Gene 561 (2015) 54–62

Contents lists available at ScienceDirect

Gene

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

# De novo sequencing analysis of the *Rosa roxburghii* fruit transcriptome reveals putative ascorbate biosynthetic genes and EST-SSR markers

Xiuqin Yan [a], Xue Zhang [a], Min Lu [a], Yong He [b], Huaming An [a,*]

[a] Guizhou Engineering Research Center for Fruit Crops, Agricultural College, Guizhou University, Guiyang 550025, PR China
[b] Guizhou Tongjitang Pharmaceutical Co. Ltd, Guiyang 550025, PR China

ARTICLE INFO

*Article history:*
Received 25 November 2014
Received in revised form 15 January 2015
Accepted 3 February 2015
Available online 19 February 2015

*Keywords:*
Transcriptome
*Rosa roxburghii* Tratt
Ascorbic acid
Simple sequence repeat

ABSTRACT

*Rosa roxburghii* Tratt. is a well-known ornamental rose species native to China. In addition, the fruits of this species are valued for their nutritional and medicinal characteristics, especially their high ascorbic acid (AsA) levels. Nevertheless, AsA biosynthesis in *R. roxburghii* fruit has not been explored in detail because of a lack of genomic resources for this species. High-throughput transcriptomic sequencing generating large volumes of transcript sequence data can aid in gene discovery and molecular marker development. In this study, we generated more than 53 million clean reads using Illumina paired-end sequencing technology. De novo assembly yielded 106,590 unigenes, with an average length of 343 bp. On the basis of sequence similarity to known proteins, 9301 and 2393 unigenes were classified into Gene Ontology and Clusters of Orthologous Group categories, respectively. There were 7480 unigenes assigned to 124 pathways in the Kyoto Encyclopedia of Gene and Genome pathway database. BLASTx searches identified 498 unique putative transcripts encoding various transcription factors, some known to regulate fruit development. qRT-PCR validated the expressions of most of the genes encoding the main enzymes involved in ascorbate biosynthesis. In addition, 9131 potential simple sequence repeat (SSR) loci were identified among the unigenes. One hundred and two primer pairs were synthesized and 71 pairs produced an amplification product during initial screening. Among the amplified products, 30 were polymorphic in the 16 *R. roxburghii* germplasms tested. Our study was the first to produce a large volume of transcriptome data from *R. roxburghii*. The resulting sequence collection is a valuable resource for gene discovery and marker-assisted selective breeding in this rose species.



## 1. Introduction

*Rosa roxburghii* Tratt. (Rosaceae), a perennial rosebush native to China, is widely distributed in the southwestern provinces of China. The fruits of this species are known for their nutritional and medicinal components, such as ascorbic acid (AsA), superoxide dismutase, flavonoids, polysaccharides, amino acids, organic acids, and mineral elements (He et al., 1984; Fan et al., 1997, 2004; An et al., 2011). The fruits are therefore believed to have valuable senescence-retarding and cancer-preventing effects (Wen et al., 2007). Compared with fruits such as kiwifruit, strawberry and orange, *R. roxburghii* fruit has very high AsA content (1100–3000 mg per 100 g of fresh weight) (Liu et al., 2013). AsA is of vital importance to humans because of its roles in collagen synthesis and protection against oxidative stress (Padayatty et al., 2003). AsA is also crucial to the function of plant cells, where it is involved in anti-oxidative reactions, regulation of cell division and expansion, and processing of defense responses (Noctor and Foyer, 1998; Davey et al., 2000; Conklin, 2001; Conklin and Barth, 2004). Although several AsA biosynthetic pathways have been proposed in higher plants, and some genes associated with AsA biosynthesis have been identified (Wheeler et al., 1998), details of the molecular mechanisms triggering AsA biosynthesis remain unknown. In addition, the biosynthetic mechanisms operating in the AsA-overproducing fruit of *R. roxburghii* may have distinctive features.

Genomic information is currently unavailable for *R. roxburghii*. As of December 2013, only 167 partial expressed sequence tag (EST) sequences and several complete mRNA sequences have been deposited in the National Center for Biotechnology Information (NCBI) database. Most of these ESTs were submitted in association with studies on vitamin C biosynthesis and resistance to rose powdery mildew. These data are insufficient to determine transcriptome complexity and the molecular mechanisms of specific traits. In addition, simple sequence repeat (SSR) markers have not yet been developed for *R. roxburghii*. These markers are needed for an in-depth understanding of the natural

*Abbreviations:* AsA, ascorbic acid; GO, Gene Ontology; COGs, clusters of orthologous groups; KEGG, Kyoto Encyclopedia of Genes and Genomes; SSR, simple sequence repeat; EST, expressed sequence tag; NCBI, National Center for Biotechnology Information; DAA, days after anthesis; HPLC, high performance liquid chromatography; qRT-PCR, quantitative real-time reverse transcription PCR; CDS, coding sequence; aa, amino acid; TFs, transcription factors; AP2, APETELA2; ZIP, zinc finger; ERF, ethylene response factor; CSN5B, COP9 signalosome subunit 5B; RAPD, random amplified polymorphic DNA; AFLP, amplified fragment length polymorphism.

* Corresponding author.
*E-mail address:* anhuaming@hotmail.com (H. An).

http://dx.doi.org/10.1016/j.gene.2015.02.054
0378-1119/

diversity of *R. roxburghii* and to develop strategies for its sustainable use. The generation of extensive EST collections will aid in the development of molecular markers for further genetic research on *R. roxburghii* and closely related species, and will help determine the molecular mechanisms related to AsA biosynthesis.

In recent years, next-generation high-throughput DNA sequencing techniques have dramatically improved the efficiency and speed of gene discovery (Ansorge, 2009). For example, Illumina sequencing technology offers millions of sequence reads from a single instrument run. ESTs derived from Illumina sequencing can be used to develop SSR markers, which are commonly used to construct linkage maps of nuclear genomes (Gai et al., 2012). In this study, we generated a normalized cDNA library prepared from *R. roxburghii* fruit and established a substantial EST dataset using high-throughput Illumina RNA sequencing. Using these data, we analyzed the *R. roxburghii* fruit transcriptome and identified candidate genes involved in AsA biosynthesis. We also designed a set of SSRs to help genetic diversity analysis and marker-assisted breeding of *R. roxburghii* and closely related species.

## 2. Materials and methods

### 2.1. Plant materials

Plants of *R. roxburghii* 'Guinong 5' (Fan et al., 2011) were grown in the fruit germplasm repository of Guizhou University, Guizhou, China. Fruits were collected at three different developmental stages: 20 days after anthesis (DAA), 60 DAA, and 100 DAA. The fruits were immediately frozen in liquid nitrogen and stored at −70 °C until use.

### 2.2. cDNA preparation and sequencing

Total RNA was isolated from the harvested fruit using the Trizol reagent (Invitrogen), according to the manufacturer's instructions. Equal volumes of RNA from 20-, 60-, and 100-DAA fruits were pooled. Enrichment of mRNA, fragment interruption, addition of adapters, size selection, PCR amplification, and RNA sequencing were performed at the Beijing Genomics Institute, Shenzhen, China. Total mRNA was isolated using oligo(dT) cellulose and broken into short fragments. Using these short fragments as templates, random hexamer-primers were used to synthesize first-strand cDNA. Second-strand cDNA was synthesized using buffer, dNTPs, RNaseH, and DNA polymerase I. Sequencing adapters were ligated to the short fragments after purification using a QIAquick gel extraction kit (Qiagen). The adapters were used to distinguish different sequencing samples. Fragments of 200 ± 25 bp were then separated by agarose gel electrophoresis and used as sequencing templates for PCR amplification. Finally, the cDNA library was sequenced using an Illumina Hiseq 2000 system.

### 2.3. De novo assembly and functional annotation

The raw reads generated by Illumina sequencing were filtered to remove adapter sequences, reads with more than 5% unknown nucleotides, and low-quality sequences containing more than 20% of nucleotides with a *Q*-value ≤10. The clean reads were then assembled de novo using the Trinity assembly program (Grabherr et al., 2011). The Trinity first combined reads with a certain length of overlap to form longer fragments, or contigs. The reads were mapped back to the contigs. Using the paired-end reads, the program was able to detect contigs from the same transcript as well as the distances between these contigs. Finally, the Trinity connected the contigs to produce sequences not extendable on either end, i.e., unigene sequences. To determine the correct sequence direction of the unigenes, BLASTx alignment (*E*-value < $10^{-5}$) was performed between unigenes and Nr, Swiss-Prot, Kyoto Encyclopedia of Genes and Genomes (KEGG), and Clusters of Orthologous Groups (COG) protein databases. In cases where results from the different databases contradicted one another, a priority order of Nr > Swiss-Prot > KEGG > COG was used to decide the unigene sequence direction. When a unigene did not align with any of the above databases, ESTScan software (Iseli et al., 1999) was used to predict its coding region and sequence orientation. The Blast2GO program (Conesa et al., 2005) was used to obtain Gene Ontology (GO) annotations of the unigenes, based on Nr annotations. After GO annotations were acquired for all unigenes, WEGO software (Ye et al., 2006) was used to perform GO functional classification of all unigenes and to visualize the macro-level distribution of gene functions of the species.

### 2.4. Development and detection of EST-SSR markers

Potential SSR markers were identified among the 106,590 unigenes using MicroSAtellite software. The program parameters were adjusted to identify perfect di-, tri-, tetra-, penta-, and hexa-nucleotide motifs with a minimum of 6, 5, 4, 4, and 4 repeats, respectively. Two or more potential EST-SSR loci within a unigene that were separated by no more than 100 bp were treated as a compound EST-SSR. Primer pairs were designed using BatchPrimer3 (Kortt et al., 1991). The major parameters for primer pair design were as follows: primer length of 18–28 bases (average of 23 bases), PCR product size of 80–160 bp, and annealing temperature of 55–65 °C (average 60 °C), with a maximum discrepancy of 2 °C among primers. In total, 102 primer pairs (Additional file 1) were synthesized and used for SSR amplification in 16 wild *R. roxburghii* germplasms in an initial screening test. PCR amplifications were performed at 94 °C for 3 min; followed by 35 cycles of 94 °C for 40 s, 55–65 °C for 40 s and 72 °C for 1 min; with a final extension at 72 °C for 10 min. PCR products were separated by electrophoresis on 8% non-denaturing polyacrylamide gels in 1 × TBE (Tris-Borate-EDTA) buffer.

### 2.5. Quantitative real-time reverse transcription PCR (qRT-PCR) analysis

Seven cDNAs encoding GME, GGP, GPP, GDH, GLDH, GUR, and MIOX proteins, all of which have potential roles in ascorbate synthesis, were chosen for qRT-PCR validation. Target gene primers were designed (Additional file 2) according to acquired sequences using Primer Express software (Applied Biosystems, USA). Total RNAs were extracted from *R. roxburghii* 'Guinong 5' fruit at 20, 60, and 100 DAA using the Trizol reagent (Invitrogen), followed by purification with an RNA purification kit (Takara). qRT-PCR was performed on an ABI ViiA 7 DX system (Applied Biosystems) using SYBR Premix Ex Taq II (TaKaRa) with the ubiquitin gene as an endogenous control. Data analysis was performed using the $2^{-\Delta\Delta CT}$ method, as described by Livak and Schmittgen (2001).

### 2.6. AsA content assays

AsA was evaluated on an HPLC system (Shimadzu, Japan) using a modification of the method described by Davey et al. (2000). Frozen samples (0.5 g) were dissolved in 10 ml of 6% metaphosphoric acid. A 20-μl aliquot of each sample was injected onto a Wondasil C18 column (4.6 mm × 150 mm; 5 μm particle size). The column was eluted at 1 ml $min^{-1}$ at 30 °C. The mobile phase was composed of 0.2% metaphosphoric acid in water. AsA quantification was performed at a UV absorption wavelength of 254 nm. A calibration curve was constructed using an AsA standard solution (Sigma-Aldrich) dissolved in 0.2% metaphosphoric acid, with concentrations ranging between 4 and 400 μg $ml^{-1}$. AsA concentrations were quantified by the standard curve and expressed in mg per 100 g fresh weight. All samples were extracted and analyzed in triplicate.