Mol Genet Genomics (2015) 290:365–375
DOI 10.1007/s00438-014-0920-7

ORIGINAL PAPER

# Characterization of the 'Xiangshui' lemon transcriptome by de novo assembly to discover genes associated with self-incompatibility

**Shuwei Zhang · Feng Ding · Xinhua He · Cong Luo · Guixiang Huang · Ying Hu**

Received: 25 February 2014 / Accepted: 11 September 2014 / Published online: 25 September 2014


**Abstract** Seedlessness is a desirable character in lemons and other citrus species. Seedless fruit can be induced in many ways, including through self-incompatibility (SI). SI is widely used as an intraspecific reproductive barrier that prevents self-fertilization in flowering plants. Although there have been many studies on SI, its mechanism remains unclear. The 'Xiangshui' lemon is an important seedless cultivar whose seedlessness has been caused by SI. It is essential to identify genes involved in SI in 'Xiangshui' lemon to clarify its molecular mechanism. In this study, candidate genes associated with SI were identified using high-throughput Illumina RNA sequencing (RNA-seq). A total of 61,224 unigenes were obtained (average, 948 bp; N50 of 1,457 bp), among which 47,260 unigenes were annotated by comparison to six public databases (Nr, Nt, Swiss-Prot, KEGG, COG, and GO). Differentially expressed genes were identified by comparing the transcriptomes of no-, self-, and cross-pollinated stigmas with styles of the 'Xiangshui' lemon. Several differentially expressed genes that might be associated with SI were identified, such as those involved in pollen tube growth, programmed cell death, signal transduction, and transcription. *NADPH oxidase* genes associated with apoptosis were highly upregulated in the self-pollinated transcriptome. The expression pattern of 12 genes was analyzed by quantitative real-time polymerase chain reaction. A putative *S-RNase* gene was identified that had not been previously associated with self-pollen rejection in lemon or citrus. This study provided a transcriptome dataset for further studies of SI and seedless lemon breeding.

**Keywords** Lemon · Self-incompatibility · RNA-seq · Programmed cell death · S-RNase

Communicated by S. Hohmann.

Shuwei Zhang and Feng Ding contributed equally to this paper.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00438-014-0920-7) contains supplementary material, which is available to authorized users.

S. Zhang · X. He (✉) · C. Luo · G. Huang · Y. Hu
Agricultural College of Guangxi University, Nanning 530004, China
e-mail: honest66222@163.com

S. Zhang
e-mail: zsw152@126.com

C. Luo
e-mail: 22003luocong@163.com

G. Huang
e-mail: hguixiang@163.com

Y. Hu
e-mail: huying19863008@163.com

S. Zhang · F. Ding · X. He
Guangxi Crop Genetic Improvement and Biotechnology Key Lab, Nanning 530007, China
e-mail: dingfeng418@163.com

F. Ding
Horticulture College of South China Agricultural University, Guangzhou 510642, China

## Introduction

Seedlessness is an important economic trait of lemon and other citrus species. Self-incompatibility (SI) is one method of producing seedless fruit in citrus, especially coupled with parthenocarpy. Therefore, understanding the molecular mechanisms of SI is important for seedless fruit breeding.

SI is widely observed in higher order plants, preventing self-fertilization and promoting genetic diversity. The

predominant form of SI is a gametophytic SI system (GSI), which is controlled by a single multi-allelic S-locus. In recent years, great progress has been made in understanding the mechanism of GSI, including in Rosaceae, Solanaceae and Papaveraceae. Many fruit tree species are GSI type, such as *Malus*, *Pyrus*, and some *Citrus* species. For GSI, two systems have been reported: GSI based on the S-ribonuclease (S-RNase) in Rosaceae, Scrophulariaceae, and Solanaceae; and GSI based on a $Ca^{2+}$ signaling cascade that results in the programmed cell death (PCD) of self-pollen tubes in Papaveraceae. The S-locus contains at least two genes, one expressed in pistils and the other in pollen. S-RNases secreted by the pistil act as cytotoxins to degrade the RNA of self-pollen tubes while allowing normal cross-pollen growth. Pollen-expressed *S* genes belong to a class of *F-box* genes *(S-locus F-box*, *SLF*). Pollen-expressed *S* genes can participate in an SCF E3 ubiquitin ligase complex, which is important for protein ubiquitination and degradation. S-locus genes of many fruit tree species have been identified and cloned, particularly in Rosaceae (Wu et al. 2013), such as in apple (Sakurai et al. 2000), pear (Zisovich et al. 2004), plum (Zhang et al. 2008), cherry (Gu et al. 2010; Schuster 2012) and apricot (Wang et al. 2012). However, in citrus, the SI-related genes have not yet been identified.

*Citrus* is an important economic crop worldwide, and seedlessness is a desired trait because seedless fruits are popular in the market. Several pummelo and mandarin-like cultivars are self-incompatible. Several studies have been performed to identify putative genes involved in citrus SI. Suppression subtractive hybridization, cDNA-amplification length polymorphisms, and microarray analysis have been performed to detect differentially expressed genes and identify candidate genes involved in SI (Distefano et al. 2009; Chai et al. 2011b; Caruso et al. 2012; Miao et al. 2013a, b). Two S-like RNase genes were isolated from the 'Shatian' and 'Zigui shatian' pummelos. However, it is not clear whether these two *S-like RNase* genes are involved in SI. In clementine (*Citrus clementina* Hort. Ex Tan.), some candidate genes have been isolated, such as receptor-like kinases, stress-induced genes, transcription factors, calcium-related genes and hormone signaling genes. In the self-incompatible 'Wuzishatangju' (*C. reticulata* Blanco), it has been demonstrated that candidate genes involved in the SI reaction act by regulating signaling pathways, pollen development, receptor kinases, the ubiquitin pathway, calcium ion binding, gibberellin stimulus and transcription (Miao et al. 2013a, b). Moreover, $Ca^{2+}$ is considered to be important in SI signal transduction. Furthermore, a novel *F-box* gene and three different *aspartic-acid rich protein* genes that potentially regulate $Ca^{2+}$ homeostasis during self-pollen recognition were identified in *Citrus clementina* (Caruso et al. 2012).

RNA sequencing (RNA-seq) is a powerful method to investigate the molecular basis of many agronomically important traits in citrus and other plants. RNA-seq has been used successfully for transcript annotation in tea (Shi et al. 2011), garlic (Sun et al. 2012), banana (Li et al. 2012), olive (Gil-Amado and Gomez-Jimenez 2012) and grape (Venturini et al. 2013), among other plants.

The 'Xiangshui' lemon [*C. limon* (L.) Burm. F.], which flowers throughout the year, is seedless, with a strong sweet smell. Our previous study demonstrated that it is self-incompatible because of inhibited fertilization mainly in the base of the stigma (Zhang et al. 2012). In this study, RNA-seq was used to analyze the transcriptomes of stigmas with styles of the 'Xiangshui' lemon 24 h after no-, self-, and cross-pollination. This transcriptome dataset could provide useful information to understand the molecular mechanism of SI, and can be used for future genomics studies of this plant.

## Materials and methods

### Plant materials and RNA extraction

In this study, stigmas with styles were collected from the 'Xiangshui' lemon (self-incompatible) during full bloom, 24 h after no- (Sample_XS), self- (Sample_XX) and cross- (Sample_XB) pollination, respectively. All flowers were bagged before flowering. The pollen of the 'Baihua' lemon [*C. limon* (L.) Burm. F.] (self-compatible) was used for cross-pollination. The trees were grown in Guangxi University, Nanning, China. The samples were immediately frozen in liquid nitrogen and stored at −80 °C until use.

The TRIzol reagent (Takara) was used to isolate total RNA according to the manufacturer's instructions. RNA samples were treated with RNase-free DNase I (Fermentas Life Sciences) to remove contaminating DNA. A NanoDrop ND1000 instrument (Thermo Scientific) was used to assess RNA quantity and quality. Three biological replicates per treatment were diluted at equal concentrations and amounts and then pooled to obtain a final quantity of 25 μg of RNA per condition.

### Preparation of cDNA library and transcriptome sequencing

mRNA was isolated using magnetic beads with Oligo (dT) and then mixed with the fragmentation buffer. mRNA was fragmented into short fragments that were then used as templates for cDNA synthesis. We used random hexamer primers and reverse transcriptase (Invitrogen) to synthesize first-strand cDNA, the second-strand cDNA was synthesized using RNase H (Invitrogen) and DNA polymerase (Invitrogen). Short fragments were purified and then

resolved with EB buffer to repair the ends and add a single A (adenine) nucleotide. The short fragments were then connected to adapters. The products were purified and amplified via polymerase chain reaction (PCR) to create the final cDNA libraries. An Agilent 2100 Bioanalyzer (Agilent, USA) and ABI StepOnePlus Real-Time PCR system (ABI, USA) were used to assess the quantity and quality of the final cDNA libraries. An Illumina HiSeq™ 2000 performed the sequencing reactions.

### Bioinformatics analysis

Raw reads produced from sequencing machines included dirty reads that contain adapters and unknown or low-quality bases. Therefore, before assembly, we filtered out these dirty reads. Clean reads were mapped to the citrus genome and gene sequences (http://citrus.hzau.edu.cn/orange/index.php) using SOAP2, permitting no more than five base mismatches in the alignment. The Trinity program was used to perform de novo assembly. Trinity combines three independent software modules, Inchworm, Chrysalis, and Butterfly (Grabherr et al. 2011), which are applied sequentially to process large volumes of RNA-seq reads. The clean reads were first assembled into longer contigs. Then, the pair-end reads were mapped back to the contigs, taking the distance of paired-end reads as a frame. The contigs were connected to access the sequences that could not be extended on both ends, and the sequence of the unigene was then produced. The unigenes were further spliced and assembled to acquire non-redundant unigenes that were as long as possible. The unigenes were then divided into two classes: clusters with the prefix CL and singletons with the prefix unigene. In the final step, BLASTx (e-value <0.00001) was performed to annotate the unigenes based on protein databases, including Nr, Swiss-Port, KEGG, and COG. The best results were used to determine the direction of the unigenes. If results based on different databases conflicted with one another, a priority order of Nr, Swiss-Port, KEGG, and COG was followed. The blast2go (Conesa et al. 2005) program was employed to obtain GO annotations for the unigenes. After obtaining GO annotations for every unigene, we used WEGO (Ye et al. 2006) software to perform GO functional classification for all unigenes and to gain an understanding of the distribution of gene functions of the species from the macro-level. The BLASTx algorithm (E-value threshold: $1e^{-5}$) was used to search against the Cluster of Orthologous Groups (COG) database and the Kyoto Encyclopedia of Genes and Genomes pathway (KEGG) database to produce the COG and KEGG pathway annotations.

Gene expression levels were calculated using the reads per kilobase per million reads method described by Mortazavi et al. (2008). Differential expression of genes across the three samples was identified using the method described by Audic and Claverie (1997). In this study, genes were deemed differently expressed at a false discovery rate (FDR $\leq$0.001) and ratio greater than 2. Differently expressed genes were then used for GO functional analysis and KEGG pathway analysis.

### Quantitative real-time PCR (qPCR) analysis

Stigmas were sampled from 0 (non-pollinated flowers) to 6 days after pollination. Total RNA was extracted using the method mentioned above and treated with RNase-free DNase I (Fermentas, Switzerland) to remove contaminating DNA. M-MLV reverse transcriptase (Fermentas) was used to synthesize single-strand cDNA, according to the manufacturer's instructions. The SYBR Green Real-time PCR Master Mix (Takara) with the ABI 7500 Real Time System (PE Applied Biosystems, Foster City, CA, USA) was used to perform qPCR. Reactions were performed as the following: 95 °C for 30 s, followed by 40 cycles of 95 °C for 5 s and 60 °C for 30 s. Three biological replicates were analyzed. Actin was used to normalize mRNA levels. Gene-specific primers were designed according to the reference unigene sequences (Table S1). Quantitative variation was analyzed using the formula $2^{-\Delta\Delta CT}$ (Livak and Schmittgen 2001).

## Results

### Analysis of RNA-Seq datasets

After removing adaptor, ambiguous and low-quality reads, we generated approximately 54.2, 54.8, and 53.9 million high-quality clean reads from the XS, XX, and XB samples, respectively (NCBI accession number: SRA174344). Of these high-quality reads, 70.16, 69.32, and 69.04 % were aligned to the citrus mRNA reference sequences, respectively (Table 1). De novo assembly generated 87,421, 98,907, and 95,225 contigs with mean lengths of 393, 383, and 363 bp and N50 s of 820, 799, and 675 bp, respectively. Finally, 57,573, 63,387, and 62,552 unigenes were obtained with mean lengths of 761, 749, and 686 bp and N50s of 1,256, 1,242, and 1,112 bp, for samples XS, XX, and XB, respectively. A total of 61,224 unigenes were obtained with a mean length of 948 bp and N50 of 1,457 bp (Table 2). Of these, 36,203 unigenes were greater than 500 bp, and 21,359 unigenes (34.88 % of the total 61,224 unigenes) were greater than 1,000 bp (Figure S1).

### Functional annotation and classification

All of the unigenes were compared to sequences in public databases, including the NCBI non-redundant protein (Nr)

**Table 1** Statistics of the reads obtained and their mapping results to citrus genes

| Sample | Total clean reads | Q20 percentage | *N* percentage | GC percentage | Total mapped reads to reference genes (%) |
|---|---|---|---|---|---|
| Sample XS | 54,215,858 | 96.87 | 0.00 | 46.37 | 70.16 |
| Sample XX | 54,895,612 | 96.95 | 0.00 | 45.68 | 69.32 |
| Sample XB | 53,989,360 | 96.84 | 0.00 | 46.88 | 69.04 |

Q20 percentage: a sequencing error probability of 0.01

**Table 2** De novo assembly statistics

| Sample | Contigs | | | Unigenes | | | | |
|---|---|---|---|---|---|---|---|---|
| | Total number | Mean length (bp) | N50 (bp) | Total number | Mean length (bp) | N50 (bp) | Clusters | Singletons |
| XS | 87,421 | 393 | 820 | 57,573 | 761 | 1,256 | 19,179 | 38,395 |
| XX | 98,907 | 383 | 799 | 63,383 | 749 | 1,242 | 21,070 | 42,317 |
| XB | 95,225 | 363 | 675 | 62,552 | 686 | 1,112 | 19,574 | 42,978 |
| All | | | | 61,224 | 948 | 1,457 | 27,128 | 34,096 |

database, the NCBI COG database, the Swiss-Port protein database, and the KEGG database. A total of 45,265 unigene-matched sequences in the Nr databases. Statistical analysis of the top hits in the Nr database demonstrated that 60 % of the mapped sequences had significant homology (<1.0E−45) (Figure S2 A), and 29.5 % of the sequences had alignment identities greater than 80 % (Figure S2 B).

More specifically, the assembled unigenes were compared against COGs for in-depth analysis of phylogenetically widespread domain families. COGs consist of protein sequences encoded in 21 complete genomes, including bacteria, algae, and eukaryotes, and were built on classifications according to phylogenetic relationships. Each COG consists of individual proteins or groups of paralogs from at least three lineages and thus corresponds to an ancient conserved domain. A total of 17,224 unigenes were assigned COG classifications (Fig. 1).

Among the 25 COG categories, the cluster of general function prediction represented the highest number (5,631, 17.23 %), followed by transcription (2,955, 9.05 %), replication, recombination and repair (2,589, 7.93 %), signal transduction mechanisms (2,265, 6.93 %), post-translational modification, protein turnover and chaperones (2,240, 6.86 %), and translation, ribosomal structure, and biogenesis (1,957, 5.99 %). The categories of nuclear structure (2, 0.006 %) and extracellular structures (6, 0.018 %) had the fewest corresponding genes.

Gene ontology (GO) is an international standardized gene functional classification system that offers a dynamic-updated controlled vocabulary and strictly defined concept to comprehensively describe properties of genes and their products in any organism. GO has three ontologies: molecular function, cellular component and biological process. A total of 37,760 unigenes were categorized into 57 subcategories (Fig. 2).

To further investigate biological behavior, the unigenes were assigned to the biochemical pathways described in the KEGG database. A total of 26,394 unigenes were assigned to the 128 KEGG pathways, including metabolic pathways, biosynthesis of secondary metabolites, plant hormone signal transduction, ubiquitin-mediated proteolysis, and RNA degradation.

## Identification of differentially expressed genes

To identify genes associated with SI in the 'Xiangshui' lemon, the significance of gene expression differences was determined using the threshold of FDR $\leq$0.001 and $|\log_2\text{Ratio}| \geq 1$. Totals of 4,785, 7,677, and 5,168 differentially expressed genes were obtained between samples XS and XX, samples XS and XB, samples XB and XX, respectively. For these up- and down-regulated genes, GO and KEGG analyses were performed. Specifically, the expression levels of 3,342 genes were upregulated and 1,443 genes were downregulated in sample XX compared with sample XS. Additionally, 2,849 of these differentially expressed genes were associated with 51 subcategories, and 2,212 were mapped to 116 pathways. The expression levels of 3,372 genes were upregulated and 4,305 genes were downregulated in sample XB compared with sample XS; 4,850 of these differentially expressed genes were associated with 52 sub-categories, and 3,516 were mapped to 122 pathways. The expression levels of 4,098 genes were upregulated and 1,070 genes were downregulated in sample XX compared with sample XB; 3,370 of these differentially expressed genes were associated with 51 sub-categories, and 2,599 were mapped to

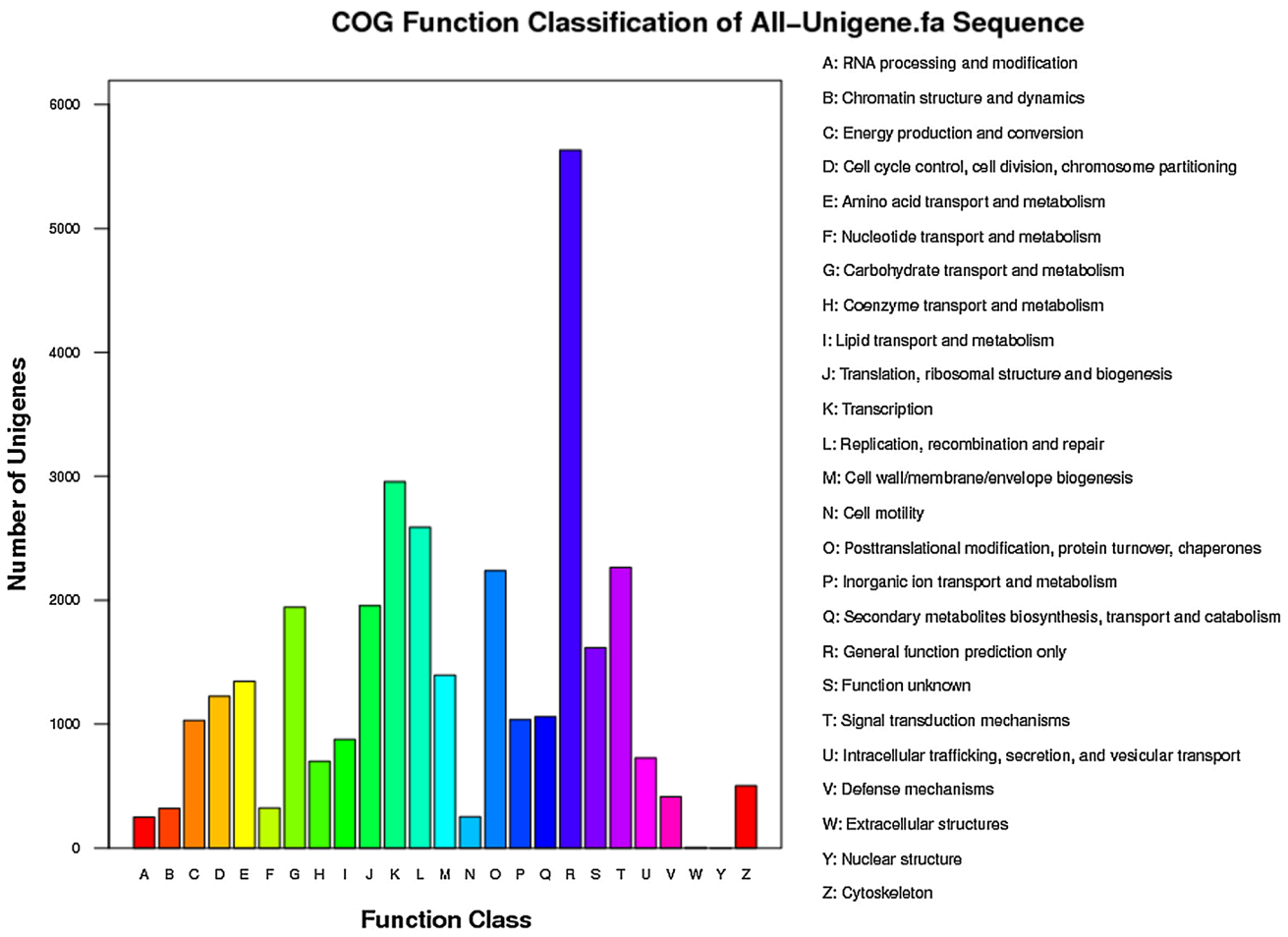



**Fig. 1** COG functional classification of all unigenes. 17,224 unigenes were assigned to 25 categories in the COG classification

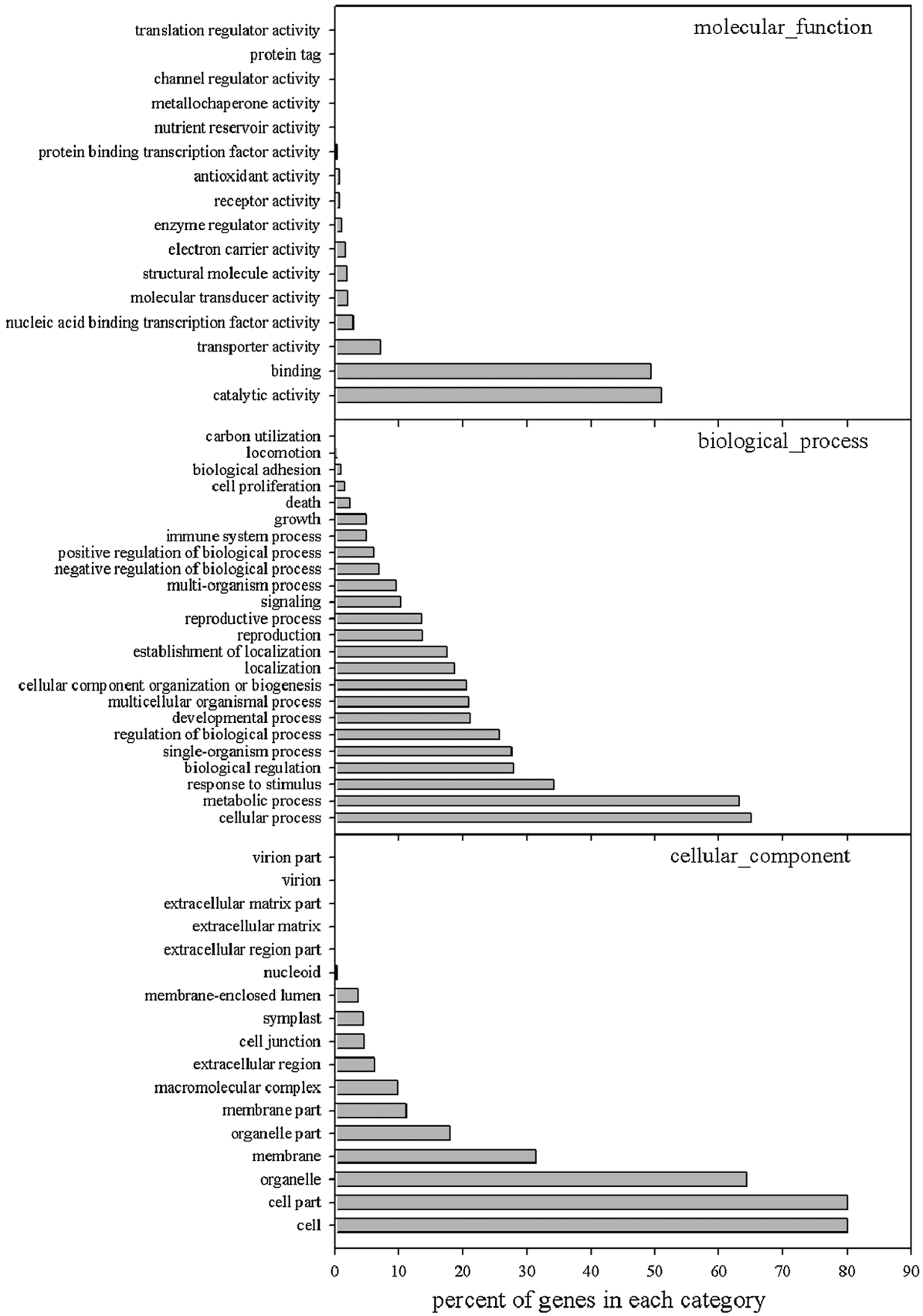
123 pathways. In the above three comparisons, specific pathways were observed in all three comparisons, such as ubiquitin-mediated proteolysis, RNA degradation, and plant hormone signal transduction. Additionally, several stress resistance, transcription-associated and unknown genes were identified. Notably, these defense-related genes might function not only in defense but also in response to pollination, which is consistent with previous studies in *Arabidopsis* (Tung et al. 2005) and rice (Lan et al. 2005). Further study is necessary on those genes that may be associated with SI.

## Candidate genes involved in SI and expression analysis of 12 candidate genes by qPCR

Pollen tube growth and elongation play an important role in sexual reproduction. The genes or factors that associate with pollen tube growth may be involved in abnormal growth of self-pollen tubes along with S-locus genes. According to the results of GO and KEGG analyses of differentially expressed genes, several genes may be associated with SI (Table S2).

A putative *S-RNase* gene (unigene 14715) was cloned that is homologous with ribonuclease $T_2$ family genes. Phylogenetic analysis was conducted by aligning entire predicted protein sequences, and the neighbor-joining phylogenetic tree demonstrated that unigene 14175 most closely related to the *S-RNase* from *Nicotiana sylvestris* (Genebank: CAA05306) and other self-incompatibility-associated ribonucleases from *Prunus* (Fig. 3).

In total, 12 candidate genes were chosen for expression analysis. qPCR analysis was performed on the 12 candidate genes using gene-specific primer pairs. The transcription levels of these genes were determined in stigmas at different times after self- and cross-pollination (Fig. 4). The expression levels of unigene 29710, unigene 29755, unigene 29829, unigene 28158, unigene 222, unigene 29240, unigene 28930 and CL8452 contig1 were upregulated after self-pollination. Their highest expression levels were at 3 and 4 days after self-pollination. However, their


molecular_function
translation regulator activity
protein tag
channel regulator activity
metallochaperone activity
nutrient reservoir activity
protein binding transcription factor activity
antioxidant activity
receptor activity
enzyme regulator activity
electron carrier activity
structural molecule activity
molecular transducer activity
nucleic acid binding transcription factor activity
transporter activity
binding
catalytic activity
biological_process
carbon utilization
locomotion
biological adhesion
cell proliferation
death
growth
immune system process
positive regulation of biological process
negative regulation of biological process
multi-organism process
signaling
reproductive process
reproduction
establishment of localization
localization
cellular component organization or biogenesis
multicellular organismal process
developmental process
regulation of biological process
single-organism process
biological regulation
response to stimulus
metabolic process
cellular process
cellular_component
virion part
virion
extracellular matrix part
extracellular matrix
extracellular region part
nucleoid
membrane-enclosed lumen
symplast
cell junction
extracellular region
macromolecular complex
membrane part
organelle part
membrane
organelle
cell part
cell
0
10
20
30
40
50
60
70
80
90
percent of genes in each category

◀ **Fig. 2** Histogram representation of GO classifications. All unigenes were annotated in three categories, as depicted on the *y*-axis. The *x*-axis indicates the percentage of genes in each category

expression levels were weak in cross-pollinated stigmas, with the highest expression at 4 h. The highest expression levels of unigene 14715 and CL2009 contig1 were at 4 h and 3 days after self-pollination, respectively; however, their expression levels were weak in cross-pollinated stigmas. Unigene 14175 was expressed specifically in pistils and most highly expressed in stigmas compared with styles and ovaries. CL2009 contig1 was expressed specifically in anthers. These results suggested that these genes could play important roles in self-incompatibility. The selected genes showed the same expression trend in qPCR analysis as they did in the RNA-seq analysis.

Significant differences in unigene 14175 in expression were analyzed between different tissues of 'Xiangshui' lemons (Fig. 5). The highest expression was detected in stigmas, where its expression was approximately 2-, 3-, and 10-fold higher than in styles, ovaries, and anthers, respectively. There was almost no expression in leaves, stems, petals, and filaments.

## Discussion

RNA-seq is a powerful tool to investigate the molecular basis of agronomic traits and provides a new strategy to identify candidate genes involved in SI. In this study, RNA-seq was used to discover genes associated with SI in 'Xiangshui' lemon. From among 61,224 assembled unigenes, 47,260 were annotated. Several genes, such as putative *S-RNase*, *F-box*, *NADPH oxidase*, and *DELLA* were identified as SI related (Table S2). There have been no previous reports of their association with SI in lemons. The data are a valuable addition to the publicly available, lemon genomic information.

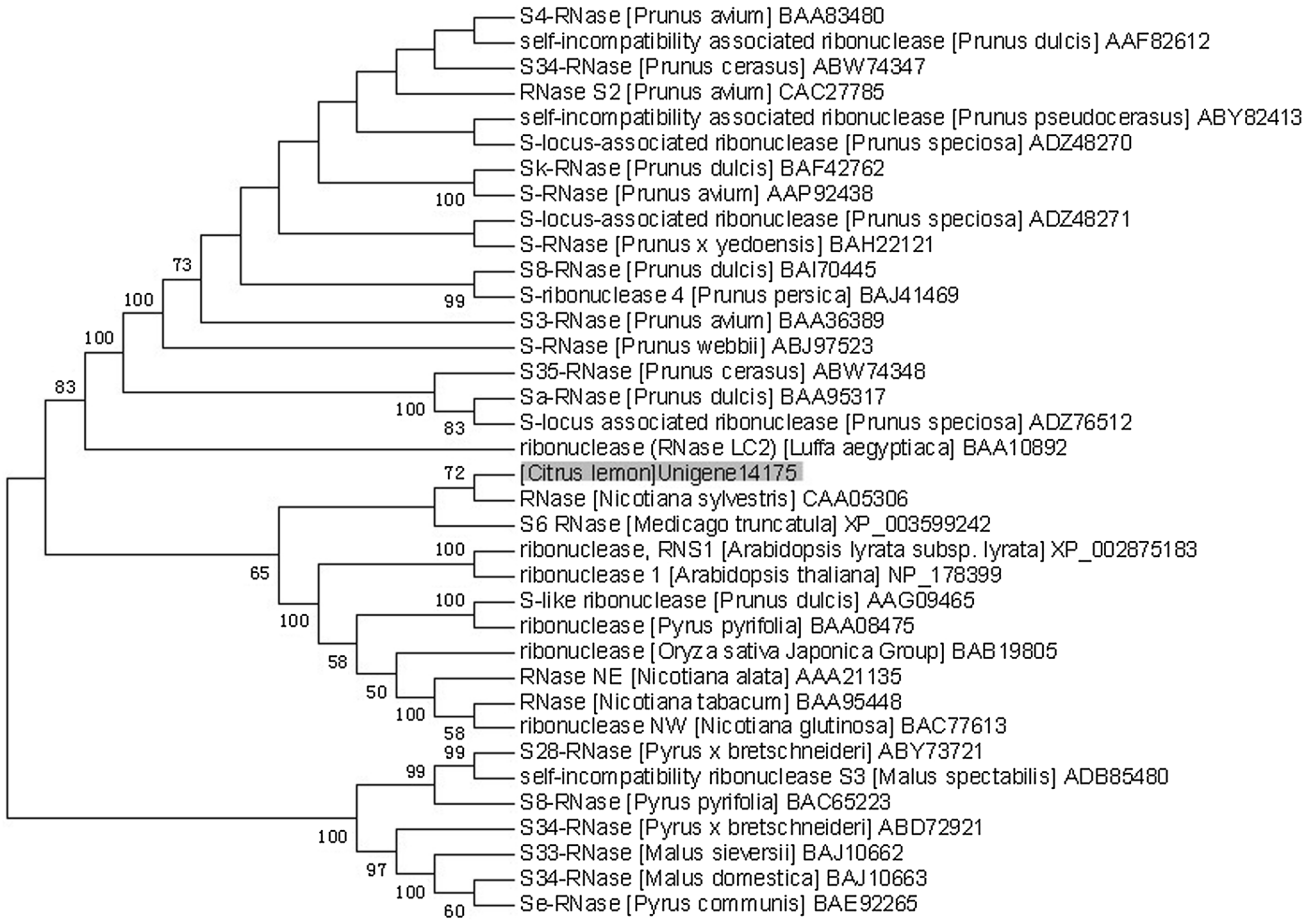



**Fig. 3** Phylogenetic analysis of unigene 14175 and other S-RNases from different species. The neighbor-joining tree based on p-distance and pairwise deletions of gaps indicates phylogenetic relationships among the putative S-RNase proteins in different species. The numbers at the branch points indicate bootstrap support values above 50 % (1,000 replicates). A putative lemon S-RNase protein is *highlighted*

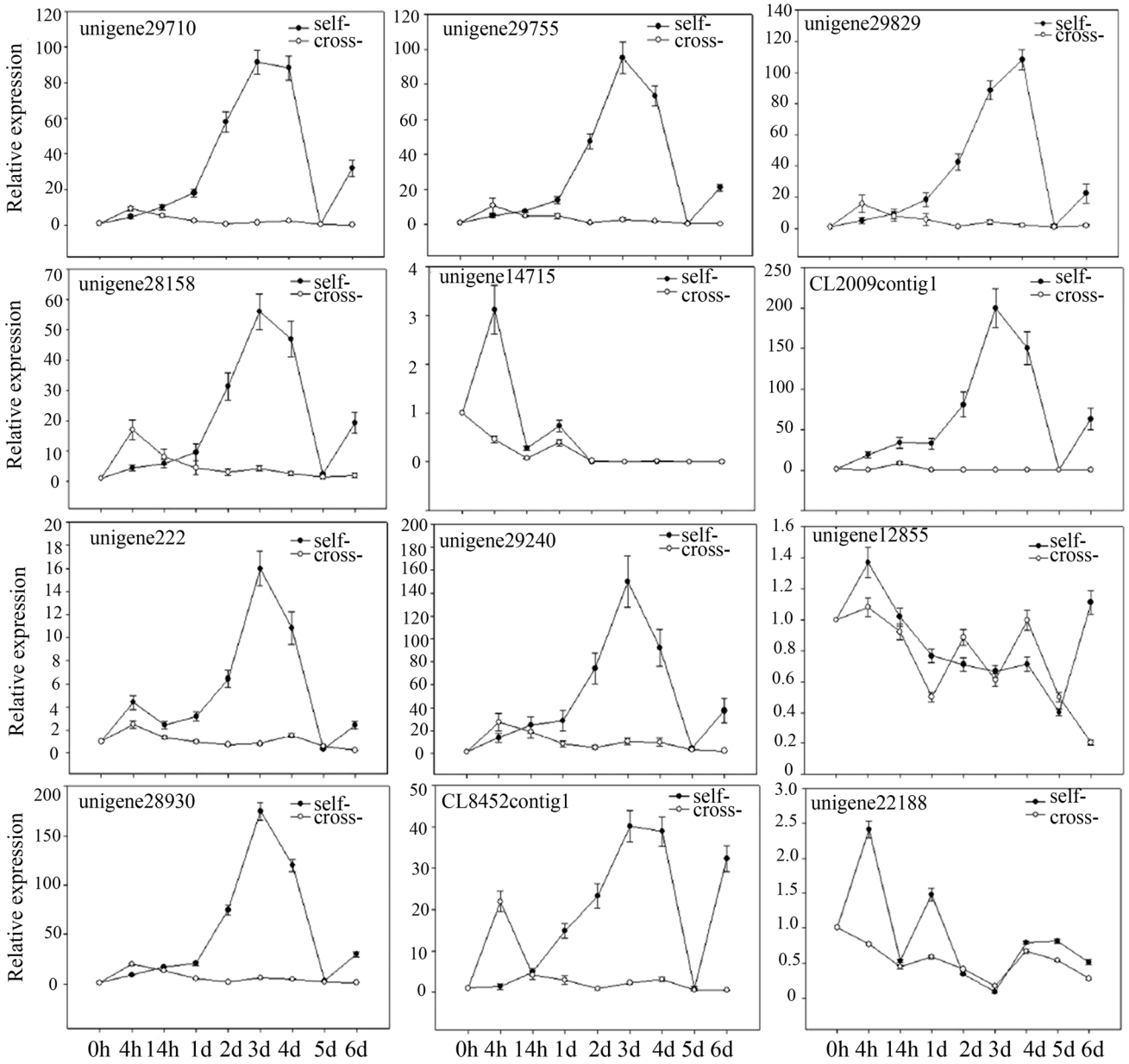



**Fig. 4** Relative gene expressions of selected genes. Relative expression was defined as the expression level, and the *x*-axis indicates days after pollination. *Bars* represent the standard deviations about the mean. Unigene 29710 (U-box domain-containing protein 35-like); Unigene 29755 (Syntaxin-124); Unigene 29829 (cyclic nucleotide-gated ion channel protein); Unigene 28158 (phosphatidylinositol-4-phosphate 5-kinase); Unigene 14715 (a putative *S-RNase*); CL2009 contig1 (F-box protein); Unigene 222 (F-box protein); Unigene 29240 (respiratory burst oxidase homolog protein A); Unigene 12855 (chitin-inducible gibberellin-responsive protein); Unigene 28930 (gibberellin 2-oxidase); CL8452 contig1 (cationic amino acid transporter); Unigene 22188 (F-box family protein)

In the GSI system, a ribonuclease (RNase) plays a key role in self-pollen rejection (McClure 2006). Pistil-expressed S-RNases act as cytotoxins to degrade the RNA of self-pollen tubes. A recent report suggested that S-RNase is associated with the pollen endomembrane system in *Nicotiana* (Goldraij et al. 2006). Pollen-expressed S-locus *F-box* genes act as a component of the SCF (Skp-1-Cullin1-F-box) complex and uniquely ubiquitinate non-self S-RNases, resulting in their degradation by the 26S proteasome. Recently, many S-RNase and F-box genes have been identified in fruit trees, particularly Rosaceae trees (De Franceschi et al. 2011; Gu et al. 2013). Among SI citrus species, an *S-like RNase* gene, *CgSL2* (NCBI accession No. FJ917371) was isolated from pummelo (Chai et al. 2011b). Further studies suggested that these genes might be associated with senescence rather than the incompatibility

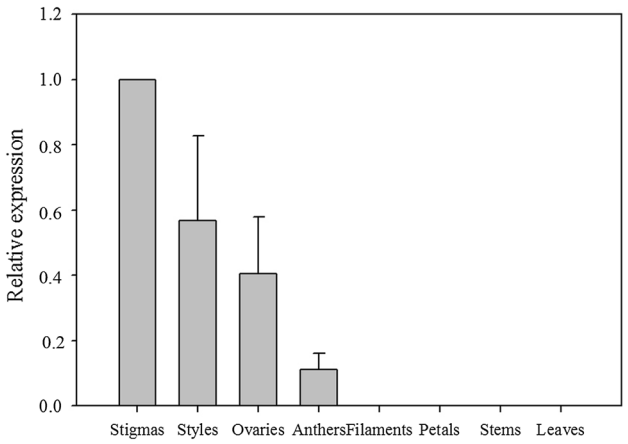



**Fig. 5** Expression analyses of a putative *S-RNase* gene (unigene 14715) in different tissues of 'Xiangshui' lemon

response. Another S-like RNase in 'Wuzishatangju' was cloned from mature pistils. An *F-box* gene designated as *CgF-box* was isolated from the 'Zigui shatian' pummelo (Chai et al. 2011a). *CgF-box* is involved in flower maturation and is thought to play different roles in different tissues. In the self-pollinated sample XX, several Ribonuclease $T_2$ homologous genes were identified. Moreover, *F-box* genes were differentially expressed among the three samples. A putative *S-RNase* gene (unigene 14715) was cloned. Unigene 14715 was highly expressed in self-pollinated stigmas and weakly expressed in cross-pollinated ones (Fig. 4). In addition, the abundance of unigene 14715 in the stigma was 2- and 3-fold higher than that in the style and ovary. It exhibited very low or no expression at other tissues, which was different to pummelo *CgSL2*. Expression of *CgSL2* was detected in petals, stamens, and stem. In pear and apricot, the expression of *S-RNase* genes was highest in the style, and self-pollen tubes were arrested in the styles (Sakurai et al. 2000; Feng et al. 2006; Wang et al. 2010).

Differentially expressed genes were annotated in various biological processes that might be involved in self-pollen rejection. Among these processes, pollen germination and pollen tube elongation are vital for sexual reproduction. The pollen tube growth-related genes might be involved in the abnormal growth of self-pollen in the 'Xiangshui' lemon. In addition, $Ca^{2+}$ plays a key role in pollen germination and tube growth. The tip-focused apical $Ca^{2+}$ gradient is characteristic of growing pollen tubes, whereas it is not observed in incompatible pollen tubes because of an influx of extracellular $Ca^{2+}$ (Franklin-Tong et al. 2002). Pollen tube growth is determined by the fusion of vesicles and the structure of actin filaments. However, to a certain extent, $Ca^{2+}$ levels regulate these processes. During pollen tube growth, $Ca^{2+}$-CaM affects phosphorylation, callose synthesis and vesicle fusion. Nevertheless, high $Ca^{2+}$ levels may interrupt pollen tube growth and cause callose accumulation in the tube tip. As a secondary messenger, $Ca^{2+}$ is also involved in signal transduction, transferring extracellular signals to intercellular signaling pathways. Unigene 30558, unigene 30258, and unigene 29593, which are involved in calcium-mediated signaling, were upregulated in self-pollinated sample XX. DEAD-box ATP-dependent RNA helicase (unigene 28511) and expansion (unigene 2232) were identified and may play an important role in abnormal tube growth. Other genes were also identified that encoded callose synthesis, calmodulin, calcium-dependent protein kinase, and NADPH oxidase. The involvement of these genes may explain our previous observations that pollen tubes in self-pollinated plants were twisted and enlarged (Zhang et al. 2012). Moreover, an exocyst complex component gene (unigene 29572) was upregulated in sample XX and downregulated in sample XB. The exocyst complex is involved in exocytosis and vesicle transport. A *Brassica napus* exocyst complex subunit, named Exo70A1, has been identified. In yeast and animal systems, this subunit is involved in targeted secretion at the plasma membrane by tethering vesicles on the membrane. Exo70A1 is a compatibility factor, and therefore, overexpression in *Brassica* is sufficient to partially overcome the self-pollen rejection response in self-incompatible *Brassica* plants. Further studies are required to clarify the function of the exocyst complex component gene that was upregulated in sample XX.

PCD is an essential element in development in animal and plant cells. However, little is understood about PCD's role during plant reproduction. In *Papaver rhoeas*, the pistil S-locus product is a small protein that interacts with incompatible pollen, triggering a $Ca^{2+}$-dependent signaling network that causes pollen arrest and PCD. By contrast, in the S-RNase-based GSI, proteolysis by the ubiquitin–proteasome pathway and PCD are involved in pollen tube growth. In pear (*Pyrus pyrifolia* L.), the S-RNase depolymerizes the actin cytoskeleton, which triggers mitochondrial alteration and DNA degradation in the incompatible pollen tube, indicating that incompatible pollen growth was terminated by PCD (Wang et al. 2009). In the self-incompatible olive tree (*Olea eurpaea* L.), experiments suggested that PCD is involved in pollen-growth arrest (Irene et al. 2010). In this study, PCD- and cell death-associated proteins, such as myosin heavy chain, hexokinase, protein disulfide isomerase, cationic amino acid transporter, vesicle-associated membrane protein, aspartic proteinase, and cysteine proteases were identified. In sample XX, NADPH oxidase (unigene 29751) was upregulated compared with the other samples. One of the major sources of reactive oxygen species (ROS) in plants is a reaction mediated by NAPDH oxidase. The activity of NADPH oxidase is regulated by calcium ions, signaling phospholipids and possibly Rac/Rop GTPases (Potocky et al. 2012). ROS-mediated signaling

is strongly associated with plant growth and developmental processes such as tip growth, pollen–pistil interaction, defense responses and PCD (Torres et al. 2002; Foreman et al. 2003; Takeda et al. 2008). ROS are necessary for pear pollen tube growth, because pollen tube growth was inhibited when an NADPH oxidase inhibitor is added to the medium (Wang et al. 2010). However, excess ROS can be cytotoxic. Further studies are required to determine whether NADPH oxidase contributes to self-pollen rejection in the 'Xiangshui' lemon.

Gibberellins (GAs) are classical plant hormones essential for plant growth and development. In the pathway of GA biosynthesis, the gibberellin 2-oxidase gene is one of the key enzymes regulating GA biosynthesis. In addition, the *DELLA* gene is central to GA signaling (Peng et al. 1997). At low GA levels, the DELLA protein acts as a growth repressor by blocking the transcriptional response to the hormone. However, the binding of GA to the GA receptor, GID1, promotes the formation of a GA–GID1–DELLA complex that is subsequently recognized by the SCF complex, resulting in the ubiquitin-marked DELLA protein being degraded by the 26S proteasome. A *DELLA* gene was previously isolated in the self-incompatible *Citrus clementina* (Distefano et al. 2009). In this study, gibberellin 2-oxidase (unigene 28930) and DELLA (unigene 12855) were upregulated in sample XX; however, further transcriptional analysis should be performed to identify their functions in the SI response.

The results of transcriptome analysis suggested that multiple genes are associated with self-pollen rejection. The most likely candidate S-locus genes were identified, which have not been reported in citrus. Based on the transcriptome analysis we hypothesize that $Ca^{2+}$, GA, and PCD might be crucial for self-pollen rejection of the 'Xiangshui' lemon. However, further studies are required to fully understand the role of the candidate genes in SI.

## Conclusion

In this study, 61,224 unigenes were obtained with an average length of 948 bp by RNA-seq and their expression level was compared in self-, cross-, and non-pollinated samples. A number of differentially expressed genes were identified. Expression levels of 12 candidate genes were observed to be quite different in self- and cross-pollinated stigmas. A putative *S-RNase* gene (unigene 14715) was expressed at its highest level in stigmas where the self-pollen tubes are mainly arrested in 'Xiangshui' lemon. Further studies are required to determine the functions of these genes in SI. Our research provided abundant transcription level data associated with SI in lemons and could provide clues for the further exploration of SI in lemon species.

**Acknowledgments** We are grateful for the technical support for Illumina sequencing and the initial data analysis we received from the Beijing Genome Institute at Shenzhen, China. This research was supported by Guangxi Science Research and Technology Development Project (1123003-3C), the National Natural Science Foundation of China (31460508), and the Innovation Team of Guangxi Citrus Industry Project.

**Conflict of interest** All authors have read and approved the manuscript. It is not being submitted to any other journal. The authors have declared that no competing interests exist.

## References

Audic S, Claverie JM (1997) The significance of digital gene expression profiles. Genome Res 7(10):986–995

Caruso M, Merelo P, Distefano G, La Malfa S, Piero ARL, Tadeo FR, Talon M, Gentile A (2012) Comparative transcriptome analysis of stylar canal cells identified novel candidate genes implicated in the self-incompatibility response of *Citrus clementina*. BMC Plant Bio 12:20

Chai LJ, Ge XX, Biswas MK, Deng XX (2011a) Molecular analysis and expression of a floral organ-relative F-box gene isolated from 'Zigui shatian' pummelo (*Citrus grandis* Osbeck). Mol Biol Rep 38:4429–4436

Chai LJ, Ge XX, Xu Q, Deng XX (2011b) *CgSL2*, an S-like RNase gene in 'Ziguishatian' pummelo (*Citrus grandis* Osbeck), is involved in ovary senescence. Mol Biol Rep 38:1–8

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21(18):3674–3676

De Franceschi P, Pierantoni L, Dondini L, Grandi M, Sanzol J, Sansavini S (2011) Cloning and mapping multiple S-locus *F-box* genes in European pear (*Pyrus communis* L.). Tree genet Genomes 7:231–240

Distefano G, Caruso M, La Malfa S, Gentile A, Tribulato E (2009) Histological and molecular analysis of pollen-pistil interaction in Clementine. Plant Cell Rep 28:1439–1451

Feng JR, Chen XS, Wu Y, Liu W, Liang Q, Zhang LJ (2006) Detection and transcript expression of S-RNase gene associated with self-incompatibility in apricot (*Prunus armeniaca* L.). Mol Biol Rep 33:215–221

Foreman J, Demidchik V, Bothwell JHF, Mylona P, Miedema H, Torres MA, Linstead P, Costa S, Brownlee C, Jones JDG, Davies JM, Dolan L (2003) Reactive oxygen species produced by NADPH oxidase regulate plant cell growth. Nature 422:442–446

Franklin-Tong VE, Holdaway-Clarke TL, Straatman KR, Kunkel JG, Hepler PK (2002) Involvement of extracellular calcium influx in the self-incompatibility response of *Papaver rhoeas*. Plant J 29:333–345

Gil-Amado JA, Gomez-Jimenez MC (2012) Transcriptome analysis of mature fruit abscission control in olive. Plant Cell Physiol 54:244–269

Goldraij A, Kondo K, Lee CB, Hancock CN, Sivaguru M, Vazquez-Santana S, Kim S, Phillips TE, Cruz-Garcia F, McClure B (2006) Compartmentalization of S-RNase and HT-B degradation in self-incompatible *Nicotiana*. Nature 439:805–810

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD, Chen ZH, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652

Gu C, Zhang SL, Huang SX, Heng W, Liu QZ, Wu HQ, Wu J (2010) Identification of S-genotypes in Chinese cherry cultivars (*Prunus pseudocerasus* Lindl.). Tree Genet Genomes 6:579–590

Gu C, Wu J, Du YH, Yang YN, Zhang SL (2013) Two different *Prunus* SFB alleles have the same function in the self-incompatibility reaction. Plant Mol Biol Rep 31:425–434

Irene S, Salvatore P, Adela O (2010) Programmed-cell-death hallmarks in incompatible pollen and papillar stigma cells of *Olea europaea* L. under free pollination. Plant Cell Rep 29:561–572

Lan LF, Li MN, Lai Y, Xu WY, Kong ZS, Ying K, Han B, Xue YB (2005) Microarray analysis reveals similarities and variations in genetic programs controlling pollination/fertilization and stress responses in rice (*Oryza sativa* L.). Plant Mol Biol 59:151–164

Li CY, Deng GM, Yang J, Viljoen A, Jin Y, Kuang RB, Zuo CW, Lv ZC, Yang QS, Sheng O, Wei YR, Hu CH, Dong T, Yi GJ (2012) Transcriptome profiling of resistant and susceptible Cavendish banana roots following inoculation with *Fusarium oxysporum* f. sp. *cubense* tropical race 4. BMC Genom 13:374

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2 (-Delta Delta C (T)) method. Methods 25:402–408

McClure B (2006) New views of S-RNase-based self-incompatibility. Curr Opin Plant Biol 9:639–646

Miao HX, Qin YH, Ye ZX, Hu GB (2013a) Molecular characterization and expression analysis of ubiquitin-activating enzyme E1 gene in *Citrus reticulata*. Gene 513:249–259

Miao HX, Qin YH, Da Silva JAT, Ye ZX, Hu GB (2013b) Identification of differentially expressed genes in pistils from self-incompatible *Citrus reticulata* by suppression subtractive hybridization. Mol Biol Rep 40(1):159–169

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5:621–628

Peng JR, Carol P, Richards DE, King KE, Cowling RJ, Murphy GP, Harberd NP (1997) The *Arabidopsis GAI* gene defines a signaling pathway that negatively regulates gibberellin responses. Genes Dev 11:3194–3205

Potocky M, Pejchar P, Gutkowska M, Jimenez-Quesada MJ, Potocka A, Alche JD, Kost B, Zarsky V (2012) NADPH oxidase activity in pollen tubes is affected by calcium ions, signaling phospholipids and Rac/Rop GTPases. J Plant Physiol 169:1654–1663

Sakurai K, Brown SK, Weeden N (2000) Self-incompatibility alleles of apple cultivars and advanced selections. Hort Sci 35:116–119

Schuster M (2012) Incompatible (S-) genotypes of sweet cherry cultivars (*Prunus avium* L.). Sci Hortic 148:59–73

Shi CY, Yang H, Wei CL, Yu O, Zhang ZZ, Jiang CJ, Sun J, Li YY, Chen Q, Xia T, Wan XC (2011) Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. BMC Genom 12:131

Sun XD, Zhou SM, Meng FL, Liu SQ (2012) De novo assembly and characterization of the garlic (*Allium sativum*) bud transcriptome by Illumina sequencing. Plant Cell Rep 31:1823–1828

Takeda S, Gapper C, Kaya H, Bell E, Kuchitsu K, Dolan L (2008) Local positive feedback regulation determines cell shape in toot hair cells. Science 319:1241–1244

Torres MA, Dangl JL, Jones JDG (2002) *Arabidopsis* gp91phox homologues AtrbohD and AtrbohF are requited for accumulation of reactive oxygen intermediates in the plant defense response. Proc Natl Acad Sci 99(1):517–522

Tung CW, Dwyer KG, Nasrallah ME, Nasrallah JB (2005) Genome-wide identification of genes expressed in *Arabidopsis* pistils specifically along the path of pollen tube growth. Plant Physiol 138:977–989

Venturini L, Ferrarini A, Zenoni S, Tornielli GB, Fasoli M, Dal Santo S, Minio A, Buson G, Tononi P, Zago ED, Zamperin G, Bellin D, Pezzotti M, Delledonne M (2013) De novo transcriptome characterization of *Vitis vinifera* cv. Corvina unveils varietal diversity. BMC Genomics 14:41

Wang CL, Xu GH, Jiang XT, Chen G, Wu J, Wu HQ, Zhang SL (2009) S-RNase triggers mitochondrial alteration and DNA degradation in the incompatible pollen tube of *Pyrus pyrifolia* in vitro. Plant J 57:220–229

Wang CL, Wu J, Xu GH, Gao YB, Chen G, Wu JY, Wu HQ, Zhang SL (2010) S-RNase disrupts tip-localized reactive oxygen species and induces nuclear DNA degradation in incompatible pollen tubes of *Pyrus pyrifolia*. J Cell Sci 123:4301–4309

Wang PP, Shi T, Zhuang WB, Zhang Z, Gao ZH (2012) Determination of S-Rnase genotypes and isolation of four novel S-RNase genes in Japanese apricot (*Prunus mume* Sieb. Et Zucc.) native to China. J Hortic Sci Biotech 87(3):266–270

Wu J, Gu C, Khan MA, Wu JY, Gao YB, Wang CL, Korban SS, Zhang SL (2013) Molecular determinants and mechanisms of gametophytic self-incompatibility in fruit trees of Rosaceae. Crit Rev Plant Sci 32(1):53–68

Ye J, Fang L, Zheng HK, Zhang Y, Chen J, Zhang ZJ, Wang J, Li ST, Li RQ, Bolund L, Wang J (2006) WEGO: a web tool for plotting GO annotations. Nucleic Acids Res 34:W293–W297

Zhang SJ, Huang SX, Heng W, Wu HQ, Wu J, Zhang SL (2008) Identification of S-genotypes in 17 Chinese cultivars of Japanese plum (*Prunus salicina* Lindl.) and molecular characterization of 13 novel S-alleles. J Hortic Sci Biotech 83(5):635–640

Zhang SW, Huang GX, Ding F, He XH, Pan JC (2012) Mechanism of seedlessness in a new lemon cultivar 'Xiangshui' [*Citrus limon* (L.) Burm. F.]. Sex Plant Reprod 25:337–345

Zisovich AH, Stern RA, Sapir G, Shafir S, Goldway M (2004) The RHV region of S-RNase in the European pear (*Pyrus communis*) is not required for the determination of specific pollen rejection. Sex Plant Reprod 17:151–156