Illumina RNA-sequencing (RNA-Seq) has become a cost-effective, accurate, and routine method. Nevertheless, NGS
generated relatively short reads, which poses a great challenge when it comes to assembly and annotation, especially without a well-annotated reference genome [17]. Recently, the advent of a third-generation sequence platform, PacBio long-read single-molecule real-time (SMRT) sequencing approach was released to overcome the limitation of short read sequence, and provided opportunities to generate reliable genome-wide full-length (FL) cDNA sequences eliminating the need for assembly [17]. Despite the importance and specificity of A. latifolia with extremely high content AsA [13–16], neither reference genome nor transcriptome of A. latifolia are not available, which was a major bottleneck in understanding functional genomics and molecular genetics of A. latifolia.

To our knowledge, no reporting regarding the accumulation pattern of AsA of this special cultivar, A. latifolia, has been made. To elucidate the mechanisms regulating AsA concentration, it is of utmost importance to monitor how AsA accumulates during fruit development and ripening processes of cultivars with different AsA contents, especially with extremely high content. In addition, much remains to be learned about the associated molecular mechanisms underlying the dynamic accumulation of AsA in A. latifolia. Herein, the Iso-seq protocol with PacBio SMRT sequencing-based de novo transcriptome was initiated with foremost emphasis to construct a high-quality reference FL transcriptome for A. latifolia. The dynamic changes of AsA content of six different stages of A. latifolia fruit development and ripening was then evaluated by high performance liquid chromatography (HPLC) to characterize the accumulation pattern. In addition, the associated transcriptional changes were quantified by second-generation Illumina paired end sequencing technology to completely elucidate the molecular mechanisms underlying. Furthermore, a weighted gene co-expression network analysis (WGCNA) of the differentially expressed genes (DEGs) identified was performed to decipher the network modules of genes highly correlated.

The paucity of A. latifolia genomic information means that the reference FL transcriptome atlas obtained in this research is vital for future genome annotation and studying gene function, especially related to some economically important traits. A thorough elucidation of AsA accumulation and associated gene expression and metabolic pathways of A. latifolia fruit development and ripening will be useful speeding up genetic or cultural improvement. Overall, the results of this study will be of paramount importance for improving AsA content of kiwifruit, and will allow extrapolation to other edible horticulture plants in the near future.

**Results**

**Morphological changes and ascorbic acid accumulation at six different stages of Actinidia latifolia fruit development and ripening**

Visual inspection of six different stages corresponding to 30, 60, 90, 120, 150, and 170 days after flowering (DAF) of A. latifolia fruit development and ripening is presented in Fig. 1a. The quantification of AsA content was evaluated by HPLC (Fig. 1b). Statistical evaluation results of ANOVA showed that the different stages of A. latifolia fruit development and ripening had no significant effect on the AsA content, however, it can be found that the initial expansion stage was characterized by a rapid increase of AsA content, reaching the maximum level at 60 DAF (1194.21 ± 69.25 mg 100 g⁻¹ FW). Later, a decrease at the veraison stage (60–90 DAF) was followed by a gradual increase in the early ripening stage (90–120 DAF). The AsA content showed a declining trend from 120 to170 DAF and reached the minimum levels (1028.76 ± 31.19 mg 100 g⁻¹ FW) at full maturity. The details of AsA content changes over the time course expressed as mean ± standard error (SE) in FW basis are shown in Fig. 1b.

**PacBio SMRT sequencing-based full-length transcriptome atlas of Actinidia latifolia fruit**
To capture as many transcripts of *A. latifolia* fruit as possible, a total of 18 RNA samples from six stages of *A. latifolia* fruit maturation were equally pooled together for library preparation (1–6 kb libraries). A total of 19.50 Gb subreads were obtained from offline data. The strict screening criteria (full passes greater than 1.0 and accuracy greater than 0.90) led to 358,138 CCS reads, comprising 743,399,287 read bases, with an average read length of 2,087 bp. After error correction, a total of 34,148 polished high-quality isoforms were obtained, among which, 91.80% (326,926 reads) were filtered as the FLNC reads of *A. latifolia* fruit transcriptome. The FLNC read length distribution ranged from 700 to 3500 bp, and its overall distribution of each size bin agreed with the size of its cDNA library (Fig. 2a).

**Functional annotation of Actinidia latifolia transcripts with multiple databases**

Transcripts were scanned against and successfully annotated by Nr, GO, COG, KOG, eggNOG, KEGG, Pfam, and Swiss-Prot databases, whose integrated alignment results are summarized in Table S1. Based on Nr functional annotation, the best blast hit of homologous species with *A. latifolia* fruit was *Vitis vinifera* (5689 isoforms, 31.33%), followed by *Sesamum indicum* (1260, 6.73%) and *Theobroma cacao* (1082, 5.78%) (Fig. 2b). A total of 15,113 transcripts were successfully annotated with GO database and classified into three categories, including biological process (BP), cellular component (CC) and molecular function (MF). The genes under cellular process (8291 matched genes, 54.86%), metabolic process (8072, 53.41%), and single-organism process (5540, 36.66%) were highly represented in the BP category. Cell (8924 matched genes, 59.05%) was the most abundant subcategory within CC, followed by cell part (8911, 58.96%) and organelle (6340, 41.95%). In the category of MF, catalytic activity (7823 matched genes, 51.76%) was the most prominently represented, followed by binding (7672, 50.76%), transporter activity (1023, 6.77%) (Fig. 2c).

**Structural analysis of the full-length transcriptome of Actinidia latifolia**

A total of 15,505 CDS were identified. The frequencies for each length of CDS were evaluated with the most frequent length ranging from 100 to 1200 bp (Fig. S1a). Furthermore, a total of 2436 genes were predicted to be TFs by predicting non-redundant transcript via iTAK software. The TFs were classified into different families, among which the most abundant type identified was C3H (84 matched genes), followed by GRAS (76), AP2/ERF-ERF (68), MYB-related (56), B3-ARF (56), CAMK_CDPK (55), C2H2 (54), and RLK_Pelle_LRR_Xl-1 (51) (Fig. S1b). A total of 3328 IncRNA candidates were identified, and 447, 445, 2459, and 1332 of them were identified by CNCI, CPC, Pfam and CPAT prediction results, respectively. Comparison results proved that 169 transcripts were simultaneously identified by the four computational approaches (Fig. 2d). In addition, transcripts were subjected to SSR analysis via MISA, and a total of 18,797 SSR including seven SSR types (i.e., mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, hexanucleotide and compound nucleotides) were detected. The dinucleotide SSR loci exhibited the highest frequency, followed by mononucleotide and the compound SSR types (Fig. S1c). A detailed breakup showing SSR types from PacBio is summarized in Table S2.

**Differentially expressed genes identified in the comparative Illumina transcriptome analysis**

The non-redundant transcripts obtained above was used a reference for Illumina sequence alignment and subsequent analysis. A total of 18 cDNA libraries for Illumina sequencing were generated from six different stages, with three biological replicates per stage. DEG identification was independently performed using a pair-wise comparison between developmental stage and the baseline control (DAF30). Based on the adopted cutoff (FDR <
0.01 and absolute fold change ≥ 2), the numbers of all genes displaying both significantly up- and downregulated profiles are presented in Fig. 3. By using FPKM values, a total of 3788, 3485, 4315, 5265 and 6079 DEGs were identified in the comparisons of DAF30_vs_60, DAF30_vs_90, DAF30_vs_120, DAF30_vs_150, and DAF30_vs_170 (the former is used as a baseline control, and the latter is the treatment group), respectively. The highest number of DEGs was identified in the comparison of DAF30_vs_170 (baseline_vs_ripening), comprising 2632 upregulated and 3447 downregulated DEGs (Fig. 3).

Functional annotation and categorization of differentially expressed genes

The annotation, pathway and functional categorization of the DEGs were analyzed based on COG, GO, KEGG, KOG, Pfam, Swiss-Prot, eggNOG and Nr databases. Assignments indicated that at least 97.79% DEGs were functionally annotated at multiple databases (Table 1). Analysis of GO categories showed that the functional distribution of DEGs in the comparable groups was similar. In detail, the BP category of GO terms were primarily grouped into cellular process, metabolic process and single-organism process. The CC category were mainly assigned to cell, cell part and membrane. Catalytic activity, binding and transporter activity were prominent in the MF category (Fig. S2). A pathway-based categorization of orthologous genes, KEGG pathway enriched analysis, was carried out to predict the functional profiles and biological significances of DEGs identified during the six different stages of *A. latifolia* fruit development and ripening. Plant hormone signal transduction, carbon metabolism, biosynthesis of amino acids, starch and sucrose metabolism, and protein processing in endoplasmic reticulum were the most significantly enriched metabolic pathways (Fig. S3).

| DEG_set<sup>a</sup> | Annotated<sup>b</sup> | COG<sup>c</sup> | GO<sup>d</sup> | KEGG<sup>e</sup> | KOG<sup>f</sup> | Pfam<sup>g</sup> | Swiss-Prot<sup>h</sup> | eggNOG<sup>i</sup> | Nr<sup>j</sup> |
|-------------------|---------------------|----------------|-------------|----------------|--------------|----------------|------------------|----------------|-----------|
| DAF30_vs_DAF120   | 4,232               | 1,993          | 3,435       | 1,687          | 2,207        | 3,823          | 3,474            | 4,142          | 4,209     |
| DAF30_vs_DAF150   | 5,176               | 2,477          | 4,187       | 2,093          | 2,771        | 4,663          | 4,232            | 5,093          | 5,156     |
| DAF30_vs_DAF170   | 5,991               | 2,832          | 4,841       | 2,386          | 3,305        | 5,383          | 4,807            | 5,893          | 5,964     |
| DAF30_vs_DAF60    | 3,718               | 1,736          | 2,940       | 1,455          | 2,053        | 3,306          | 2,983            | 3,664          | 3,709     |
| DAF30_vs_DAF90    | 3,405               | 1,542          | 2,704       | 1,234          | 1,760        | 3,045          | 2,756            | 3,326          | 3,381     |

<sup>a</sup>DEG: differentially expressed gene.  
<sup>b</sup>Annotated: number of differentially expressed transcripts annotated.  
<sup>c</sup>COG: clusters of orthologous groups.  
<sup>d</sup>GO: gene ontology.  
<sup>e</sup>KEGG: kyoto encyclopedia of genes and genomes.  
<sup>f</sup>KOG: eukaryotic orthologous groups.  
<sup>g</sup>Pfam: protein family.  
<sup>h</sup>Swiss-Prot: a manually annotated and reviewed protein sequences database.  
<sup>i</sup>eggNOG: evolutionary genealogy of genes, non-supervised orthologous groups.  
<sup>j</sup>Nr: non-redundant protein sequence database.

Identification of key network modules associated with AsA content by weighted gene co-expression network analysis

To further unveil the crucial regulatory genes associated with AsA accumulation during *A. latifolia* fruit development and ripening, we performed WGCNA with all DEGs identified. The co-expression network was constructed, and the genes with similar expression patterns were clustered together into eight distinct modules with the average linkage
hierarchical clustering (Fig. 4a). Eight different colors including blue, brown, cyan, yellow, salmon, red, tan and grey were used to represent distinct modules, containing 764, 2688, 316, 212, 77, 993, 80, and 27 transcripts, respectively (Table S3). WGCNA also allowed the assessment of the overall correlation of the modules with AsA content. The blue and brown modules that possibly related to AsA content were found (Fig. 4b). Analysis of the module and AsA content relationship revealed the blue module ($r = 0.5, p = 0.03$) was significantly ($p < 0.05$) associated with AsA content during *A. latifolia* fruit development and ripening (Fig. 4c). In the following analysis, the gene significance and module membership of the blue (cor = 0.7, p = 1.6e-113), brown (cor = 0.019, p = 0.32), cyan (cor = 0.2, p = 0.00035), yellow (cor = 0.14, p = 0.042), red (cor = 0.45, p = 1.1e-50), salmon (cor = 0.031, p = 0.79), and tan (cor = -0.25, p = 0.025) modules were calculated (Fig. S4).

The highest association in the module-trait relationship was found between blue module and AsA content, followed by brown module (Fig. 4b). Therefore, these two modules were selected as module of interest and studied in subsequent biological and functional interpretations (Fig. 5). Other modules without statistical significance with AsA content were not further considered. To evaluate the biological significance of module of interest, KEGG pathway enrichment analysis was performed. The blue module was enriched with genes associated with protein processing in endoplasmic reticulum (ko04141), glycolysis/gluconeogenesis (ko00010), and carbon metabolism (ko01200) (Fig. 5a). However, genes associated with ascorbate and aldarate metabolism (ko00053) was only enriched in the brown module, and the absolute value of gene significance in the brown module enriched with genes involved in ascorbate and aldarate metabolism (ko00053) is generally lower than that in several pathways including plant hormone signal transduction (ko04075), starch and sucrose metabolism (ko00500), phenylpropanoid biosynthesis (ko00940), and alpha-linolenic acid metabolism (ko00592) (Fig. 5b).

**Expression patterns of genes involved in ascorbic acid biosynthesis during fruit development and ripening of *Actinidia latifolia* verified by qRT-PCR analysis**

To investigate the molecular mechanism regulating AsA biosynthesis in *A. latifolia* fruit development and ripening, the expression levels of AsA biosynthesis genes in different developmental stages were further examined by qRT-PCR, as illustrated in Fig. 6. L-galactose pathway, L-gulose, myo-inositol and D-galacturonic acid pathways were identified. The overall gene expression levels of eight enzymes, including L-galactose dehydrogenase (GalDH), GDP-L-galactose phosphorylase (GGP), GDP-mannose-3',5'-epimerase (GME), L-galactose-l-phosphate phosphatase (GPP), GDP-mannose pyrophosphorylase (GMP), phosphoglucoisomerase (PGI), myo-inositol oxygenase (MIOX), and ascorbate peroxidase (APX), exhibited a similar trend with the cumulative content of AsA described above (Fig. 1). The L-galactose pathway and myo-inositol pathway were the predominant pathways regulating the high AsA content in *A. latifolia* fruit development and ripening (Fig. 6).

**Discussion**

The first high-quality functionally annotated reference transcriptome for *Actinidia latifolia*

*A. latifolia* has been identified as a promising kiwifruit species containing remarkably high concentration of AsA [13–16]. The lack of a high-confidence transcriptome atlas of *A. latifolia* greatly hindered scope of investigation into the molecular genetic basis of this important cultivar. In this study, we successfully built the high-quality transcriptome for *A. latifolia* by PacBio SMRT sequencing method for the first time (Fig. 2), which will be a crucial source towards exploring genome mining and understanding gene functions.
The reference transcriptome or draft genome data of *A. latifolia* has not been completely sequenced yet. It is crucial to annotate transcripts to biological functions and metabolic pathways. Therefore, sequence-based alignments against eight functional databases were conducted (Fig. 2) for understanding high-level functions and utilities of the biological systems. Another important aspect of our study was to predict the CDS, TF, and lncRNA from the non-redundant transcripts of *A. latifolia* (Fig. S1). LncRNA represents a novel class of non-coding RNA but plays a regulatory role in a range of biological process, such as plant growth and development and stress responses [18]. In the present study, IncRNA was identified by a combination of CPC, CNCI, Pfam, and CPAT databases (Fig. 2d), which will be useful for subsequent studies of biological functions of lncRNA in *A. latifolia*.

The dinucleotide SSR type was the most frequent in *A. latifolia*, which was in agreement with the expressed sequence tags (EST) derived SSR distribution reported in *Actinidia* species previously [19]. SSR markers have proved to be the most favored molecular markers to estimate genetic diversity, phylogenetic relationships, genotype identification and discrimination, marker-phenotype association, and genetic map construction [20]. FL transcriptome atlas contains a large amount of genetic information and is a rich source for SSR discovery [21]. Transcriptome-based SSR development have increased potential for association with functional genes or even agronomic phenotypes because of the close linkage to expressed genes of transcriptome [22]. This study reported the use of PacBio SMRT sequencing method for discovery of a set of SSR loci in *A. latifolia* for the first time (Table S2). The SSR identified in this study will contribute a valuable resource facilitating marker-assisted breeding of *A. latifolia*.

**Dynamics of Ascorbic acid content accompanying fruit development and ripening of *Actinidia latifolia***

AsA plays a plethora of roles in biological functions of both plants and humans [2–5]. Beside the cultivar-dependent differences [13–16], kiwifruit also shows tissue- and developmental-specific variability in AsA content [23]. The highest accumulation of AsA content was recorded at four to seven weeks after anthesis in *A. deliciosa*, *A. chinensis*, and *A. eriantha* [23]. Li et al. demonstrated a similar pattern of ascorbate accumulation and associated gene expression profile during fruit development of *A. chinensis var. deliciosa* ‘Qinmei’ [24]. Zhang et al. reported that *A. chinensis var. chinensis* ‘Hongyang’ kiwifruit exhibited a maximal ascorbate level at its immature green stage due to the high biosynthesis rate, it decreased as it ripened and then remained fairly stable until complete ripening [25].

In the present study, we quantified the dynamic changes of AsA by HPLC occurring during different stages of fruit development and ripening in *A. latifolia* (Fig. 1). Although no significant changes of AsA content were detected by Tukey's HSD test accompanying fruit development and ripening of *A. latifolia*, there is a rapid increasing profile during the initial expansion stage with a peak around the 60 days after anthesis, followed by a progressive, albeit not significant decrease tendency (Fig. 1). Overall, the accumulation dynamic of AsA in *A. latifolia* positively correlated with fruit developmental stages, as in other kiwifruit varieties [23–25]. On the other hand, our results demonstrated the extremely high levels of AsA content of *A. latifolia* (1028.76 ± 31.19 mg 100 g⁻¹ FW, Fig. 1b), which was consistent with previous studies [13–16].

**An elucidation of the molecular mechanisms regulating AsA accumulation of *Actinidia latifolia***

Ascorbate biosynthesis and metabolism is a complex set of reactions dependent on the co-expression and coordination of a cluster of genes [1, 4, 12]. It is accepted that the genes associated with the same metabolic function probably exhibit co-expression patterns [26]. The versatility of co-expression network analysis for inferring gene function have extensively been supported [26–29], although gene co-expression does not necessarily mean a
direct regulatory relationship [30]. WGNCA is a powerful algorithm to construct free-scale gene co-expression networks by comprehensively capturing the relationships between gene sets or between gene sets and complex traits [27, 28]. Therefore, in the present study, we quantified the global gene expression accompanying fruit development and ripening of *A. latifolia*. (Fig. 3, Table S4) and presented a WGCNA approach using these 18 cDNA libraries data to identify key functional genes involved in AsA accumulation of *A. latifolia* (Fig. 4–5). By classifying gene expression pattern in this large datasets here, we demonstrated that these hub genes in blue and brown modules were associated with AsA accumulation in *A. latifolia* fruit development and ripening (Fig. 4).

In plants, at least four distinct metabolic pathways, including the L-galactose, L-gulose, galacturonic, and myo-inositol pathways, constructed a complex network for AsA biosynthesis [1]. The well represented pathways of AsA biosynthesis in *A. latifolia* discovered in this study included the L-galactose, and galacturonic, myo-inositol pathway (Fig. 6). In theory, the changes of AsA contents should be closely related to the gene expression of the AsA biosynthesis pathway [30]. However, the co-expression of the genes in the blue module was explained by their roles in protein processing in endoplasmic reticulum (ko04141), glycolysis/gluconeogenesis (ko00010), and carbon metabolism (ko01200) (Fig. 5a). An intriguing interpretation is that the carbohydrates such as sucrose, glucose, mannose, galactose, myo-inositol, pectin polymers, and cellulose are employed as the substrates for biosynthesis of AsA [1, 12, 30]. For example, the initial substrate for the L-galactose pathway is a glucose molecule [1]. The content of soluble carbohydrates such as the glucose, fructose and sucrose regulates the AsA content emerged in tomato fruits [31].

Previous studies have demonstrated the correlation between GMP transcripts level and the AsA accumulation rate to some extent [23, 32]. In this study, the GMP transcript level tended to higher in the young fruits, and had a similar trend with the AsA accumulation occurring early in fruit developmental stages (Fig. 6). GMP plays an important role in the cell wall biosynthesis and protein glycosylation. Therefore, it has been postulated that those two processes also are involved in the high GMP transcript level in young fruits [32]. Transformation of plants with over-expression of the GGP gene in *Arabidopsis* resulted in more than 4-fold increment in AsA content [33]. As an early committed step of L-galactose biosynthesis pathway, GGP contributed greatly to the rapid increase of AsA accumulation in several fruit species [23, 34, 35]. GGP has been considered as the best regulatory control point of AsA biosynthesis [36]. In the present study, GGP had relatively high expression level in the early stage of fruit development (Fig. 6), evidencing the contribution of GGP for the AsA biosynthesis. In addition, the gene expression level of GME was also correlated with AsA content (Fig. 6), emphasizing the central role and positive regulation of GME for AsA accumulation in immature fruit [23]. In addition to the three upstream enzymes, we observed the positive correlation of GalDH, GPP, PGI, MIOX, and APX and cumulative content of AsA (Fig. 6), suggesting that the high co-expression of these enzymes resulted in the high AsA accumulation in *A. latifolia*.

The regulation pathways of AsA content accompanying with fruit development and ripening is characteristic of high complexity and often integrate a cluster of transcriptional and post-transcriptional regulatory processes, and feedback inhibition [1, 12, 37]. Therefore, a plausible explanation for the observed decrease of AsA content towards full maturity (Fig. 1) is the feedback inhibition. Overall, the current study provides insights into the molecular mechanisms regulating AsA accumulation of *A. latifolia* fruit, which are expected to be highly useful for breeding cultivars with super-high AsA content in the future.

**Conclusions**
The lack of comprehensive genome sequence information limits the scope of investigation into the molecular genetic basis of *Actinidia latifolia* with extremely high ascorbic acid content. Here, we used PacBio SMRT sequencing technology to generate a high-quality functionally annotated reference transcriptome for this special kiwifruit species. This is the first *A. latifolia* full-length transcriptome release covering the fruit tissues extracted from six stages of fruit development and ripening, which will be of great importance for both basic and applied research on biotechnology assays and genetic improvement in the future. Based on the transcriptome, pair-wise comparisons between developmental stage and the base control (DAF30) were performed, and differentially expressed genes (DEGs) were detected. Co-expression network modules highly correlated with ascorbic acid content accompanying during fruit development and ripening of *A. latifolia* was built by a weighted gene co-expression network analysis (WGCNA). Functional annotation revealed the particularly enriched KEGG pathways for the two modules that had the highest association in the module-trait relationship. In addition, the expression patterns of genes involved in ascorbic acid biosynthesis and metabolism were further validated by qRT-PCR, which was an important explanation for the detected high AsA content in *A. latifolia* fruit. Our study provides insights into the molecular mechanism regulating AsA accumulation of *A. latifolia* fruit, which are expected to be highly useful for breeding cultivars with super-high AsA content in the future.

**Methods**

**Plant materials and sampling**

This study was conducted using five-year-old (in 2019) bearing *A. latifolia* kiwifruit trees in an experimental block of Sichuan Provincial Academy of Natural Resource Sciences, Deyang city, Sichuan Province, China (N31°30′, E104°23′). Fruit samples were randomly collected from the fruiting branches (5 per tree) of six different trees for AsA content determination and gene expression analysis. Fruit sampling included six representative stages of *A. latifolia* fruit development and ripening corresponding to 30, 60, 90, 120, 150, and 170 days after flowering (DAF). In all cases, three different biological replicates per stage were obtained with ten fruits in each replication. Samples were stored in a cold chamber and transported to the laboratory within 2.0 h. Upon arrival laboratory, samples were cut into slices and immediately frozen in liquid nitrogen and kept at -80 °C until for sample preparation.

**Chemicals and solvents**

Authentic standards of AsA and oxalic acid were of HPLC grade and purchased from Beijing Solaribio Sciences & Technologies Co. Ltd. (Beijing, China). Ultrapure water with an electrical resistivity of 18.2 MΩ cm was prepared using a Milli-Q Gradient water purification system (Millipore corporation, Bedford, MA, USA) via a 0.22 µm filter, and this purified water was used for preparing all solutions.

**Determination of AsA using HPLC coupled with UV detection**

The determination of AsA content was executed by HPLC coupled with UV detection using the methods described by [38] with minor modifications. In brief, a portion of 2.0 g of frozen samples was fully grounded with 5.0 ml 0.1% oxalic acid to homogenous slurry and made it up until 25.0 ml of total volume. The extracts obtained were filtered through a 0.45 µm membrane and thus were ready for injection in HPLC system for quantifying AsA content. AsA standard solution (1.0 mg/ml) was prepared by dissolving 25.0 mg AsA in 0.1% oxalic acid and diluted to 25.0 ml with the same solvent. HPLC analysis employed an Agilent 1260 HPLC instrument and a variable wavelength detector (VWD) (Agilent, Santa, Clara, CA, U.S.A). The chromatographic separation was carried out on a Zobax Stablebond Analytical SB-C18 column (250.0 × 4.6 mm, 5.0 µm). The mobile phase used was a 0.1% solution of
oxalic acid. The flow rate of mobile phase was kept constant at 1.0 ml/min and 10.0 µl injection volume of samples and standard were used in quantitative analysis. AsA was monitored at a wavelength of 265.0 nm, and quantified from calibration curve.

**Assays of APX, DHAR, and MDHAR activities**

Enzyme activity were prepared and detected according to our previously published method [39].

**RNA extraction and quality evaluation**

Total RNA was extracted using a PureLink RNA Mini Kit (Invitrogen Inc., Carlsbad, CA, USA) and purified with an on-column PureLink DNase treatment (Invitrogen) according to the manufacturer's instructions. RNA purity was determined by A260 absorbance using a Nanodrop Spectrophotometer (Thermo Scientific NanoDrop 2000a). RNA concentration was quantified by a Qubit 2.0 Fluorometer (Invitrogen) and integrity was assessed using an RNA Nano 6000 Assay Kit on an Agilent Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA, USA). The samples were sent to Biomarker Technologies Co. Ltd (Beijing, China) for sequencing using PacBio- and Illunina RNA-Seq technology.

**Construction of Iso-seq cDNA library and PacBio sequencing**

The mRNA was enriched using oligo-dT magnetic beads from 4.0 µg total RNA and reverse transcribed into cDNA using the SMARTer™ PCR cDNA Synthesis Kit (Clontech, now Takara, http://www.takarabio.com). The size-selected cDNA library was constructed according to the BluePippin Size Selection System protocol as described by PacBio (PN 100-092-800-03) and sequenced on the PacBio Sequel platform.

**Reads processing and error collection of BacBio Iso-seq reads**

Row data acquired after SMRT sequencing were firstly processed using SMRTlink v5.0 software. The circular consensus sequence (CCS) reads were yield from subread BAM files, and the full-length non-chimeric (FLNC) reads and non-full-length reads were determined by the simultaneous presence of the poly-A tail signal and the 5’ and 3’ cDNA primers from reads of insert (ROIs). The short reads were discarded. Subsequently, FLNC sequences were isoform-level clustered with iterative clustering and error correction (ICE) Quiver algorithm and herein generated one consensus isoform [40]. The non-full-length CCSs were polished with the Quiver algorithm. Finally, isoform with a minimum Quiver accuracy of 0.99 was considered high quality isoform and used for subsequent analyses.

**Gene functional annotation**

All isoforms were subjected to functional annotations using multiple protein and nucleotide databases, including the National Center for Biotechnology Information (NCBI) non-redundant protein (Nr, cutoff E-value ≤ 1e⁻⁵) [41], gene ontology (GO, E-value ≤ 1e⁻¹⁰) [42], kyoto encyclopedia of genes and genomes (KEGG, E-value ≤ 1e⁻³) [43], clusters of orthologous groups (COG, E-value ≤ 1e⁻³) [44], eukaryotic orthologous groups (KOG, E-value ≤ 1e⁻³) [45], protein family (Pfam, E-value ≤ 0.01) [46] and a manually annotated and reviewed protein sequences database (Swiss-Prot, E-value ≤ 1e⁻⁵) [47].

**Transcript structure analysis**

Potential coding sequence (CDS) regions within transcripts were predicted by TransDecoder (https://github.com/TransDecoder/TransDecoder/releases). Simple sequence repeats (SSRs) within transcriptome were identified by MISA program (http://pgrc.ipk-gatersleben.de/misa/). LncRNA was screened via coding-non-coding-index (CNCI) with default parameters [48] and coding potential calculator (CPC) with NCBI eukaryotes'
protein database (E-value < $10^{-10}$) [49]. Each transcript was translated in three possible frames, and Pfam Scan with default parameters of -E 0.001 --domE 0.001 was utilized to determine whether there exists a domain of known protein family. Transcription factors (TFs) were predicted with iTAK software from putative protein sequences [50].

**Illumina cDNA library construction and second-generation sequencing for transcriptome of fruit development and ripening stages**

A total of 18 cDNA libraries (6 representative stages × 3 biological replicates) were constructed and used for second-generation high-throughput sequencing. The RNA extraction and quality detection, cDNA synthesis, library preparation, high-throughput sequencing, identification of DEGs, functional categorization, and pathway analysis of DEGs followed our previously published protocol [9].

**Validation of DEGs by quantitative reverse transcription PCR (qRT-PCR)**

Specific primers for qRT-PCR are presented in Table S5, which were designed using the Primer Premier 5.0 software (Premier Biosoft, Palo Alto, CA, U.S.A.). The qRT-PCR experiment and data normalization were performed as described by [9].

**Weighted gene co-expression network analysis (WGCNA) for construction of modules**

The R package for WGCNA was applied for co-expression network analysis to identify modules where genes showed high correlations [27, 28]. The adjacency matrix was performed on the basis of the normalized fragments per kilobase of transcript per million mapped reads (FPKM) value over 0.1 and variation of FPKM over 0.05. The scale-free topology criterion was set as a power of $\beta = 7$. The adjacency matrix was then transformed to a topological overlap matrix (TOM) that was used as input for the hierarchical clustering analysis, and its corresponding dissimilarity was calculated. The transcripts with similar expression trends were grouped into one module with a minimum module size cutoff of 30, hierarchical cluster tree was built based on the dynamic hybrid tree cut algorithm, and highly correlated (or hub) genes for different modules were calculated. The minimum height cut-off for merging modules was set as 0.26435.

**Declarations**

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Availability of data and materials**

All data generated or analysed during this study are included in this published article and its supplementary information files.

**Competing interests**
The authors declare that they have no competing interests.

**Funding**

This work was supported by the Sichuan Science and Technology Program (2016NZ0105 and 2019YFSY0041).

**Authors’ contributions**

H.X. and H.D. contributed equally to this work. D.L. and H.X. conceived and designed the experiments. H.D. performed the bioinformatics analysis and wrote the paper. H.X., R.H., L.L., and J.W. collected the plant materials and performed the RNA extraction and purification. H.X., H.D., Q.D., K.X., Z.W., and X.L. performed the HPLC analysis and did the assays of enzyme activities. All authors have read and approved the final manuscript.

**Acknowledgements**

Not applicable.

**References**

1. Mellidou I, Kanellis AK. Genetic control of ascorbic acid biosynthesis and recycling in horticultural crops. Front Chem. 2017;5:50.
2. Gallie DR. L-Ascorbic Acid: A multifunctional molecule supporting plant growth and development. Scientifica (Cairo). 2013; doi:10.1155/2013/795964.
3. Akram NA, Shafiq F, Ashraf M. Ascorbic acid-a potential oxidant scavenger and its role in plant development and abiotic stress tolerance. Front Plant Sci. 2017;8:613.
4. Granger M, Eck P. Dietary vitamin C in human health. Advances in food and nutrition research. 2018; doi: 10.1016/bs.afnr.2017.11.006.
5. Miranda-Massari JR, González MJ, Marcial-Vega VA, Soler JD. A possible role for ascorbic acid in COVID-19. J Restor Med. 2020;9:1–7.
6. Bvenura C, Sivakumar D. The role of wild fruits and vegetables in delivering a balanced and healthy diet. Food Res Int. 2017;99:15–30.
7. Wallace TC, Bailey RL, Blumberg JB, Burton-Freeman B, Chen Cy O, Crowe-White KM, et al. Fruits, vegetables, and health: A comprehensive narrative, umbrella review of the science and recommendations for enhanced public policy to improve intake. Crit Rev Food Sci. 2020;60:2174–211.
8. Huang S, Ding J, Deng D, Tang W, Sun H, Liu D, et al. Draft genome of the kiwifruit Actinidia chinensis. Nat Commun. 2013;4:2640.
9. Liang D, Deng H, Deng Q, Lin L, Lv X, Wang J, et al. Dynamic changes of phenolic compounds and their associated gene expression profiles occurring during fruit development and ripening of the Donghong kiwifruit. J Agric Food Chem. 2020;68:11421–33.
10. Xia H, Wang X, Su W, Jiang L, Lin L, Deng Q, et al. Changes in the carotenoids profile of two yellow-fleshed kiwifruit cultivars during storage. Postharvest Biol Technol. 2020;164:111162.
11. Vissers MCM, Carr AC, Pullar JM, Bozonen SM. The bioavailability of vitamin C from kiwifruit. Advances in food and nutrition research. 2013; doi: 10.1016/B978-0-12-394294-4.00007-9.
12. Bulley SM, Laing W. Ascorbic acid-related genes. The Kiwifruit genome, compendium of plant genomes. 2016; doi: 10.1007/978-3-319-32274-2_13.

13. Du G, Li M, Ma F, Liang D. Antioxidant capacity and the relationship with polyphenol and vitamin C in Actinidia fruits. Food Chem. 2009;113:557–62.

14. Rassam M, Laing W. Variation in ascorbic acid and oxalate levels in the fruit of Actinidia chinensis tissues and genotypes. J Agric Food Chem. 2005;53:2322–6.

15. Nishiyama I, Yamashita Y, Yamanaka M, Shimohashi A, Fukuda T, Oota T. Varietal difference in vitamin C content in the fruit of kiwifruit and other Actinidia species. J Agric Food Chem. 2004;52:5472–5.

16. Latocha P, Krupa T, Wołosiak R, Worobiej E, Wilczak J. Antioxidant activity and chemical difference in fruit of different Actinidia species. Int J Food Sci Nutr. 2010;61:381–94.

17. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. Genome Biol. 2020;21:1–16.

18. Wu L, Liu S, Qi H, Cai H, Xu M. Research progress on plant long non-coding RNA. Plants. 2020;9:408.

19. Fraser LG, Harvey CF, Crowhurst RN, De Silva HN. EST-derived microsatellites from Actinidia species and their potential for mapping. Theor Appl Genet. 2004;108:1010–6.

20. Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK. Microsatellite markers: An overview of the recent progress in plants. Euphytica. 2011;177:309–34.

21. Deng K, Deng R, Fan J, Chen E. Transcriptome analysis and development of simple sequence repeat (SSR) markers in Zingiber striolatum Diels. Physiol Mol Biol Plants. 2018;24:125–34. doi: 10.1007/s12298-017-0485-0.

22. Taheri S, Abdullah TL, Yusof MR, Hanafi MM, Sahebi M, Azizi P, et al. Mining and development of novel SSR markers using Next Generation Sequencing (NGS) data in plants. Molecules. 2018;23:399.

23. Bulley SM, Rassam M, Hoser D, Otto W, Schüinemann N, Wright M, et al. Gene expression studies in kiwifruit and gene over-expression in Arabidopsis indicates that GDP-L-galactose guanyltransferase is a major control point of vitamin C biosynthesis. J Exp Bot. 2009;60:765–78.

24. Li M, Ma F, Liang D, Li J, Wang Y. Ascorbate biosynthesis during early fruit development is the main reason for its accumulation in kiwi. PLos One. 2010;5:e14281.

25. Zhang JY, Pan DL, Jia ZH, Wang T, Wang G, Guo ZR. Chlorophyll, carotenoid and Vitamin C metabolism regulation in Actinidia chinensis “Hongyang” outer pericarp during fruit development. PLoS One. 2018;13:1–17.

26. El-Sharkawy I, Liang D, Xu K. Transcriptome analysis of an apple (Malus × domestica) yellow fruit somatic mutation identifies a gene network module highly associated with anthocyanin and epigenetic regulation. J Exp Bot. 2015;66:7359–76.

27. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol. 2005;4.

28. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.

29. Zhang L, Zhang Q, Li W, Zhang S, Xi W. Identification of key genes and regulators associated with carotenoid metabolism in apricot (Prunus armeniaca) fruit using weighted gene coexpression network analysis. BMC Genomics. 2019;20.

30. Gao C, Ju Z, Li S, Zuo J, Fu D, Tian H, et al. Deciphering ascorbic acid regulatory pathways in ripening tomato fruit using a weighted gene correlation network analysis approach. J Integr Plant Biol. 2013;55:1080–91.
31. Badejo AA, Wada K, Gao Y, Maruta T, Sawa Y, Shigeoka S, et al. Translocation and the alternative D-galacturonate pathway contribute to increasing the ascorbate level in ripening tomato fruits together with the D-mannose/L-galactose pathway. J Exp Bot. 2012;63:229–39.

32. Li M, Chen X, Wang P, Ma F. Ascorbic acid accumulation and expression of genes involved in its biosynthesis and recycling in developing apple fruit. J Amer Soc Hort Sci. 2011; 136:231–238.

33. Dowdle J, Ishikawa T, Gatzek S, Rolinski S, Smirnoff N. Two genes in Arabidopsis thaliana encoding GDP-L-galactose phosphorylase are required for ascorbate biosynthesis and seedling viability. Plant J. 2007;52:673–89.

34. Mellidou I, Keulemans J, Kanellis AK, Davey MW. Regulation of fruit ascorbic acid concentrations during ripening in high and low vitamin C tomato cultivars. BMC Plant Biol. 2012;12:1.

35. Mellidou I, Chagné D, Laing WA, Keulemans J, Davey MW. Allelic variation in paralogs of GDP-L-galactose phosphorylase is a major determinant of vitamin C concentrations in apple fruit. Plant Physiol. 2012;160:1613–29.

36. Laing WA, Martínez-Sánchez M, Wright MA, Bulley SM, Brewster D, Dare AP, et al. An upstream open reading frame is essential for feedback regulation of ascorbate biosynthesis in arabidopsis. Plant Cell. 2015;27:772–86.

37. Bulley S, Laing W. The regulation of ascorbate biosynthesis. Curr Opin Plant Biol. 2016;33:15–22.

38. Li H, Liu ZW, Wu ZJ, Wang YX, Teng RM, Zhuang J. Differentially expressed protein and gene analysis revealed the effects of temperature on changes in ascorbic acid metabolism in harvested tea leaves. Hortic Res. 2018;5:65.

39. Liang D, Zhu T, Ni Z, Lin L, Tang Y, Wang Z, et al. Ascorbic acid metabolism during sweet cherry (Prunus avium) fruit development. PLoS One. 2017;12.

40. Gordon SP, Tseng E, Salamov A, Zhang J, Meng X, Zhao Z, et al. Widespread polycistronic transcripts in fungi revealed by single-molecule mRNA sequencing. PLoS One. 2015;10:1–15.

41. Yangyang D, Jiangqi L, Songfeng W, Yunping Z, Chen Y, Fuchu H. Integrated Nr database in protein annotation system and its localization. Comput Eng. 2006;32:71–4.

42. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: Tool for the unification of biology. Nat Genet. 2000;25:25–9.

43. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. Nucleic Acids Res. 2004;32:s277-280.

44. Tatusov RL, Galperin MY, Natale DA, Koonin E V. The COG database: A tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 2000;28:33–6.

45. Koonin E V., Fedorova ND, Jackson JD, Jacobs AR, Krylov DM, Makarova KS, et al. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. Genome Biol. 2004;5:R7.

46. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: The protein families database. Nucleic Acids Res. 2014;42:D222-230.

47. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, et al. UniProt: The universal protein knowledgebase. Nucleic Acids Res. 2004;32:D115-119.

48. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, et al. CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 2007;35(suppl.2):W345–349.
49. Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. Nucleic Acids Res. 2013;41:e166.

50. Zheng Y, Jiao C, Sun H, Rosli HG, Pombo MA, Zhang P, et al. iTAK: A program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. Mol Plant. 2016;9:1667–70.

**Figures**

**Figure 1**

Phenotypic characterization (a) and ascorbic acid accumulation (b) of Actinidia latifolia fruit development and ripening. The X axis represented different stages of fruit development and ripening, which is expressed as days after flowering (DAF). The scale bar in (a) denotes 1 cm. Vertical error bars in (b) represent standard error (SE) of three biological replicates. Statistical difference was assessed by one-way analysis of variance (ANOVA) and Tukey’s HSD test with 95% confidence.
Figure 2

Characteristics of Actinidia latifolia fruit full-length transcriptome. (a) Consensus isoforms read length distribution of 1-6K size bin. (b) Homologous species distribution of A. latifolia transcripts annotated in the non-redundant database. (c) The alignment results against gene ontology database. (d) Candidate IncRNAs predicted by CNCl, CPC, Pfam, and CPAT database.
Figure 3

Stack bar charts showing differentially expressed genes in six stages of Actinidia latifolia fruit development and ripening. C1, DAF30\_vs\_60, C2, DAF30\_vs\_90, C3, DAF30\_vs\_120, C4, DAF30\_vs\_150, C5, DAF30\_vs\_170. DAF, days after flowering. The former is used as a baseline, and the latter is considered as the treatment group.
Figure 4

Weighted gene co-expression network analysis of differentially expressed genes identified over the time course of Actinidia latifolia fruit development and ripening. (a) Hierarchical cluster tree showing eight modules of co-expressed genes. The tree was built based on dynamic hybrid tree cut algorithm. (b) Relationships among modules. Hierarchical clustering of module eigengenes that summarize the eight modules detected in the clustering analysis is located on the top and below was heatmap plot of the adjacencies in the eigengene network. (c) Correlation heatmap among the identified eight modules and the ascorbic acid accumulation. Ascorbic acid trait is presented in column and its association with module eigengene (rows) is determined by a Pearson’s correlation coefficient and a p-value within parentheses.
Figure 5

Histograms of KEGG enrichment classification of blue (a) and brown (b) modules. Abscissas represents the number of annotated genes in corresponding modules, and ordinate represents KEGG metabolic pathways. Different colors are used to distinguish biological processes.
Figure 6

Expression analysis of candidate differentially expressed genes associated with ascorbic acid biosynthesis and metabolism in Actinidia latifolia. The blue and orange line graphs depicted the qRT-PCR validation of gene expression and enzyme activity levels, respectively. The green rectangular box represented the relative expression level of transcripts identified in modules, and the numbers in green rectangular box represented the enzyme commission number.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS15.xlsx
- Fig.S1S4.pdf