Gene 540 (2014) 96–103

Contents lists available at ScienceDirect

Gene

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

# Transcriptome and gene expression analysis during flower blooming in *Rosa chinensis* 'Pallida'

Huijun Yan [a,b], Hao Zhang [a,b,1], Min Chen [a,b], Hongying Jian [a,b], Sylvie Baudino [c], Jean-Claude Caissard [c], Mohammed Bendahmane [d], Shubin Li [a,b], Ting Zhang [a,b], Ningning Zhou [a,b], Xianqin Qiu [a,b], Qigang Wang [a,b], Kaixue Tang [a,b,*,1]

[a] Flower Research Institute of Yunnan Academy of Agricultural Sciences, Kunming, Yunnan 650205, PR China
[b] Yunnan Flower Breeding Key Lab., Kunming Yunnan 650205, PR China
[c] Laboratoire BVpam, EA3061, Université de Saint-Etienne, Université de Lyon, 23 rue du Dr Michelon, Saint-Etienne F-42023, France
[d] Laboratoire Reproduction et Développement des Plantes, UMR INRA-CNRS, Lyon1-ENSL, Ecole Normale Supérieure, 46 allée d'Italie, 69364 Lyon Cedex 07, France

ARTICLE INFO

*Article history:*
Accepted 8 February 2014
Available online 12 February 2014

*Keywords:*
*Rosa chinensis* 'Pallida'
Transcriptome
Digital gene expression
Scent-related genes

ABSTRACT

*Rosa chinensis* 'Pallida' (*Rosa* L.) is one of the most important ancient rose cultivars originating from China. It contributed the 'tea scent' trait to modern roses. However, little information is available on the gene regulatory networks involved in scent biosynthesis and metabolism in *Rosa*. In this study, the transcriptome of *R. chinensis* 'Pallida' petals at different developmental stages, from flower buds to senescent flowers, was investigated using Illumina sequencing technology. De novo assembly generated 89,614 clusters with an average length of 428 bp. Based on sequence similarity search with known proteins, 62.9% of total clusters were annotated. Out of these annotated transcripts, 25,705 and 37,159 sequences were assigned to gene ontology and clusters of orthologous groups, respectively. The dataset provides information on transcripts putatively associated with known scent metabolic pathways. Digital gene expression (DGE) was obtained using RNA samples from flower bud, open flower and senescent flower stages. Comparative DGE and quantitative real time PCR permitted the identification of five transcripts encoding proteins putatively associated with scent biosynthesis in roses. The study provides a foundation for scent-related gene discovery in roses.



## 1. Introduction

Roses are cultivated worldwide and they are mainly used as cut flowers as garden ornamental and for the perfume industry (Zhang and Zhu, 2006). Ancient Chinese cultivars, including *Rosa chinensis* 'Pallida', *R. chinensis* 'Semperflorens', *Rosa odorata* 'Park's yellow tea-scented China', and *R. odorata* 'Hume Blush tea-scented' were hybridized with European roses to breed modern rose cultivars (Chen, 2001). *R. chinensis* 'Pallida' is closely related to *R. chinensis* 'Old Blush'. These two rose cultivars are part of the 'China roses' composed of both natural and cultivated hybrids that have evolved over more than a thousand years in Chinese gardens, under different ecological and environment conditions.

Among ornamental plants, the rose is of interest as a model species because of its traits such as scent production, recurrent blooming and double flower characters (Bendahmane et al., 2013). Several groups have recently initiated molecular approaches aimed at providing new genetic and transcriptomic tools applied to rose. In the past few years efforts have been made to identify *Rosa* sp. expressed genes (Bendahmane et al., 2013; Channelière et al., 2002; Dubois et al., 2011, 2012; Foucher et al., 2008; Guterman et al., 2002; Kim et al., 2012; Pei et al., 2013). Very recently, the transcriptome in most organs of *R. chinensis* 'Old Blush' was investigated using a combination of 454 and Illumina sequencing technologies (Dubois et al., 2012). This study yielded valuable sequence dataset information representing about 20,000 *Rosa* sp. expressed genes, as well as digital gene expression (DGE) for the identified rose transcripts (Dubois et al., 2012). However, we are far from having information on all *Rosa* sp. expressed gene, estimated to about 30,000 genes. For example, some known genes associated with scent production and emission are missing in this dataset.

In the past few years, there has been an increasing demand from consumers worldwide for scented rose flowers (Bergougnoux et al., 2007). However, modern rose breeding has mainly focused on cold tolerance and disease resistance, flower form and recurrent blooming (Yan et al., 2011). Fragrance seems to have been largely lost during the breeding process (Channelière et al., 2002). Nowadays, only limited transcriptomic and genomic data are available on scent-related gene pathways in *Rosa* sp.

*Abbreviations:* GO, gene ontology; COG, clusters of orthologous groups; KEGG, Kyoto Encyclopedia of Genes and Genomes; NR, non-redundant databases; qRT-PCR, quantitative real-time polymerase chain reaction; OOMT, orcinol Omethyltransferases; EGS, eugenol synthase; LIS, linalool synthase; AAT, alcohol acetyltransferase; TF, transcription factor.

* Corresponding author at: Flower Research Institute of Yunnan Academy of Agricultural Sciences, Kunming, Yunnan 650205, PR China.
*E-mail address:* kxtang@hotmail.com (K. Tang).

[1] Equal contribution to this work.

0378-1119/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2014.02.008

With the objective to identify genes associated with rose scent, RNAseq was used to compare the transcriptome at three flower developmental stages of *R. chinensis* 'Pallida', namely flower bud, open flower and senescing flower stages. The EST datasets and DGE data will serve as a valuable resource for the discovery of candidate genes associated with scent in *Rosa* sp.

## 2. Results and discussion

### 2.1. Illumina sequencing and de novo assembly

mRNAs were purified from three rose flower development stages exhibiting contrasted scent emission; the open flowers at which high peak of scent emission is observed and floral buds and senescing flowers with much less scent emission (Fig. 1). A cDNA library was generated from an equal mixture of RNA isolated from the above three flower development stages and then used for Illumina sequencing. Using 90 bp pair-end sequencing based on Illumina sequencing approach 66,523,228 reads were obtained and then assembled using Trinity Software (Grabherr et al., 2011). The longest assembled sequences containing blocks of unknown bases (Ns) were called contigs (Li et al., 2010). A total of 155,708 contigs ranging in length from 75 to 5236 bp were assembled with an average length of 380 bp (Table 1). The RNA-seq data in this study have been deposited in the Gene Expression Omnibus (GEO) database (GSE54486).

The de novo assembly yielded 89,641 unisequences with a total length of 38,355,533 bp and an average length of 428 bp (Table 1), which showed a similar average length to that previously published for *R. chinensis* 'Old Blush' (average length of 444 bp) (Dubois et al., 2012). However, the assembly in this study extended the range of sequence length ranging from 200 to 7326 bp. The length of sequences ranged from 200 to 500 nucleotides for 64,907 unisequences (72.4%), from 501 to 1000 nucleotides for 19,711 unisequences (22.0%) and over 1000 nucleotides for 5023 unisequences (Fig. S1). Further, the assembled sequences were compared with the 80,714 transcript clusters of *R. chinensis* 'Old Blush' using custom PERL scripts and plots were drawn by R with ggplot2 (Ito and Murphy, 2013). The percentages of sequences with more than 760 bp in length were slightly higher than those of the previous sequence assembly based on 454 sequencing (Dubois et al., 2012). This is likely due to the fact that here we performed 90 bp pair-end Illumina sequencing approach and to the depth of sequencing (Fig. 2). Therefore, the data here complement the previously published work and provide novel information on *Rosa* sp. expressed genes.

### 2.2. Annotation of predicted proteins

Sequence similarity search was conducted against non-redundant database (Nr), UniProtKB/Swiss-Prot (SwissProt), Gene Ontology (GO), Clusters of Orthologous Groups (COG) and Kyoto Encyclopedia of Genes and Genomes (KEGG). We found a total of 55,442 (61.9% of all clusters) unisequences providing significant BLAST hits, while 38,256 unisequences had similarity to proteins in the Swiss-Prot database. Altogether, 56,378 unisequences were successfully annotated in the Nr or Swiss-Prot databases. GO assignments were used to classify the predicted functions of *R. chinensis* 'Pallida' genes. Based on sequence homology, 25,705 unisequences were categorized into 43 functional groups (Fig. 3). In each of the three main categories (biological process, cellular component, and molecular function) of the GO classification, the metabolic process, various cellular activities and catalytic activity terms were dominant, respectively. A high-percentage of genes was assigned to the categories of cellular process, cell components and DNA binding (Fig. 3). Only few genes were assigned to other categories such as antioxidant activity (Fig. 3). Flavones, anthocyanin, coumarin lignans and catechins contribute to the majority of the antioxidant activity, they are usually present at high levels in medicinal plants (Škrovánková et al., 2012). In addition, all clusters were subjected to a search against the Cluster of Orthologous Groups (COG) database for functional prediction and classification. In total, 37,159 of 56,378 sequences showing Nr hits were assigned to COG classifications (Fig. S2). Among the 25 COG categories, the cluster for 'General function prediction' represents the largest group (5219; 14.0%) followed by 'Transcription' (3728; 10.0%) and 'Replication, recombination and repair' (2636; 7.1%), with the following categories extracellular structures (4; 0.01%), and nuclear structures (7; 0.02%) being the smallest groups (Fig. S2).

In order to identify the active biological pathways in *R. chinensis* 'Pallida' flower buds, blooming and senescing flowers, 22,605 annotated transcripts were mapped to 121 KEGG pathways (Table S1). Transcripts representing metabolic pathways (5812 members), biosynthesis of secondary metabolites (2931 members), plant–pathogen interaction (1512 members), starch and sucrose metabolism (622 members) and phenylpropanoid biosynthesis (450 members) were represented in our dataset. These annotations provide a valuable resource for studying specific processes, and pathways in *Rosa* research.

### 2.3. R. chinensis 'Pallida' dataset as a resource for scent-related gene identification

In the annotated *R. chinensis* 'Pallida' transcriptome dataset, we identified multiple transcripts encoding almost all known enzymes mainly

**Table 1**
Summary of data generated for *R. chinensis* 'Pallida' transcriptome.

| | |
|---|---|
| Total number of reads | 66,523,228 |
| Total clean nucleotides (bp) | 4,822,934,030 |
| Number of contigs | 155,708 |
| Average length of all contigs (bp) | 380 |
| Range of contig length (bp) | 75–5236 |
| Number of unisequences | 89,614 |
| Length of all unigenes (bp) | 38,355,533 |
| Range of unigene length (bp) | 200–7326 |
| Average unigene length (bp) | 428 |


**Fig. 1.** Different developmental stages of *R. chinensis* 'Pallida'. A. Floral bud. B. Open flower. C. Senescent flower.