Shot-gun proteome and transcriptome mapping of Jujube floral organ and identification of a pollen-specific S-locus F-box gene

Ruihong Chen [1], **Guoliang Chen** [2], **Jian Huang** Corresp. [3]

[1] Shaanxi Province Key Laboratory of jujube, College of Life Science, Yan'an University, Yan'an, China
[2] Shaanxi Province Key Laboratory of jujube, College of Life Science, Yan'an University, Yan'an, China
[3] College of Forestry, Northwest A & F University, Yangling, China

Corresponding Author: Jian Huang
Email address: huangj@nwsuaf.edu.cn

The flower is a plant reproductive organ that forms part of the fruit produced as the flowering season ends. While the number and identity of proteins expressed in a Jujube (*Ziziphus jujuba* Mill.) flower is currently unknown, integrative proteomic and transcriptomic analyses provide a systematic strategy of characterizing the floral biology of plants. We conducted a shotgun proteomic analysis on Jujube flowers by using a filter-aided sample preparation tryptic digestion, followed by LC-MS/MS. In addition, transcriptomics analyses were performed on HiSeq2000 sequencers. In total, 7853 proteins were identified accounting for nearly 30% of the 'Junzao' gene models (27,443). Genes identified in proteome generally showed higher RPKM (Reads Per Kilobase per Million mapped reads) values than undetected genes. Gene ontology categories showed that ribosomes and intracellular organelles were the most dominant classes and accounted for 17.0% and 14.0% of the proteome mass, respectively. The top-ranking proteins with iBAQ > 10^{10} included non-specific lipid transfer proteins, histones, actin-related proteins, fructose-bisphosphate aldolase, Bet v I type allergens, etc. In addition, we identified one pollen-specificity S-locus F-box-like gene located on the same chromosome as the S-RNase gene. Both of these may activate the behaviour of gametophyte self-incompatibility in Jujube. These results reflected the protein profile features of Jujube flowers and contributes new information important to the Jujube breeding system.

Shot-gun proteome and transcriptome mapping of Jujube floral organ and identification of a pollen-specific S-locus F-box gene

Ruihong Chen[1] Guoliang Chen[1] Jian Huang[1,2]

Ruihong Chen: chenruihong328 @163.com;

Guoliang Chen: glc9359@163.com

Jian Huang: huangj@nwsuaf.edu.cn

Correspondence: huangj@nwsuaf.edu.cn

[1] Shaanxi Province Key Laboratory of jujube, College of Life Science, Yan'an University, Yan'an,716000, China

[2] College of Forestry, Northwest A & F University, Yangling, 712100, China

Abstract

The flower is a plant reproductive organ that forms part of the fruit produced as the flowering season ends. While the number and identity of proteins expressed in a Jujube (*Ziziphus jujuba* Mill.) flower is currently unknown, integrative proteomic and transcriptomic analyses provide a systematic strategy of characterizing the floral biology of plants. We conducted a shotgun proteomic analysis on Jujube flowers by using a filter-aided sample preparation tryptic digestion, followed by LC-MS/MS. In addition, transcriptomics analyses were performed on HiSeq2000 sequencers. In total, 7853 proteins were identified accounting for nearly 30% of the 'Junzao' gene models (27,443). Genes identified in proteome generally showed higher RPKM (Reads Per Kilobase per Million mapped reads) values than undetected genes. Gene ontology categories showed that ribosomes and intracellular organelles were the most dominant classes and accounted for 17.0% and 14.0% of the proteome mass, respectively. The top-ranking proteins with iBAQ > 10^{10} included non-specific lipid transfer proteins, histones, actin-related proteins, fructose-bisphosphate aldolase, Bet v I type allergens, etc. In addition, we identified one pollen-specificity S-locus F-box-like gene located on the same chromosome as the S-RNase gene. Both of these may activate the behaviour of gametophyte self-incompatibility in Jujube. These results reflected the protein profile features of Jujube flowers and contributes new information important to the Jujube breeding system.

Abbreviations

LC-MS/MS – Liquid Chromatography-tandem Mass Spectrometry

MS – Mass Spectrometry

RPKM – Reads Per Kilobase per Million mapped reads

FDR – False discovery rate

CBB – Coomassie Brilliant Blue

SDS-PAGE – Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis

MW – Molecular Weight

GO – Gene Ontology

qRT-PCR – Quantitative real-time reverse transcription-PCR

iBAQ – Intensity Based Absolute Quantification

Introduction

Chinese Jujube (*Ziziphus jujuba* Mill.) is one of the most popular fruit trees in China. The fruits are valued for their good taste, high nutritional and medical value (Liu et al. 2014; Huang et al. 2016). Increasing economic value of Chinese Jujube has resulted in an increase in cultivation acreage.

Flowers are a critical reproductive organ and are the origin of the valuable fruits. They begin to develop in early June, originating from the leaf axil with shoot growth, and continue to be produced until the end of October in northern China (Niu et al. 2015). Jujube flowers also produce abundant nectar, and honey made from Jujube nectar is one of the most popular types of honey in China. However, Jujube flowers are small and drop easily; 87.9-99.9% of the flowers and 68.5% of young fruits usually drop during development, and only 1.1% of the fruits reach maturity on average (Liu et al. 2009). Therefore, understanding Jujube floral biology is important and may provide insights into fruit setting and Jujube breeding. Although several studies have focused on the biology of flower differentiation (He et al. 2009; Song et al. 2013; Niu et al. 2011), a global analysis of protein expression in Jujube flowers is needed to understand floral biology.

Proteomics provide a comprehensive and quantitative approach to identify proteins that are expressed in a given organ, tissue, or cell line. With the advance in high-resolution mass spectrometry (MS) - based shotgun proteomics (Mann et al. 2013), several complete proteome maps have been assembled in recent years. However, these studies were restricted to model organisms (Wilhelm et al. 2014; Kim et al. 2014; Baerenfaller et al. 2008). Recently, whole-genome sequencing has been applied to more organisms, and the developed gene models have provided information on the theoretical protein-coding capacity of an organism. These advances have enabled complete proteome analysis for more species. Since the first proteome map of a model plant species, *Arabidopsis thaliana,* was assembled (Baerenfaller et al. 2008), rapid progress in plant proteome has been made in various plant species at different tissue levels, such as fruit, seeds, flowers and leaves roots (Zeng et al. 2011; Girolamo et al. 2012; Dai et al. 2013; Zhu et al 2015). Undoubtedly, sequencing of the Jujube genome will also contribute to the proteomics study of jujube (Liu et al. 2014).

Some genes specifically expressed in floral organs have been shown to contribute to flower development and reproductive function. In a previous study, gametophyte self-incompatibility (GSI) induced in *Z. jujuba* led to high percentage of seedless fruits (Asatryan and Tel-Zur 2013). Self-incompatibility/compatibility is controlled by the pistil S determinant ribonuclease (S-RNase) and the pollen S determinant F-box (SFB) gene(s). The GSI has been well characterized in some core-eudicots, such as *Solanaceae*, *Plantaginaceae* and *Rosaceae*. We have previously identified an S-RNase candidate gene in the chromosome 1 of the assembled 'Junzao' Jujube genome (Huang et al. 2016). However, the SFB candidate genes specifically expressed in pollens have not been identified.

A combination of transcriptome and proteome technologies offers a new comprehensive approach to identify the genes involved in developmental systems and environmental responses. This study

100 is the first to generate a draft proteome map of Jujube flowers based on the liquid chromatography-
101 tandem mass spectrometry (LC-MS/MS). To further improve the molecular knowledge of GSI in
102 *Z. jujuba*, we attempt to identify the SFB gene(s) specially expressed in anthers/pollens linked to
103 the S-RNase gene. Information from our study will provide a better understanding of Jujube flower
104 development and the breeding system.
105

106 **Materials and methods**

107 **Sample collection**

108 Budding flowers were collected from nine year old Jujube trees (*Ziziphus jujuba* 'Junzao') in 2013.
109 We collected two biological replicates of flower samples from each of two trees. The collected
110 flowers were immediately immersed in liquid nitrogen, and stored at - 80°C for transcriptome and
111 proteome analysis.
112

113 **Protein extraction and quality determination**

114

115 We selected one sample replicate for proteome analysis. The trichloroacetic acid (TCA)/acetone
116 precipitation method was used to extract protein from Jujube flowers (Wu et al. 2014). Briefly,
117 frozen flowers were ground in liquid nitrogen and precipitated with TCA/acetone (1:9) overnight
118 at −20 °C. Then the homogenate was centrifuged (7000× g, 30 min) and washed three times with
119 acetone. The precipitated proteins were freeze-dried and then re-suspended in 800 μl of SDT buffer
120 (4% SDS, 100 mM DTT, 150 mM Tris-HCl, pH 8.0). After 15 min of incubation in boiling water,
121 the homogenate was fractionated by ultrasound (80 W, 10 s, 10 times) on ice. The homogenate
122 was centrifuged at 14 000 × g at 4 °C for 30 min, and the supernatant was collected. Finally, the
123 protein concentration was determined using a bicinchoninic acid protein assay kit (BCA,
124 Beyotime, China) and adjusted to 5 μg/μl. Protein was also analysed by electrophoresis on a 12.5%

sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) gel (7 cm × 20 cm) and stained with Coomassie Brilliant Blue (CBB) staining buffer (Gelcode blue, Pierce, USA).

In-solution protein digestion

Proteins were digested by following the filter-aided sample preparation (FASP) procedure (Wisniewski et al. 2009). Total-protein samples (50 µg) were incorporated into 30 µl of SDT buffer (4% SDS, 100 mM DTT, 150 mM Tris-HCl, pH 8.0) and then heated at 95 °C for 5 min. Thereafter, 200 µl of UA buffer (8 M urea, 150 mM Tris-HCl pH 8.0) was added to the protein solution, mixed to remove the detergent, DTT, and passed through an ultrafiltration unit with a nominal molecular weight (MW) cut-off of 30 kDa (Cat No. MRCF0R030, Millipore) to remove other low-molecular-weight components. Then, 100 µl of 0.05 M iodoacetamide in UA buffer was added, and the samples were incubated for 20 min in the dark. The filter was washed three times with 100 µl of UA buffer, followed by two washes with 100µl of 25 mM NH_4HCO_3. Finally, 2 µg of trypsin (Promega) was added, the protein suspension was digested overnight at 37 °C and the resulting peptides were collected as the filtrate.

Peptide Fractionation with Strong Cation Exchange (SCX) Chromatography

Peptides were fractionated by SCX chromatography using an AKTA Purifier system (GE Healthcare). The dried peptide mixture was reconstituted and acidified with 2 ml of buffer A (10 mM KH_2PO_4 in 25% of acetonitrile (ACN), pH 2.7) and loaded onto a Poly SULFOETHYL 4.6 × 100 mm column (5 µm, 200 Å, PolyLC Inc, Maryland, U.S.A.). The peptides were eluted at a flow rate of 1 ml/min using a gradient of 0 –10% buffer B (500 mM KCl, 10 mM KH_2PO_4 in 25% of ACN, pH 2.7) for 2 min, 10–20% buffer B for 25 min, 20 – 45% buffer B for 5 min, and 50% – 100% buffer B for 5 min. The eluents were monitored by absorbance at 214 nm and fractions were collected every 1 min. The collected fractions (approximately 30) were combined into 25 pools and desalted on C18 Cartridges (Empore™ SPE Cartridges C18 (standard density), bed I.D. 7 mm,

151 volume 3 ml, Sigma). Each fraction was concentrated by vacuum centrifugation and reconstituted
152 in 40 µl of 0.1% (v/v) trifluoroacetic acid. All samples were stored at −80 °C until LC-MS/MS
153 analysis could be conducted.

154 **Liquid Chromatography (LC) - Electrospray Ionization (ESI) Tandem MS (MS/MS)**
155 **Analysis by Q Exactive.**

156 Experiments were performed on a Q Exactive mass spectrometer that was coupled to a nanoflow
157 HPLC (Easy nLC, Proxeon Biosystems, now Thermo Fisher Scientific). Each fraction (6µl) was
158 injected for Nano LC-MS/MS analysis. The peptide mixture (5 µg) was loaded onto a C18-
159 reversed phase column (Thermo Scientific Easy Column, 10 cm long, 75 µm inner diameter, 3µm
160 resin) in buffer A (0.1% formic acid) and separated using a linear gradient of buffer B (80%
161 acetonitrile and 0.1% formic acid) at a flow rate of 250 nl/min controlled by IntelliFlow technology
162 over 60 min.
163
164 MS data were acquired dynamically using a data-dependent top10 method by choosing the most
165 abundant precursor ions from the survey scan (300 – 1800 m/z) for HCD fragmentation. Target
166 value determination was based on predictive automatic gain control (pAGC). The dynamic
167 exclusion duration was 60 s. Survey scans were acquired at a resolution of 70 000 at 200 m/z , and
168 for HCD spectra, the resolution was set to 17 500 at 200 m/z. The normalized collision energy was
169 30 eV, and the underfill ratio, which specifies the minimum percentage of the target value likely
170 to be reached at maximum fill time, was defined as 0.1%. The instrument was run in peptide
171 recognition mode.
172

173 **RNA extraction and Transcriptome analyses**

174 Total RNA of whole budding flowers was isolated using an RNA prep Pure Tissue Kit (Tiangen,
175 Beijing) for RNA extraction and then treated with RNase-free DNase I treatment (Promega, USA).
176 RNA purity was checked using the NanoPhotometer® spectrophotometer (IMPLEN, CA, USA).
177 RNA concentration was measured using Qubit 2.0 Flurometer (Life Technologies, CA, USA).
178 RNA integrity was assessed using an RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system
179 (Agilent Technologies, CA, USA). Finally, RNA degradation and contamination was visualized
180 on 1% agarose gels. A total of 3 μg of RNA per sample was used as input material for the RNA
181 sample preparations. Sequencing libraries were generated using NEBNext® Ultra™ RNA Library
182 Prep Kit for Illumina® (NEB, USA) according to manufacturer's recommendation. The libraries
183 were sequenced on an Illumina HiSeq 2000 platform and 100 bp paired-end reads were
184 generated. The raw-sequence reads data are deposited in NCBI Sequence Read Archive
185 (SRX1518648 and SRX1518646).
186
187 After quality control of raw reads, filtered reads were mapped to the 'Junzao' genome using
188 Tophat2 (Kim et al. 2013). Total aligned read numbers were normalized by gene length and
189 sequencing depth for an accurate estimate of expression levels (Robinson et al. 2010). HTSeq 0.6.1
190 was used to count the reads mapping to each gene (Anders et al. 2014). In addition, the RPKM
191 (Reads Per Kilobase per Million mapped reads) of each gene was calculated based on the length
192 of the gene to represent the expression level (Mortazavi et al. 2008).
193
194 **Sequence Database Searching and Data Analysis**
195
196 The mass spectrometric raw data from the top10 method were analysed using MaxQuant 1.3.0.5
197 (Cox and Mann 2008). MS/MS data were searched against the gene predicted gene models of
198 *Ziziphus jujuba* 'Junzao' (Accession: LPXJ00000000 and Dryad Digital Repository.

199 doi:10.5061/dryad.83fr7). An initial search was set at a precursor mass window of 6 ppm. The
200 search followed an enzymatic cleavage rule of Trypsin/P and allowed a maximum of two missed
201 cleavage sites and a mass tolerance of 20 ppm for fragment ions. Enzyme = Trypsin, Missed
202 cleavage = 2, Fixed modification: Carbamidomethyl (C), Variable modification: Oxidation (M),
203 Decoy database pattern = decoy. The cut-off of global false discovery rate (FDR) for peptide and
204 protein identification was set to 0.01. The iBAQ algorithm was used to calculate the approximate
205 abundance of each protein, which normalizes the summed peptide intensities by the number of
206 theoretically observable peptides of the protein (Schwanhausse et al. 2011).

207 **Gene Ontology (GO) term enrichment analysis of expressed genes and proteins**

208 GO term enrichment analysis was conducted for expressed genes and proteins in the Jujube flower
209 using the R package TopGO, and is available from Dryad Digital Repository
210 (doi:10.5061/dryad.83fr7). GO term enriched results for the transcriptome and proteome were
211 compared for the three aspects of cell components, molecular function and biological progress.

212 **Gene expression validation by quantitative Real-time PCR (qRT-PCR) Analysis**

213 We performed a qRT-PCR analysis to validate the relative gene expression levels in the flowers.
214 RNA samples from the same flower samples used for RNA-sequencing analysis were extracted
215 using the same method as described above. cDNA was synthesized using a PrimeScript^{TM} RT
216 reagent kit with gDNA Eraser (TaKaRa, Dalian, China) according to the manufacturer's
217 instruction. We selected 30 genes with different levels of RPKM and iBAQ values for qRT-PCR
218 analyses (Table S1). The primers were designed using Primer-BLAST online software
219 (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). Primer specificities were initially verified by
220 PCR amplification of a single product and product size was visualized after electrophoresis on a
221 1.2% agarose gel. The cDNAs were then amplified in a Bio-Rad CFX Real-Time System (Bio-
222 Rad, Hercules, CA, USA) with SYBR Premix Ex Taq TM II (Takara, Dalian, China). The qRT-
223 PCR amplification protocol was set at 95°C for 3 min, 95 °C for 10 s, 40 cycles of 95 °C for 10 s

and 58 °C for 10 s, and finally, 72°C for 30 s. Three replicates were performed for each gene. The UBQ gene (Zj.jz021445712) of Jujube was chosen as an internal reference gene for normalization (Zhang et al. 2015). The average threshold cycle (Ct) was normalized, and the relative expression was calculated using the $2^{-\Delta Ct}$ method.

Identification of S-locus F-box （SFB） genes

We used previously reported S-locus F-box genes as queries to BLAST search 'Junzao' gene models, such as those from *Plantaginaceae* (6 S-locus F-box (SLF) genes), Solanaceae (35 SLFs), and Maleae (22 S-locus F-box brothers (SFBB)), 8 SLF-like genes and 3 SFB genes of *Prunus*. Because we have identified the S-RNase gene (Zj.jz035833030) located on the linkage group (LG) 1 of the assembled 'Junzao' genome, we treated the newly identified F-box genes located on LG1 as potential SFB genes and then subjected them to phylogenetic analysis with the known reported S-locus F-box genes. All the amino acid sequences were aligned and a neighbour-joining tree was constructed using Mega 6.0.6 (Tamura et al. 2013). Then, RPKM values of those F-box genes were calculated based on the transcriptomic analysis of leaves, flowers, shoot and fruits based on the previous report (Huang et al. 2016). We further performed tissue-specific expression analysis of those F-box genes among different tissues/organs (leaves, shoot, young fruit, ripe fruit, pollen, pistil, petal, flower disc) using RT-PCR. RNA extraction, cDNA synthesis and primer design (Table S3) were performed as described above. PCR was performed in using 2 × GoldStar Best MasterMix (CWBIO, Beijing) according to its protocol. PCR products were visualized on a 1.2% agarose gel after electrophoresis.

Results

Transcriptome and Proteome analysis of Jujube flowers

RNA sequencing of two independent flower samples (flower 1 and 2) on an Illumina Hiseq 2000 platform generated 52.8 and 59.4 million clean paired-end reads after quality control for flower 1 and 2, respectively (Table S1). In general, 90.1% of the reads from flower 1 and 92.2% of flower 2 were mapped onto the 'Junzao' Jujube gene models, and 88.1% and 90.3% of reads were mapped to unique sites, respectively. A total of 18 646 (67.9%) and 17 995 (65.6%) genes had RPKM values ≥ 1 in two samples, respectively (Table S2). The two samples showed a strong correlation (Pearson $r = 0.96$, $P < 0.01$) on the RPKM metrics. In addition, the rank pattern of expression levels of the selected 30 genes generated by qRT-PCR analysis showed a linear correlation with their RPKM values ($r = 0.95$, $P < 0.01$). Taken together, both correlation analyses and independent qRT-PCR evaluation confirmed the reproducibility and validity of the relative expression levels based on the RPKM methods in the transcriptome analysis (Table S2).

In total, 80 802 MS/MS spectra were obtained and evaluated against the 'Junzao' Jujube gene prediction database. We identified 7853 proteins with 42 925 unique peptides originating from 80 394 tandem mass spectrometry spectrum assignments with a false-discovery rate (FDR) below 1%, which accounts for nearly 30% of all predicted 'Junzao' gene models (27 443).

Comparison of the proteome and transcriptome

We compared the transcriptome and proteome data generated from the flower samples based on the 'Junzao' gene models. As shown in the Venn diagram (Fig. 1), 94% of genes identified by MS as protein could also be detected at the mRNA expression level (RPKM ≥1), and only 80 genes were missing. The identified proteins accounted for 41.2% of the transcripts with RPKM values ≥ 1.

268

269 The RPKM values of all detected genes show a bimodal distribution (Fig. 2A) in which 81.27%
270 (17 995) and 64.08% (14, 182) of genes had signals of RPKM values ≥ 1 and 5, respectively. When
271 the criterion was set at 5, the left shoulder of the bimodal distribution line disappeared and
272 transformed into a Gaussian distribution. By comparison, the identified protein abundance showed
273 a broader distribution range than the filtered mRNA abundance (Fig. 2A, B). Among the 7 497
274 genes identified in the proteomic analysis, only 129 (1.6%) and 473 (6.0%) of genes had RPKM
275 values < 1 and 5, respectively. RPKM values of the genes identified in the proteome or those not
276 identified differ clearly (Fig. 2C). We also found that the RPKM-based transcript abundance values
277 correlated (Spearman's correlation 0.5) well with iBAQ-based protein abundance values in Jujube
278 flowers (Fig. 2D). The top10 categories of GO for the proteome and the corresponding
279 transcriptome in the three aspects of cell components, biological progress and molecular functions
280 are shown in figure 3. The ranking patterns were generally similar in the transcriptome and
281 proteome in the three aspects, and the highest-ranking categories were consistent among the three
282 aspects.

283 **Protein abundance**

284 The iBAQ values are considered as estimates of the absolute amount of each protein. The most
285 abundant 224 proteins (2.85%) constituted 50% of the Jujube flower proteome, and the 691 most
286 abundant proteins, representing 80% abundance, dominated the total protein abundance (Fig.4).
287 The 41 most abundant proteins comprised 20% of the proteome. The top7 proteins with iBAQ
288 values > 10^{10} included two non-specific lipid transfer proteins, histones, actin-related proteins,
289 fructose-bisphosphate aldolase, Bet v I type allergens and unknown proteins. However, ribosomal
290 proteins with six members showed the most abundant protein mass (50 381 000 000) within the
291 top20 proteins (Supplementary Data).

292

293 Abundance values of each protein reflect their contribution to the total proteome (Nagaraj et al.
294 2011). As shown in the GO category, more members of those proteins did not necessarily possess
295 a higher proportion of the protein mass (Fig S1. A). Some protein classes contributed a
296 significantly higher proportion of the total protein mass compared with the corresponding
297 identified protein number. Based on the cellular components, we identified 175 (2.3%) different
298 ribosomes accounting for 17% of the total protein mass in our data (FigS1. A). Similarly, the
299 intracellular components with 230 genes (3.1%) contributed 13.96% of the proteome mass. This
300 large difference was also observed in the "structural constituents of ribosomes" belonging to
301 'molecular function' and 'translation' belonging to 'biological progress' (Fig S1. B, C). In contrast,
302 integral membrane proteins accounted for 2.3% of the protein mass but contributed much less to
303 the proteome (1.0% of total protein mass).

304 **Identification of SFB gene**

305 We first identified seven candidate SFB genes (Table S3) in LG1 around S-RNase
306 (Zj.jz035833030) by Blastp analysis. Phylogenetic analysis showed that S-locus F-box genes
307 belonging to *Plantaginaceae*, *Solanaceae*, and *Maleae* clustered into one group. Jujube SFB genes
308 were located between *Prunus* SFB genes and SLFL genes (Fig. 5). According to the transcriptomic
309 analysis, we found that No. 5 (Zj.jz026591005) and No. 7 (Zj.jz039541036) were slightly
310 expressed in budding flowers while the other four genes were expressed in several tissues (Fig.
311 6A). RT-PCR analyses showed that Zj.jz039541036 was strongly expressed in pollens while
312 Zj.jz026591005 was only weakly detected in all flower organs. In contrast to RT-PCR results,
313 Zj.jz027555002 was not detected in any of the tissues (Fig. 6B). Therefore, we consider
314 Zj.jz039541036 as the candidate SFB gene of *Z. jujuba*.

315 **Discussion**

316 In this study, we used an FASP-SCX-LC-MS/MS method to resolve the proteome of Jujube flower

317 organs. In total, we identified 7497 proteins that accounted for nearly 30.0% of all predicted
318 'Junzao' gene models (27 443). This depth was similar to the results of the shotgun proteome study
319 of *Arabidopsis* flowers (Baerenfaller et al. 2008), in which 7040 proteins (26.1% of all gene
320 models) were identified and represented nearly 50% of all identified *Arabidopsis* genes in the
321 proteome map. Compared to the improved LC-MS/MS technologies, two-dimensional
322 electrophoresis/mass spectrometry (2-DGE/MS) and multi-dimensional protein identification
323 technology (MudPIT) approaches only identified 2400 floral proteins in wild *Arabidopsis* flowers
324 (Feng et al. 2009). Lack of genome resources could undermine the efficiency of protein
325 identification. In one study, only 1488 proteins were obtained from the fruit of *Punica granatum*
326 by nanoliquid chromatography high-resolution tandem mass spectrometry (Capriotti et al. 2013),
327 and 554 of the 2043 peptides could not be annotated in the proteome of the midgut of *Spodoptera*
328 *litura* larvae by shotgun ESI-MS (Liu et al. 2010). In the present study, 58% of the genes with
329 RPKM values $\geq$ 1 were still not covered by the proteome in Jujube flowers, suggesting the
330 limitation of the present research strategies. The flower is a complex plant organ that secretes many
331 sugars that likely interfere with protein separation (Righetti and Boschetti 2016). When the flower
332 is separated into different organs, such as pistil, petal, stigma, and style, the depth of the flower
333 proteome may be further increased. In general, plant proteomics often identify fewer proteins than
334 animal and human cell or organ proteomics because of higher amount of proteins in animals
335 compared to the protein levels in plant biomass (for example, cell walls contain 90–95%
336 polysaccharides and 5–10% proteins). Therefore, a new strategy for protein extraction should be
337 considered (e.g., CPLL; Righetti and Boschetti 2016).
338
339 Recent advances in genome sequencing and proteomics make it possible to study protein
340 abundance and its relationship to corresponding mRNA levels. The distribution pattern of all
341 transcripts in transcriptome showed a bimodal distribution. Transcripts in the left part of bimodal
342 distributions (Fig. 2a) possessed less than one copy and were not expressed as functional protein

in tissues (Hebenstreit et al. 2011). When the RPKM value criterion increased to 5, the RPKM values of transcripts showed a Gaussian distribution and were close to the distribution of genes detected in the proteome (Fig. 2a). In addition, the expression levels of those genes that have their corresponding proteins identified in the proteome were generally higher than those of which the proteins were not detected in the proteome (Fig. 2b). Genome-scale proteome analysis of *A. thaliana* preferentially detected proteins that are expressed at higher transcript frequencies (Baerenfaller et al. 2008). Feng et al. (2009) also found that proteins encoded by genes with relatively high levels of expression were detected with greater frequency in the proteome of wild *Arabidopsis* flower investigated by 2-DGE/MS and MudPIT approaches. These results confirm that genes with higher expression levels have a high likelihood of being translated into active protein.

Unlike the observed domination of RuBisCO (ribulose-1,5-bisphosphate carboxylase/oxygenase) in leaf proteins (> 50%) and of storage proteins in seeds (> 80%), in the Jujube flower proteome, ribosomal proteins were the most abundant protein class, accounting for 17% of the total protein mass. This reflects the different biological functions of floral organs in the role of protein composition. Other abundant proteins also contribute to the special features of flowers. Non-specific lipid transfer proteins are abundant and widely distributed in plants, representing as much as 4% of total soluble proteins (Liu et al. 2015). Plant actin-related protein are also an important component and are widely distributed in plant cells. Fructose-bisphosphate aldolase is a key enzyme involved in glycolysis, gluconeogenesis in the cytoplasm, and in the Calvin cycle. Members of fructose-bisphosphate aldolase that are highly expressed in flowers may play an important role in pollen tube growth (Lu et al. 2012; Tang. 2013). In addition, abundant Bet v I type allergens make Jujube flowers an allergen source for humans. This finding could also characterize the feature of Jujube flower based on proteomics.

GSI is an important feature of reproductive biology in core-eudicots and exhibited in *Z. jujuba*. It has been well characterized that GSI is controlled by one pistil-specificity S-RNase and a single (or multiple) pollen–specificity F-box gene(s). We found there were some differences between the RNA-Seq and RT-PCR. According to the transcriptomic analysis, we were failed to identify a flower-specificity F-box gene linked to S-RNase based on the transcriptomic analysis (Fig. 6A). In contrast, we found a single pollen-specific F-box gene (Zj.jz 039541036) based on RT-PCR analysis after dissecting the flower into its component organs. Most likely this is because the RNA-Seq of whole flowers is not reflecting the gene expression level in specific organs or tissues. Phylogenetic analysis revealed that the Jujube SFB gene was close to that of *Prunus* (Fig. 5), in which GSI exhibited a distinct system of recognition between pollen and pistil involving a single SFB gene. This contrasts with multiple F-box genes that are activated in other plants (Matsumoto and Tao 2016). With the formerly identified S-RNase gene in *Z. jujuba*, therefore, we concluded that Jujube GSI behaviour was like *Prunus* and controlled by the S-RNase and a single SFB gene together. This finding will facilitate further characterization of the underlying molecular mechanism of the reproductive biology of *Z. jujuba*.

References

Anders S, Pyl PT, Huber W(2014) HTSeq–A Python framework to work with high-throughput sequencing data, Bioinformatics .31:166-169

Asatryan A, Tel-Zur N. 2013. Pollen tube growth and self-incompatibility in three *Ziziphus* species (Rhamnaceae). Flora - Morphology, Distribution, Functional Ecology of Plants 208:390-399.

Baerenfaller K, Grossmann J, Grobei MA, Hull R, Hirsch-Hoffmann M, Yalovsky S, Zimmermann P, Grossniklaus U, Gruissem W, Baginsky S (2008) Genome-Scale proteomics reveals *Arabidopsis thaliana* gene models and proteome dynamics.Science 320:938-941

Capriotti AL, Caruso G, Cavaliere C, Foglia P, Piovesana S, Samperi R, Laganà A (2013) Proteome investigation of the non-model plant pomegranate (*Punica granatum* L.), Analytical and bioanalytical chemistry 405: 9301-9309

397 Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-
398 range mass accuracies and proteome-wide protein quantification. Nature biotechnology 26:
399 1367-72
400 Dai L, Kang G, Li Y, Nie Z, Duan C, Zeng R(2013) In-depth proteome analysis of the rubber
401 particle of *Hevea brasiliensis* (para rubber tree). Plant Mol Biol 82:155-168
402 Feng B, Li L, Zhou X, Stanley B, Ma H(2009) Analysis of the Arabidopsis floral proteome:
403 detection of over 2000 proteins and evidence for posttranslational modifications. Journal of
404 integrative plant biology 51:207-23
405 Girolamo F, D'Amato A, Righetti PG (2012) Assessment of the floral origin of honey via
406 proteomic tools. Journal of proteomics 75:3688-3693
407 He R, Zhang Q, Jiang X (2009) Observation on flower-bud morph differentiation and flower
408 structure on jujube, Journal of Tarim University 21: 11-14
409 Hebenstreit D, Fang M, Gu M, Charoensawan V, van Oudenaarden A, Teichmann SA (2011) RNA
410 sequencing reveals two major classes of gene expression levels in metazoan cells. Molecular
411 systems biology 7:497
412 Huang J, Zhang CM, Zhao X, ZJ Fei, Wan KK, Zhang Z, Pang XM, YinX, Bai Y, SunXQ, Gao
413 LZ, Li RQ, Zhang JB, Li XG(2016) The Jujube genome provides insights into genome
414 evolution and the domestication of sweetness/acidity taste in fruit trees, Plos Genetics 12(12):
415 e1006433.
416 Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. (2013) TopHat2: accurate
417 alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome
418 Biology 14:R36
419 Kim MS, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar
420 DS, Isserlin R, Jain S, Thomas JK, Muthusamy B, Leal-Rojas P, Kumar P, Sahasrabuddhe
421 NA, Balakrishnan L, Advani J, George B, Renuse S, Selvan LD, Patil AH, Nanjappa V,
422 Radhakrishnan A, Prasad S, Subbannayya T, Raju R, Kumar M, Sreenivasamurthy SK,
423 Marimuthu A, Sathe GJ, Chavan S, Datta KK, Subbannayya Y, Sahu A, Yelamanchi SD,
424 Jayaram S, Rajagopalan P, Sharma J, Murthy KR, Syed N, Goel R, Khan AA, Ahmad S, Dey
425 G, Mudgal K, Chatterjee A, Huang TC, Zhong J, Wu X, Shaw PG, Freed D, Zahari MS,
426 Mukherjee KK, Shankar S, Mahadevan A, Lam H, Mitchell CJ, Shankar SK, Satishchandra
427 P, Schroeder JT, Sirdeshmukh R, Maitra A, Leach SD, Drake CG, Halushka MK, Prasad TS,
428 Hruban RH, Kerr CL, Bader GD, Iacobuzio-Donahue CA, Gowda H, Pandey A.(2014) A draft
429 map of the human proteome.Nature 509:575-581
430 Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary
431 Genetics Analysis version 6.0. Molecular Biology and Evolution:30 2725-2729.
432 Liu F, Zhang X, Lu C, Zeng X, Li Y, Fu D, Wu G (2015) Non-specific lipid transfer proteins in
433 plants: presenting new advances and an integrated functional analysis, Journal of experimental
434 botany .66: 5663-5681

435 Liu J, Zheng S, Liu L, Li L,Feng Q (2010) Protein Profiles of the Midgut of *Spodoptera litura*
436 larvae at the sixth instar feeding stage by shotgun ESI−MS approach. Journal of Proteome
437 Research 9:2117-2147
438 Liu MJ, Zhao J, Cai QL, Liu GC, Wang JR, Zhao ZH, Liu P, Dai L, Yan G, Wang WJ, Li XS,
439 Chen Y, Sun YD, Liu ZG, Lin MJ, Xiao J, Chen YY, Li XF3, Wu B, Ma Y, Jian JB, Yang W,
440 Yuan Z, Sun XC, Wei YL, Yu LL, Zhang C, Liao SG, He RJ, Guang XM, Wang Z, Zhang
441 YY, Luo LH (2014) The complex jujube genome provides insights into fruit tree
442 biology.Nature communications 55:315
443 Liu P, Liu MJ, Zhao ZH, Liu XY, Wang JR,Yan C (2009) Investigation on the characteristics of
444 fruiting and seed development in Chinese jujube (*Ziziphus jujuba* Mill.). International Society
445 for Horticultural Science (ISHS). Leuven, Belgium. pp. 209-214
446 Lu W, Tang X, Huo Y, Xu R, Qi S, Huang J, Zheng C, Wu CA (2012) Identification and
447 characterization of fructose 1,6-bisphosphate aldolase genes in Arabidopsis reveal a gene
448 family with diverse responses to abiotic stresses. Gene 503:65-74
449 Mann M, Kulak NA,Nagaraj N,Cox J (2013) The coming age of complete, accurate, and
450 ubiquitous proteomes. Molecular cell 49:583-590
451 Matsumoto D, Tao R. 2016. Distinct Self-recognition in the *Prunus* S-RNase-based Gametophytic
452 Self-incompatibility System. The Horticulture Journal 85:289-305.
453 Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying
454 mammalian transcriptomes by RNA-Seq.Nature methods 5: 621-628
455 Nagaraj N, Wisniewski JR, Geiger T, Cox J, Kircher M, Kelso J, Pääbo S, Mann M (2011) Deep
456 proteome and transcriptome mapping of a human cancer cell line. Molecular systems biology
457 7:548
458 Niu HL, Zhang HW, Yuan B, Li XG (2015) Flower formation and endogenous hormones dynamic
459 in Chinese jujube. Acta Horticulturae Sinica 42: 655-664
460 Niu YF, Peng JY, Li L (2011) Study on morphology of flower organ and anatomical characteristics
461 of microspore different developmental period in Chinese jujube and wild jujube. Journal of
462 plant genetic resources 12:158-162
463 Righetti PG, Boschetti E (2016) Global proteome analysis in plants by means of peptide libraries
464 and applications.Journal of proteomics 143: 3-14
465 Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression
466 analysis of RNA-seq data. Genome biology 11:R25
467 Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M (2013)
468 Corrigendum: Global quantification of mammalian gene expression control. Nature 495: 126-
469 127
470 Song YQ, Li WQ, Li LL, Liu H, Lu GB(2013) Observation on jujube flowers morphology and
471 anther characteristics. Journal of Shanxi Agricultural Sciences 41:1190-1193
472 Tang X (2013) New functional studies of two cytosolic Fructose Bisphosphate Aldolase genes in
473 *Arabidopsis* [D]. Shandong agriculture university, Taian

Wilhelm M, Schlegl J, Hahne H, Gholami AM, Lieberenz M, Savitski MM, Ziegler E, Butzmann L, Gessulat S, Marx H, Mathieson T, Lemeer S, Schnatbaum K, Reimer U, Wenschuh H, Mollenhauer M, Slotta-Huspenina J, Boese JH, Bantscheff M, Gerstmair A, Faerber F, Kuster B(2014) Mass-spectrometry-based draft of the human proteome.Nature 509: 582-587

Wiśniewski JR, Zougman A, Nagaraj N, Mann M (2009) Universal sample preparation method for proteome analysis. Nature methods 6: 359-362

Wu X, Xiong E, Wang W,Scali M,Cresti M (2014) Universal sample preparation method integrating trichloroacetic acid/acetone precipitation with phenol extraction for crop proteomic analysis. Nature protocols 9:362-374

Zeng Y, Pan Z, Ding Y, Zhu A, Cao H, Xu Q, Deng X (2011) A proteomic analysis of the chromoplasts isolated from sweet orange fruits (*Citrus sinensis* (L.) Osbeck), Journal of experimental botany 62:5297-309

Zhang C, Huang J, Li X (2015) Identification of appropriate reference genes for RT-qPCR analysis in Ziziphus jujuba Mill. Scientia Horticulturae 197: 166-169

Zhu W, Hu J, Wang X, Tian J, Komatsu S (2015) Organ-specific analysis of mahonia using gel-free/label-free proteomic technique. Journal of Proteome Research 14:2669-2685

Figure legends

Fig. 1 Venn diagram of the number of expressed genes with RPKM $\geqslant$1 and the identified proteins in the proteome of Jujube flowers.

510 **Fig. 2** Comparison of proteomics and RNA-Seq data generated from Jujube flowers. (A)
511 Distribution of RPKM values in different levels. (B) Distribution of RPKM values of
512 identified and unidentified genes in proteomic. (C) Distribution of protein abundance
513 (iBAQ intensities) (FDR = 1%). (D) A density scatter plot of iBAQ versus RPKM values.
514 The colour indicates the percentage points that were included in a region.
515 **Fig.** 3 Lists of ten highly ranked GO categories of proteome and their corresponding transcriptome.
516 (A) biological progress, (B) cellular components, (C) molecular function.
517 **Fig.** 4 Cumulative protein mass from the highest to the lowest abundance proteins identified in
518 Jujube flowers.
519 **Fig. 5** Phylogenetic analysis of the S locus-linked F-box genes from different taxa exhibiting S-
520 RNase-based GSI and seven F-box genes of Ziziphus jujuba located in the same
521 chromosome as the S-RNase gene. The neighbour-joining tree was constructed based on
522 aligned amino acid sequences corresponding to those S locus-linked F-box genes from well
523 characterized S haplotypes in each plant species using Mega 6.0 (bootstrap=1,000). *Zj.*
524 *jz039541036* was the pollen-specificity candidate SFB gene of Jujube.
525 **Fig. 6** (A) Transcriptomic and (B) RT-PCR analysis of candidate S-locus F-box genes expression
526 in different organs: fruit 1: young fruit, fruit 2: red fruit, flower: opening flower. The
527 number of each photo was corresponding to the F-box genes listed in Table S2.

Figure 1(on next page)

Venn diagram of the number of expressed genes with RPKM ≥1 and the identified proteins in the proteome of Jujube flowers.

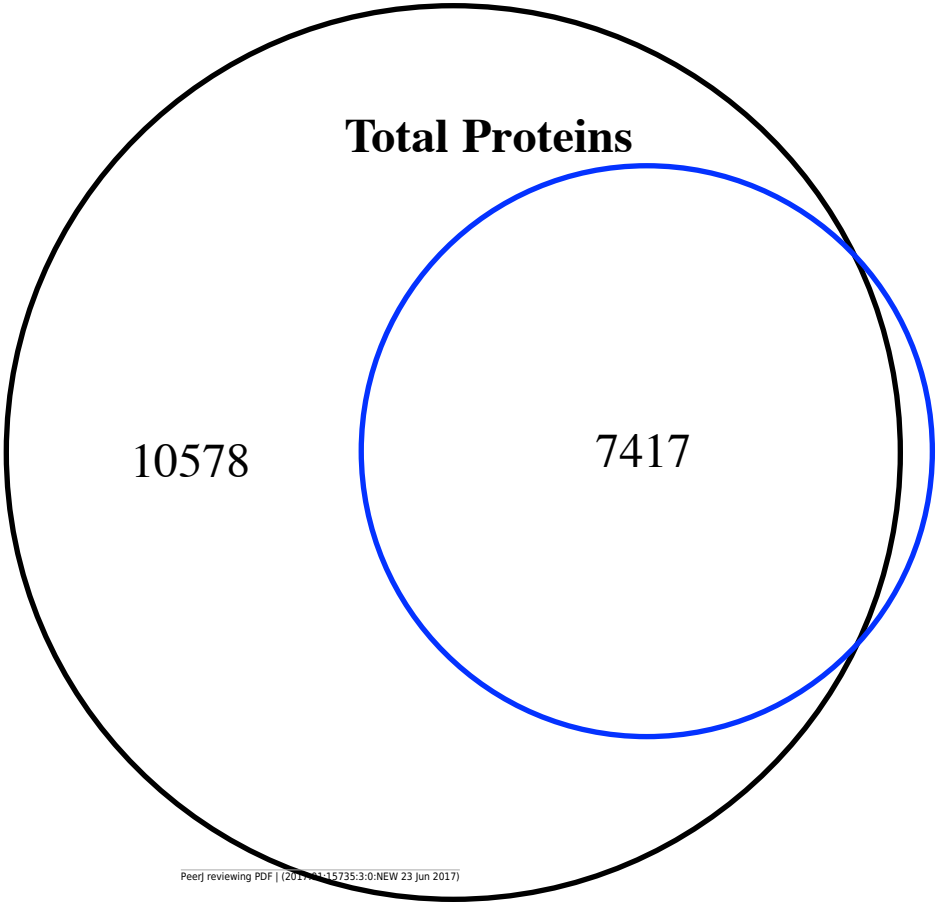

Total mRNAs
Total Proteins
10578
7417

Figure 2

Comparison of proteomics and RNA-Seq data generated from Jujube flowers.

(A) Distribution of RPKM values in different levels. (B) Distribution of RPKM values of identified and unidentified genes in proteomic. (C) Distribution of proteins abundance (iBAQ intensities,FDR = 1%). (D) A density scatter plot of iBAQ versus RPKM values. The colour indicates the percentage points that were included in a region representing a specific colour.

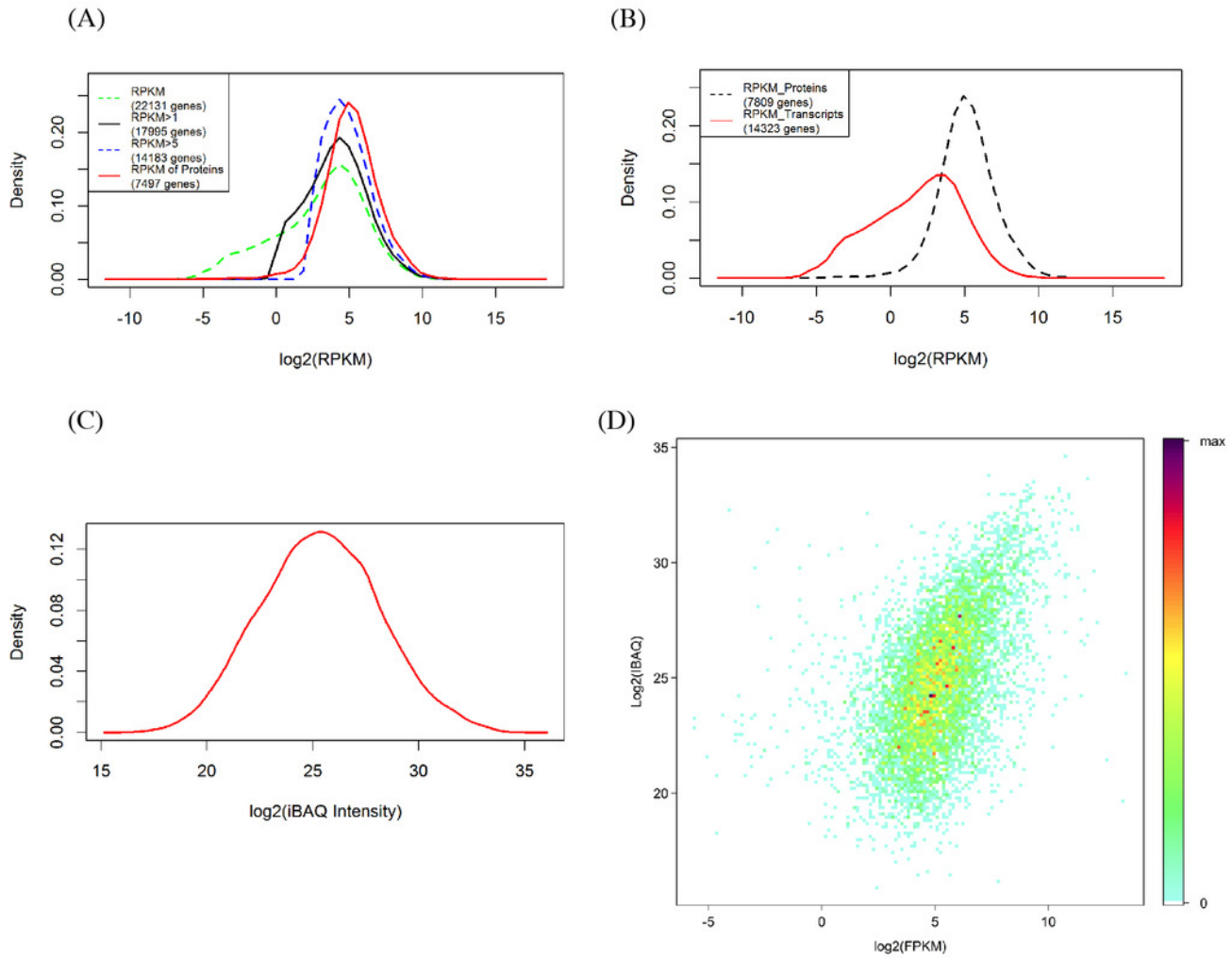
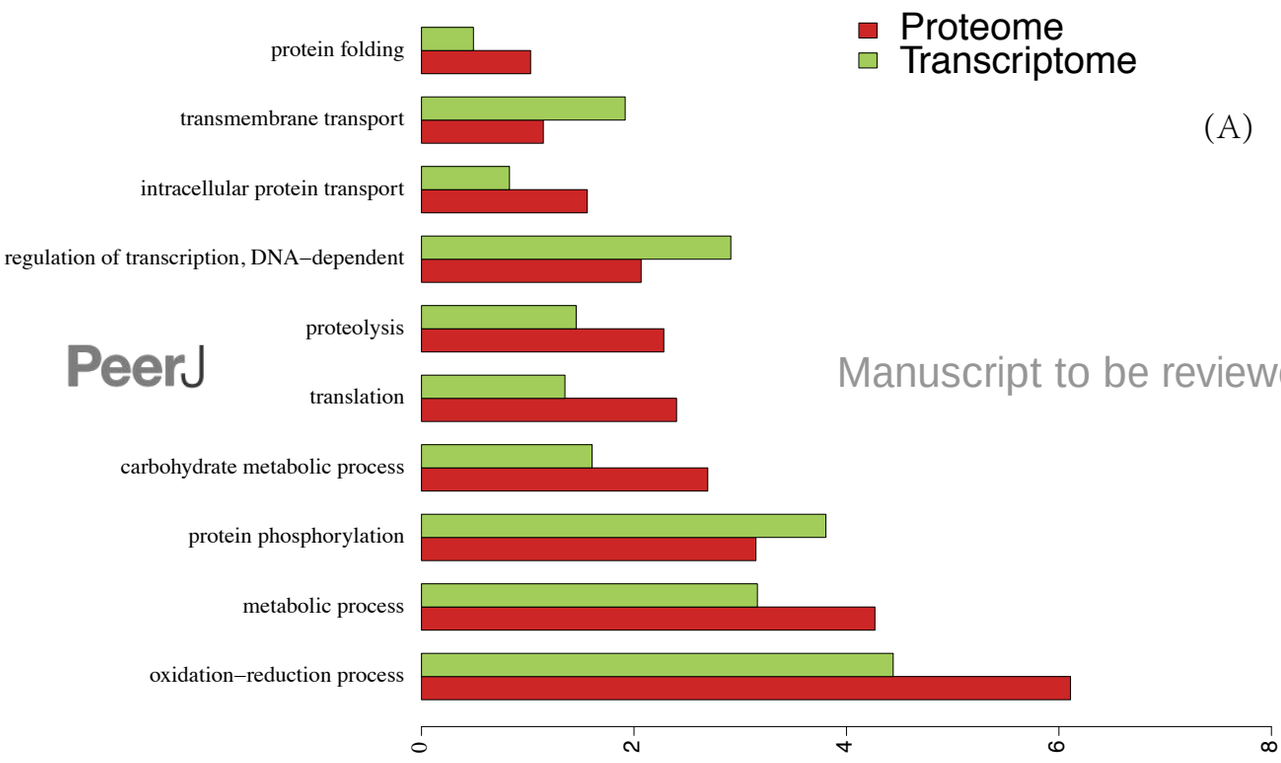


Figure 3(on next page)

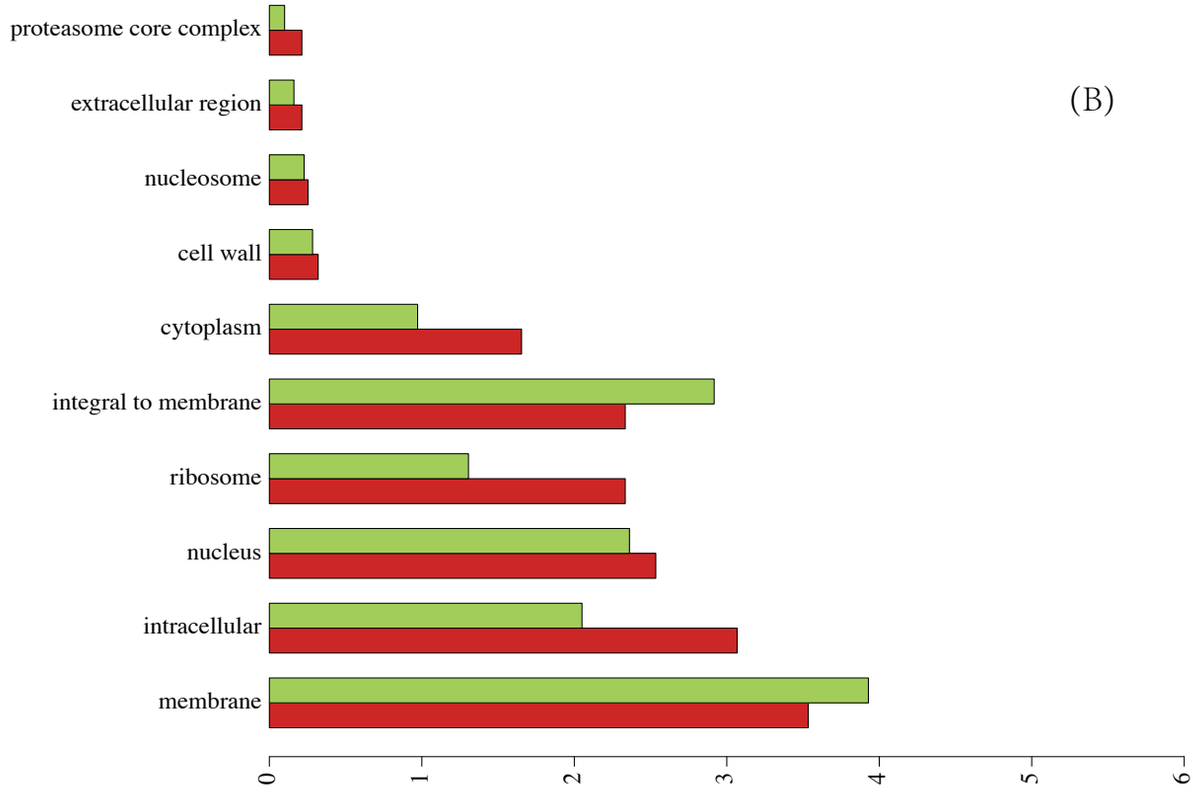
Lists of ten highly ranked GO categories of proteome and their corresponding transcriptome.

(A) biological progress, (B) cellular components, (C) molecular function.



PeerJ

Manuscript to be reviewed



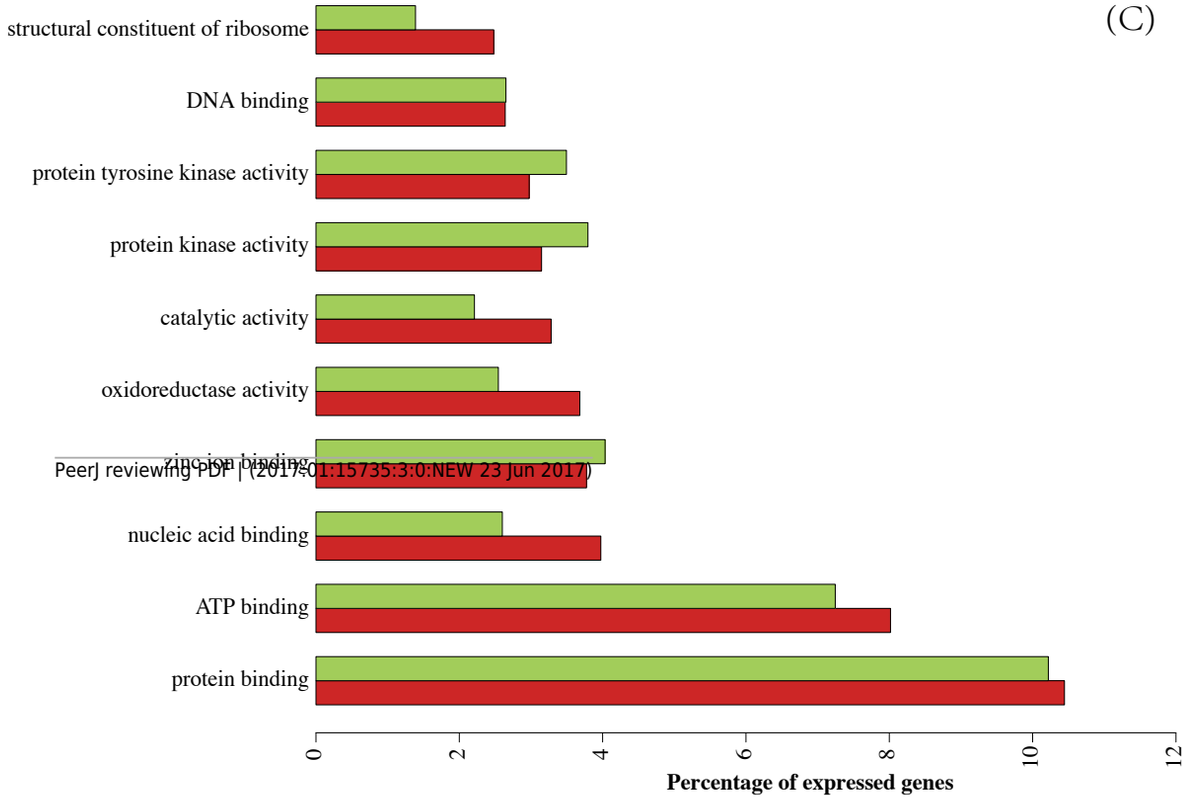


Figure 4

Cumulative protein mass from the highest to the lowest abundance proteins identified in jujube flowers.

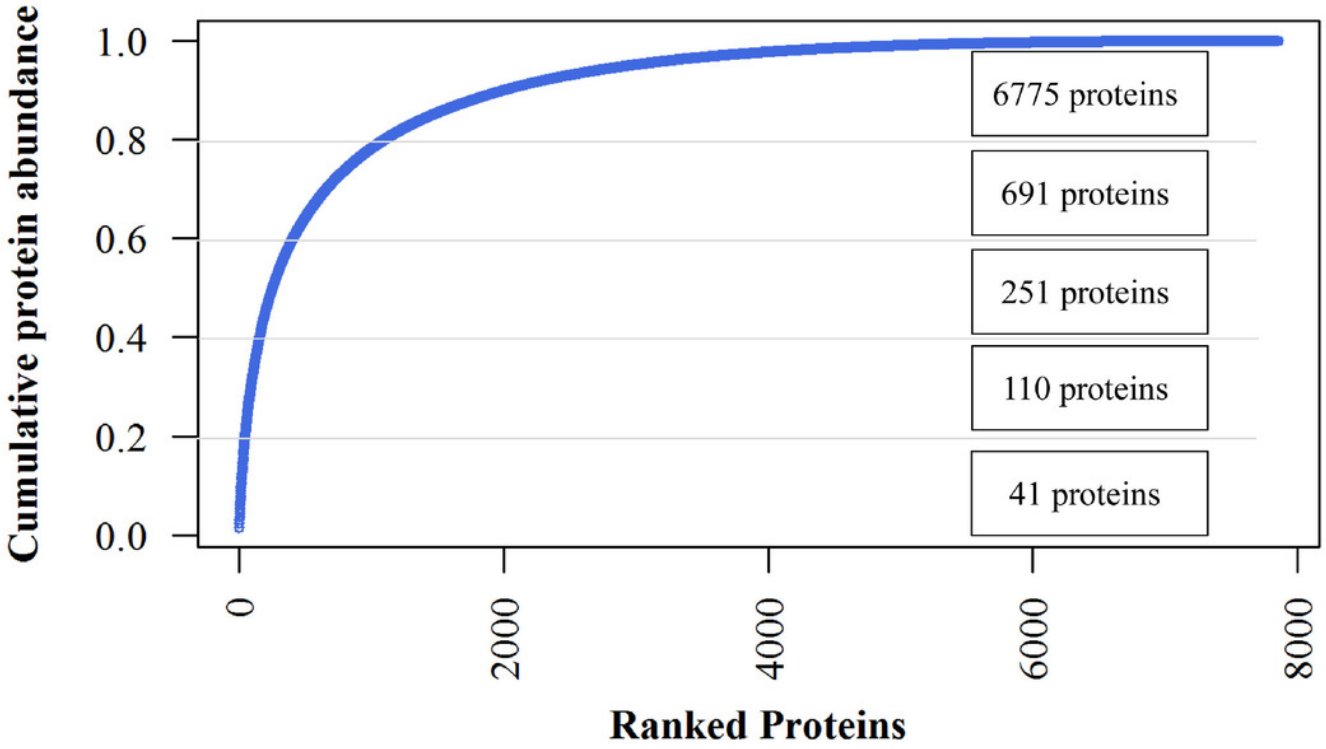


Figure 5(on next page)

Phylogenetic analysis of the S locus-linked F-box genes from different taxa exhibiting S-RNase-based GSI and seven F-box genes of Ziziphus jujuba located in the same chromosome with the S-RNase gene.

The neighbor-joining tree was constructed based on aligned amino acid sequences corresponding to those S locus-linked F-box genes from well characterized S haplotypes in each plant species using Mega 6.0 (bootstrap=1,000). Zj. jz039541036 was the pollen-specificity candidate SFB gene of Jujube.

PeerJ Manuscript to be reviewed

CAD56661 A.hispanicum SLF-S4
CAC33011 A.hispanicum SLF-S2L
CAC33010 A.hispanicum SLF-S2
CAD56660 A.hispanicum SLF-S4D
CAD56852 A.hispanicum SLF-S4A
ABE77147 A.hispanicum SLF-S2C
Plantaginaceae SLF

AIK66495 P.inflata SLF8-S3
ABX82525 P.inflata SLF8-S2
AIK66461 P.inflata SLF13-S2
AHF49538 P.inflata SLFx-S2
AIK66462 P.inflata SLF13-S3
AIK66496 P.inflata SLF3-S3
ABX82524 P.inflata SLF3-S2
AIK66452 P.inflata SLF11-S3
AIK66451 P.inflata SLF11-S2
AIK66457 P.inflata SLF12-S3
AIK66456 P.inflata SLF12-S2
AIK66500 P.inflata SLF4-S3
AHA84093 P.inflata SLF4-S2
AHA84097 P.inflata SLF6-S3
AHA84095 P.inflata SLF6-S2
AHA84096 P.inflata SLF5-S3
AHA84094 P.inflata SLF5-S2
AAR15916 P.inflata SLF10-S3
AAR15915 P.inflata SLF10-S2
AAR15913 P.inflata SLF9-S3
AAR15912 P.inflata SLF9-S2
AIK66498 P.inflata SLF2-S3
AIK66497 P.inflata SLF2-S2
AAU29055 P.inflata SLF1-S2
AAS79486 P.inflata SLF1-S3
AIK66499 P.inflata SLF7-S3
ABX82526 P.inflata SLF7-S2
AIK66472 P.inflata SLF15-S3
AIK66471 P.inflata SLF15-S2
AIK66467 P.inflata SLF14-S3
AIK66466 P.inflata SLF14-S2
AIK66482 P.inflata SLF17-S3
AIK66481 P.inflata SLF17-S2
AIK66477 P.inflata SLF16-S3
AIK66476 P.inflata SLF16-S2
Solanaceae SLF

BAJ52234 P.pyrifolia SFBB-S2
BAJ11967 M.domestica SFBB9-S9
BAJ11954 M.domestica SFBB10-S9
BAJ11961 M.domestica SFBB-S9
BAJ11959 M.domestica SFBB-S9
BAJ52235 P.pyrifolia SFBB2-S2+
BAJ52236 P.pyrifolia SFBB1-S2
BAJ11964 M.domestica SFBB1-S9
BAJ11958 M.domestica SFBB8-S9
BAF79627 P.pyrifolia SFBB8-S2
AB699122 M.domestica SFBB8-S9
BAJ11963 M.domestica SFBB6-S9
BAJ52231 P.pyrifolia SFBB7-S2
BAJ52232 P.pyrifolia SFBB5-S2
BAF47174 M.domestica SFBB3-S9
BAJ11960 M.domestica SFBB11-S9
BAJ52230 P.pyrifolia SFBB4-S2
AB699121 M.domestica SFBB4-S9
BAJ52229 P.pyrifolia SFBB2-S2
AB699120 M.domestica SFBB2-S9
BAJ52228 P.pyrifolia SFBB3-S2
BAJ11965 M.domestica SFBB3-S9
Maleae SFBB

PeerJ reviewing PDF | (2017:01:15735:3:0:NEW 23 Jun 2017)

AB092626 P.mume SLFL2-S7
AB092625 P.mume SLFL2-S1
AB280954 P.mume SLFL2-S4
AB280955 P.avium SLFL3-S4
AB092627 P.avium SLFL3-S7
AB280953 P.avium SLFL1-S4
AB092624 P.mume SLFL1-S7
AB092623 P.mume SLFL1-S1
Prunus SLFL

Zj.jz039969034
Zj.jz026591005
Zj.jz027555002
Zj.jz035833029
Zj.jz027555004
Zj.jz027555001
Zj.jz039541036 *
Z. jujuba SFB

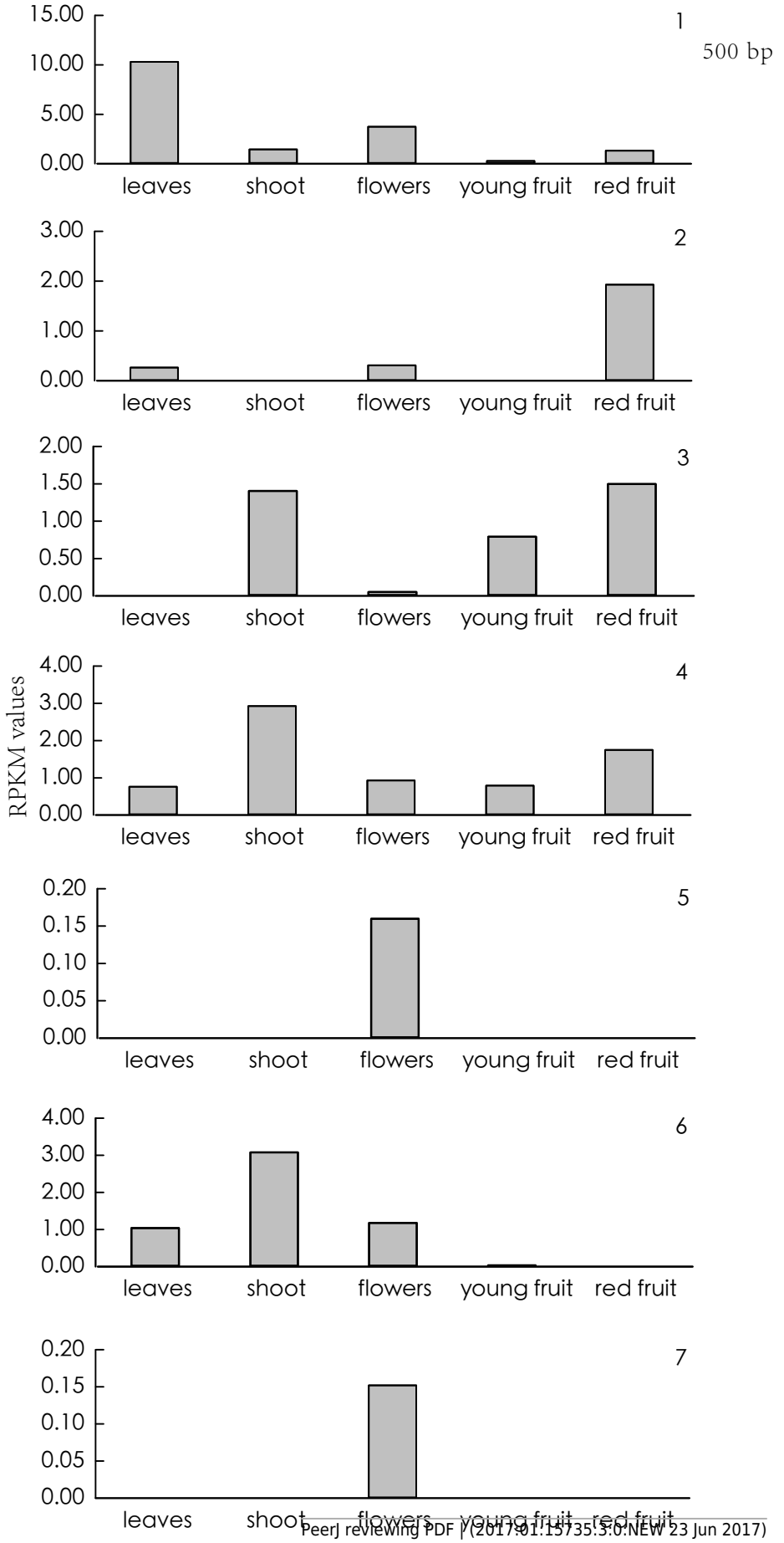
AB111521 P.avium SFB-S4
AB101440 P.mume SFB-S1
AB101441 P.mume SFB-S7
Prunus SFB

Figure 6(on next page)

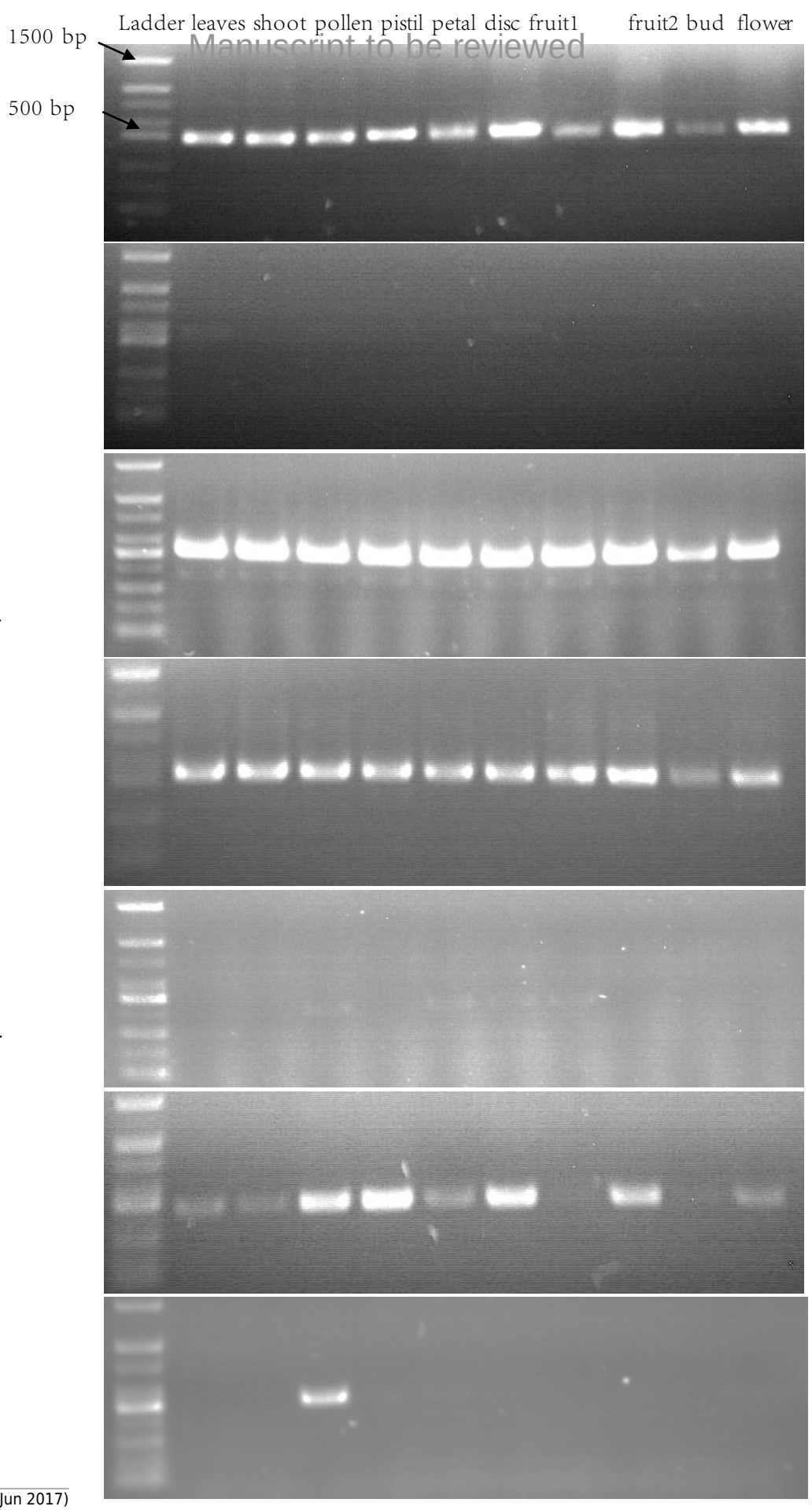
(A) Transcriptomic and (B) RT-PCR analysis of candidate S-locus F-box genes expression in different organs. fruit 1: young fruit, fruit 2: red fruit, flower: opening flower.

The number of each photo was corresponding to the F-box genes listed in Table S2.

PeerJ



(A)



(B)

Manuscript to be reviewed