Transcriptome analysis of *Chelidonium majus* elaiosomes and seeds provide insights into fatty acid biosynthesis

Jiayue Wu [1], **Linlin Peng** [1], **Shubin Dong** [2], **Xiaofei Xia** [3], **Liangcheng Zhao** [Corresp. 1, 4]

[1] College of Nature Conservation, Beijing Forestry University, Beijing, China

[2] College of Biological Sciences and Technology, Beijing Forestry University, Beijing, China

[3] Beijing Museum of Natural History, Beijing, China

[4] Museum of Beijing Forestry University, Beijing Forestry University, Beijing, China

Corresponding Author: Liangcheng Zhao
Email address: lczhao@bjfu.edu.cn

Background. Elaiosomes are specialized fleshy and edible seed appendages dispersed by ants. Lipids are the primary components of elaiosomes. *Chelidonium majus* is a well-known plant, the seeds of which are dispersed by ants. Previous studies have identified the presence of primary fatty acids in its elaiosomes and seeds. However, the molecular mechanisms underlying fatty acid biosynthesis in elaiosomes remain unknown.

Methods. In order to gain a comprehensive transcriptional profile of the elaiosomes and seeds of *C. majus*, and understand the expression patterns of genes associated with fatty acid biosynthesis, four different developmental stages, including the flower-bud (Ch01), flowering (Ch02), young seed (Ch03), and mature seed (Ch04) stages, were chosen to perform whole-transcriptome profiling through the RNA-seq technology (Illumina NGS sequencing).

Results. A total of 63,064 unigenes were generated from 12 libraries. Of these, 7,323, 258, and 11,540 unigenes were annotated with 25 clusters of orthologous groups (COGs), 43 gene ontology (GO) terms, and 373 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, respectively. In addition, 322 genes were involved in lipid transport and metabolism, and 508 genes were involved in the lipid metabolism pathways. Forty-one significantly differentially expressed genes (DEGs) involved in the lipid metabolism pathways were identified, most of which were upregulated in Ch03 compared to Ch02, indicating that fatty acid biosynthesis primarily occurs during the flowering to the young seed stages. Of the DEGs, acyl-ACP thioesterases, acyl carrier protein desaturase (*DESA1*), and malonyl CoA-ACP transacylase were involved in palmitic acid synthesis; stearoyl-CoA desaturase and *DESA1* were involved in oleic acid synthesis, and acyl-lipid omega-6 desaturase was involved in linoleic acid synthesis.

1 # Transcriptome analysis of *Chelidonium majus*
2 # elaiosomes and seeds provide insights into fatty acid
3 # biosynthesis

5 Jiayue Wu[1], Linlin Peng[1], Shubin Dong[2], Xiaofei Xia[3], Liangcheng Zhao[1,4*]

7 [1] College of Nature Conservation, Beijing Forestry University, Beijing, China

8 [2] College of Biological Sciences and Technology, Beijing Forestry University, Beijing, China

9 [3] Beijing Museum of Natural History, Beijing, China

10 [4] Museum of Beijing Forestry University, Beijing, China

13 Corresponding Author:

14 Liangcheng Zhao

15 No. 35 East Qinghua Road, Haidian District, Beijing, 100083, China

16 Email address: lczhao@bjfu.edu.cn

19 # ABSTRACT

20 **Background.** Elaiosomes are specialized fleshy and edible seed appendages dispersed by ants.

21 Lipids are the primary components of elaiosomes. *Chelidonium majus* is a well-known plant, the

22 seeds of which are dispersed by ants. Previous studies have identified the presence of primary

23 fatty acids in its elaiosomes and seeds. However, the molecular mechanisms underlying fatty

24 acid biosynthesis in elaiosomes remain unknown.

25 **Methods.** In order to gain a comprehensive transcriptional profile of the elaiosomes and seeds of

26 *C. majus*, and understand the expression patterns of genes associated with fatty acid biosynthesis,

27   four different developmental stages, including the flower-bud (Ch01), flowering (Ch02), young

28   seed (Ch03), and mature seed (Ch04) stages, were chosen to perform whole-transcriptome

29   profiling through the RNA-seq technology (Illumina NGS sequencing).

30   **Results.** A total of 63,064 unigenes were generated from 12 libraries. Of these, 7,323, 258, and

31   11,540 unigenes were annotated with 25 clusters of orthologous groups (COGs), 43 gene

32   ontology (GO) terms, and 373 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways,

33   respectively. In addition, 322 genes were involved in lipid transport and metabolism, and 508

34   genes were involved in the lipid metabolism pathways. Forty-one significantly differentially

35   expressed genes (DEGs) involved in the lipid metabolism pathways were identified, most of

36   which were upregulated in Ch03 compared to Ch02, indicating that fatty acid biosynthesis

37   primarily occurs during the flowering to the young seed stages. Of the DEGs, acyl-ACP

38   thioesterases, acyl carrier protein desaturase (*DESA1*), and malonyl CoA-ACP transacylase were

39   involved in palmitic acid synthesis; stearoyl-CoA desaturase and *DESA1* were involved in oleic

40   acid synthesis, and acyl-lipid omega-6 desaturase was involved in linoleic acid synthesis.

42   **INTRODUCTION**

43   Elaiosomes are highly specialized nutrient-rich structures that are attached to the seeds of many

44   plant species. The term elaiosome was introduced for the first time by Sernander (1906) to refer

45   to all fleshy and edible appendages of the diaspore dispersed by ants. Seed dispersal by ants is

46   known as myrmecochory, a worldwide phenomenon found in a diverse range of habitats,

47   including arid, tropical, and temperate regions (Van der Pijl, 1982; Werker, 1997; Giladi, 2006).

48   A recent study estimated that more than 11,000 species in 334 genera (about 2.5% of all genera)

49   and 77 families (around 17% of all families) of angiosperm plants could be dispersed via

50   myrmecochory (Lengyel et al., 2010).

51   Myrmecochory is regarded as a mutualistic relationship that is beneficial for both plants and

52   ants. Ants typically collect and transfer whole diaspores (seed + elaiosome) to their nests,

53   wherein the elaiosomes are consumed and the seeds are discarded unharmed within or outside

54   the nests, which are in turn generally regarded to provide optimal germinating conditions

55   (Beattie, 1985; Hughes & Westoby, 1990; Giladi, 2006). The elaiosomes can be fed to the larvae,

56 which in turn promotes ant colony reproduction, fitness, and gyne production (Morales &

57 Heithaus, 1998; Bono & Heithaus, 2002; Fischer et al., 2005; Gammans, Bullock & Schönrogge,

58 2005). Therefore, elaiosomes are key structures that increase the attractiveness of diaspores to

59 ants (Reifenrath, Becker & Poethke, 2012). Elaiosomes develop from seed tissues (chalaza,

60 funiculus, hilum, and raphe-antiraphe) through various mechanisms, but appear to have the same

61 main function of attracting ants (Lisci, Bianchini & Pacini, 1996; Morrone, Vega & Maier, 2000;

62 Ciccarelli et al., 2005; Yang et al., 2015). Hence, elaiosomes are considered suitable examples

63 for studying convergent evolution in flowering plants (Lengyel et al., 2010).

64   Many studies have described several factors that likely determine the efficiency of

65 myrmecochory. The sizes of seeds and elaiosomes, the elaiosome size/seed size ratio, and their

66 physical surfaces influence the ability of ants to disperse the diaspores (Oostermeijer, 1989;

67 Hughes & Westoby, 1992; Mark & Olesen, 1996). In addition to the morphological and

68 structural characteristics of diaspores, chemical compounds present in the elaiosomes are

69 additionally regarded as crucial factors that influence the removal of diaspores, considering that

70 elaiosomes are beneficial for diaspore transport and food acquisition in ants (Hughes, Westoby &

71 Jurado, 1994; Fischer et al., 2008; Reifenrath, Becker & Poethke, 2012; Chen et al., 2016).

72 Although protein and starch are also present in elaiosomes, lipids generally constitute the major

73 class of compounds in elaiosomes. Based on the prominent hypothesis that certain fatty acids

74 induce the collection of diaspores by ants, most studies have focused on the fatty acids of the

75 elaisomes and compared their compositions with those of seeds or even insects (Marshall,

76 Beattie & Bollenbacher, 1979; Soukup & Holman, 1987; Kusmenoglu, Rockwood &

77 Gretz,1989; Lanza, Schmitt & Awad, 1992; Hughes, Westoby & Jurado, 1994; Morrone, Vega &

78 Maier, 2000; Boulay, Coll-Toledano & Cerdá, 2006). A more recent chemical analysis of the

79 diaspores of 13 central European ant-dispersed plant species from seven families revealed that

80 elaiosomes have a more homogeneous fatty acid composition, with oleic acid as the most

81 abundant fatty acid component, followed by palmitic acid and linoleic acid (Fischer et al., 2008).

82   *Chelidonium majus* L., commonly known as greater celandine, is the only species of the

83 genus *Chelidonium* in the Papaveraceae family. *C. majus* is an herbaceous perennial plant (Fig.

84 1A) that is widely distributed in Europe and Asia. *C. majus* is rich in various types of

85 isoquinoline alkaloids and is widely used against various diseases, such as ulcer, gastric cancer,

86   oral infection, liver disorders, chronic bronchitis, asthma, and general pains, in European

87   countries and Chinese traditional medicine (Arora & Sharma, 2013; Maji & Banerji, 2015). *C.*

88   *majus* has attracted research attention not only for its medicinal value, but also for its well-

89   known role in ant-dispersal (Fig. 1I). As a typical myrmecochorus plant, *C. majus* has been

90   investigated in several studies on ant-seed interactions (Gorb & Gorb, 2000; Fischer et al., 2008;

91   Servigne & Detrain, 2008; Reifenrath, Becker & Poethke, 2012). In terms of its reproductive

92   features, *C. majus* is characterized by large yellow petals, long cylindrical capsules, and

93   numerous small seeds (ca. 1.1-1.5 mm in length) with conspicuous coronal elaiosomes (Figs. 1C-

94   1F). The elaiosomes (ca. 0.8-1.1 mm in length) are white and transparent, which contrasts with

95   the glossy black color of the seeds (Figs. 1F,1G).

96   The structure and development of elaiosomes in *C. majus* were briefly described by Lisci et

97   al. (1996), who showed that the elaiosomes of *C. majus* arise from the cells of the raphe and

98   form a ventral excrescence along the seed from the hilum and adjacent part of the micropyle to

99   the chalazal area. Recently, one of the authors of this paper (Liangcheng Zhao) conducted a

100   detailed study on the complete ontogeny of *C. majus* elaiosomes and seeds using standard

101   techniques for plant anatomy and histochemistry (see Yang et al., 2015). Different development

102   stages have been characterized based on distinct morphological and anatomical features that can

103   be used in genetic studies. Results indicated that the meristem of *C. majus* elaiosome is formed

104   by periclinal divisions of the outmost layer cells of raphe in flower-bud stage (prior to

105   fertilization); after flowering stage (post-pollination) mitotic activity ceases, and the elaiosome

106   develops solely through very active cell growth. The mature elaiosome of *C. majus* is composed

107   of two cell types, small basal cells and very elongated outer cells. Both cell types, especially the

108   outer cells, contain a large amount of lipids (Fig. 1H), as verified by the red-stained oil droplets

109   with Sudan III. In contrast, the seed primarily consists of the endosperm, which contains

110   numerous protein granules (Fig.1H). Reifenrath, Becker & Poethke (2012) performed

111   quantitative analysis of the elaiosomes by gas chromatography and provided a detailed fatty acid

112   composition of the *C. majus* elaiosomes. The most abundant fatty acid was determined to be

113   linoleic acid, followed by palmitic acid, and then oleic acid; these three fatty acids comprised of

114   nearly 96.5% of the total fatty acid content of the elaiosomes.

115   Although various studies have investigated the lipid composition of elaiosomes and seeds in

116   different plant species and identified the main fatty acids, the regulatory mechanisms related to

117   the biosynthesis of these fatty acids remain unknown. To understand the expression patterns of

118   genes associated with fatty acid biosynthesis in elaisomes and seeds, detailed genome

119   information is essential and indispensable. In the present study, RNA-seq technology was

120   employed to obtain a comprehensive overview of gene expression related to fatty acid

121   biosynthesis in *C. majus* elaiosomes and seeds for the first time. In recent years, next-generation

122   sequencing (NGS), also known as high-throughput sequencing, has emerged as a powerful tool

123   for profiling the whole transcriptome from non-model plants that do not have available reference

124   genomes (Unamba, Nag & Sharma, 2015). In the present study, we performed whole-

125   transcriptome profiling by NGS to capture the genomic landscape of *C. majus* elaiosomes and

126   seeds. The analysis not only provided important insights into fatty acid biosynthesis in *C. majus*

127   elaiosomes and seeds, but also served as a genomic resource for future evolutionary studies that

128   focus on the convergent evolution of myrmecochorous plants.

129

MATERIALS & METHODS

131   **Plant Materials and Treatments**

132   The sampled plants of *Chelidonium* majus were grown in Beijing Badaling National Forest Park,

133   China, under natural conditions (Fig. 1A). To gain a comprehensive transcriptional profile of its

134   elaiosomes and seeds, sample collection at the optimal developmental stage is crucial. Based on

135   previous observations of *C. majus* elaiosome and seed development, four different

136   developmental stages were defined(Yang et al., 2015). In the present study, we chose these four

137   developmental stages, namely flower-bud stage (pre-fertilization) (Ch01) (Fig. 1B), flowering

138   stage (post-pollination) (Ch02) (Fig. 1C), young seed stage (Ch03) (Fig. 1E), and mature seed

139   stage (Ch04) (Fig. 1F) for comparative transcriptome analysis to explore gene expression during

140   fatty acid biosynthesis in the elaisomes and seeds of *C. majus*. Corresponding to the four stages,

141   flower buds (with bracts, petals, and stamens removed), fertilized ovaries, young seeds (testa

142   white to brown), and mature seeds (testa black and hard) were collected as RNA-seq materials.

143   They were frozen in liquid nitrogen and stored at -80 °C until further use. Three samples were

144   collected as biological repeats for each developmental stage.

145   **RNA Isolation and Sequencing**

146   To obtain transcriptome expression profiles, RNA-seq transcriptome libraries were prepared

147   using the TruSeqTM RNA sample preparation Kit from Illumina (San Diego, CA, USA) and 5

148   μg of total RNA. First, messenger RNA was isolated through the polyA selection method using

149   oligo (dT) beads and fragmented in a fragmentation buffer. Next, double-stranded cDNA was

150   synthesized using a SuperScript double-stranded cDNA synthesis kit (Invitrogen, CA, USA) and

151   random hexamer primers (Illumina). Then, the synthesized cDNA was subjected to end-repair,

152   phosphorylation, and "A" base addition according to Illumina's library construction protocol.

153   Libraries were size selected for cDNA target fragments of 200-300 bp on a 2% Low Range Ultra

154   Agarose gel followed by PCR (polymerase chain reaction) amplification using Phusion DNA

155   polymerase (NEB) for 15 cycles. After quantification using TBS380, the paired-end RNA-seq

156   sequencing library was sequenced using Illumina HiSeq 4000 (2 × 150 bp read length). RNA

157   samples were named Ch01a/b/c (flower-bud stage), Ch02a/b/c (flowering stage), Ch03a/b/c

158   (young seed stage), and Ch04a/b/c (mature seed stage). Twelve cDNA libraries were then

159   constructed and sequenced using the Illumina deep sequencing platform.

160   ***De novo* assembly, Annotation, Differential expression analysis**

161   The raw paired-end reads were trimmed and quality controlled by SeqPrep and Sickle with

162   default parameters. Then, clean data were used to do *de novo* assembly with Trinity (Grabherr et

163   al., 2011). All of the assembled transcripts were searched against the NR, COG, and KEGG

164   databases using BLASTX to identify the proteins that had the highest sequence similarity to the

165   given transcripts to retrieve their function annotations; a typical cut-off E-value of less than

166   1.0×10^{-5} was set. The BLAST2GO program was used to get GO annotations of uniquely

167   assembled transcripts for describing biological processes, molecular functions, and cellular

168   components (Conesa et al., 2005). COG is an early database that is expected to be a useful

169   platform for the functional annotation of newly sequenced genomes (Tatusov et al., 2003). The

170   KEGG database is the primary public repository devoted specifically to pathways. Metabolic

171   pathway analysis was performed using KEGG (Kanehisa & Goto, 2000), and KEGG

172   classifications facilitate the exploration of categories of genes with specific functions. To

173   identify DEGs between different samples, the expression level of each transcript was calculated

174   according to the fragments per kilobase of the exon per million mapped reads (FPKM) method.

175   RSEM (RNA-Seq by Expectation-Maximization) was used to quantify gene and isoform

176   abundances (Li & Dewey, 2011). The R statistical package software, EdgeR (Empirical analysis

177   of Digital Gene Expression in R), was utilized for the differential expression analysis (Robinson,

178   McCarthy & Smyth, 2010). In addition, functional-enrichment analysis, including GO and

179   KEGG, were performed to identify which DEGs were significantly enriched in GO terms and

180   metabolic pathways at a Bonferroni-corrected P-value ≤ 0.05 compared with the whole-

181   transcriptome background. GO functional enrichment and KEGG pathway analysis were carried

182   out by Goatools and KOBAS (Xie et al., 2011).

183

184

185   **Quantitative Real-time PCR (qRT-PCR) analysis**

186   For each sample, 1 μg of total RNA was used for cDNA synthesis using the Plant RNA

187   Purification Reagent (Invitrogen). ChamQ SYBR Color qPCR Master Mix (2X) (Vazyme

188   Biotechnology, Nanjing, China) was used for qRT-PCR. qRT-PCR was performed on Line gene

189   9600 plus Fluorescence quantitative PCR instrument (Hangzhou, China). Each sample contains

190   three independent biological replicates.

191

192   **RESULTS AND DISCUSSION**

193   **Transcriptome sequencing and *de novo* assembly**

194   In this study, a total of 12 libraries were sequenced. After removing the low-quality reads and

195   adaptor sequences, the remaining clean data comprised of approximately 61 Gb of sequence data

196   (171 million reads from the flower-bud stage (Ch01); 141 million reads from flowering stage

197   (Ch02); 148 million reads from young seed stage (Ch03); and 154 million reads from mature

198   seed stage (Ch04)). The results are shown in Table 1. *De novo* assembly of the clean reads was

199   performed using Trinity to generate a reliable integrated reference sequence database (File S1).

200   Some reads comprise of a contig, and some contigs comprise of a unigene. We obtained a total

201   of 63,064 unigenes (File S2) for all studied samples; these unigenes had an average length and

202   N50 value of 1114.56 bp and 1565 bp, respectively. After sorting the assembled unigenes from

203   large to small, the N50 value was calculated as half of the total length of the unigenes. The size

204   distribution of the assembled transcripts is shown in Figure 2. As shown in Table 1 and Figure 2,

205   the sequencing data were of sufficiently high quality for subsequent analyses. Therefore, gene

206   expression data from *C. majus* elaisomes and seeds were reliable.

207   **Functional annotation and classification of unigenes**

208   To verify the potential functions of the 63,064 assembled unigenes, we performed sequence

209   similarity BLASTx searches against the NR (NCBI protein nonredundant), COG (Clusters of

210   Orthologous Groups), GO (Gene Ontology), and KEGG (Kyoto Encyclopedia of Genes and

211   Genomes) protein databases with E-values $< 1e^{-5}$. After alignment, gene annotation, functional

212   classification, and metabolic pathway assignment, a total of 30,143 (47.80%) of the unigenes

213   were successfully annotated (File S3), which indicated that their functions were relatively

214   conserved and diverse.

215   **Characterization of unigene sequencing**

216   Among all the 30,143 annotated unigenes, 29,352 (97.38%) were mapped to the NR database. In

217   addition, the distribution of E-values of the hits in the NR database indicated that 20,378

218   unigenes, representing more than half of the mapped sequences (69.43%), showed significant

219   homology (E-value $< 1.0 \times 10^{-30}$), while 8,973 unigenes (about 30.57%) had E-values ranging

220   from 1.0×10^{-30} to 1.0×10^{-5} (Fig. 3, File S3). Smaller E-values distribution (Fig. 3A) indicated a

221   higher reliability of hits. The similarity distribution showed that the majority of the mapped

222   26,160 unigenes (89.13%) had similarity values > 60%, whereas only 3,192 (10.87%) had

223   similarity values < 40% (Fig. 3B). To identify species specificity, individual unigenes were

224   annotated based on the highest BLAST scores against the NR database. Among flowering plants,

225   6,666 (22.71%) unigenes showed significant homology to *Vitis vinifera*, followed by *Theobroma*

226   *cacao* (1,983 unigenes, 6.76%), *Phoenix dactylifera* (1,230 unigenes, 4.19%), *Populus*

227   *trichocarpa* (1,224 unigenes, 4.17%), *Prunus persica* (1,090 unigenes, 3.71%), and *Jatropha*

228   *curcas* (1,065 unigenes, 3.63%) (Fig. 3C). A total of 2,980 unigenes, representing around

229   10.15% of all unigenes, did not match the NR database, and thus provided novel genomic

230   information on *C. majus*. Some of these unigenes are likely to be untranslated regions (UTRs) or

231   non-coding RNAs.

232   **Functional classification by Gene Ontology**

233   By conducting GO functional classification, a total of 258 unigenes were classified into one or

234   more terms and divided into 43 functional groups, belonging to three major categories (Fig. 4).

235   In the biological process category, the majority of the assigned proteins were involved in

236   metabolic processes (177 unigenes, 68.60%), followed by genes involved in cellular processes

237   (169 unigenes, 65.50%) and single-organism processes (156 unigenes, 60.47%). Under the

238   molecular function category, the mapped unigenes were categorized into nine groups, out of

239   which genes were predominantly associated with catalytic activity (165 unigenes, 63.95%) and

240   binding (123 unigenes, 47.67%), which exceeded the total number of the other terms. The

241   cellular component category was divided into 14 groups, in which the largest assignments were

242   found to be in the cell and cell part. There were 164 unigenes in the cell and cell part (63.57%),

243   followed by organelles (115 unigenes, 44.57%), membrane (52 unigenes, 20.16%), and

244   membrane parts (36 unigenes, 13.95%). The above results suggested that the development of

245   elaisomes and seeds in *C. majus* are closely associated with enzyme-catalyzed reactions, cell

246   structure formation, and metabolic processes.

247   **Functional classification by Clusters of Orthologous Groups**

248   To predict the potential function of genes and evaluate the completeness of the transcriptome

249   libraries, all assembled unigenes were searched against the COG database. A total of 7,323

250   mapped unique sequences were clustered into 25 functional categories (Fig. 5). Among them, the

251   largest category was 'general functions prediction only' (917, 12.52%), followed by 'signal

252   transduction mechanisms' (798, 10.90%), 'posttranslational modification, protein turnover, and

253   chaperones' (727, 9.93%), and 'translation, ribosomal structure, and biogenesis' (596, 8.14%).

254   Thus, we inferred that these genes play an important role in ribosome formation, protein

255   synthesis, and signal transduction during the development of elaisomes and seeds in *C. majus*.

256   Notably, 322 unigenes (4.40%) were classified under the 'lipid transport and metabolism'

257   category, which could provide useful information for further studies on lipid metabolism in *C.*

258   *majus* elaisomes and seeds.

259   **Biological pathways identification by KEGG**

260   To identify the biological pathways that were actively involved in lipids accumulation in *C.*

261   *majus* elaisomes and seeds, all annotated unigenes were mapped to the reference pathways in the

262   KEGG. The analysis identified a total of 11,540 unigenes that were assigned to five clusters and

263   distributed across 373 KEGG pathways (Fig. 6). Most of the unigenes were grouped under the

264   metabolism cluster (A). Within this cluster, the majority of unigenes were assigned under the

265   "global and overview maps pathway" (2,736, 23.71%), followed by "carbohydrate metabolism"

266   (1022, 8.85%), "amino acid metabolism" (800, 6.93%), and "energy metabolism" (590, 5.11%)

267   categories. The relative abundance of unigenes involved in 'amino acid metabolism' pathways is

268   probably related to the biosynthesis of massive storage protein in the *C. majus* endosperm.

269   Notably, in the "energy metabolism" cluster, some pathways were found to be closely related to

270   the lipid metabolism. A total of 508 unigenes, including glycerophospholipid metabolism (205,

271   40.35%), glycerolipid metabolism (108, 21.26%), fatty acid degradation (100, 19.69%),

272   biosynthesis of unsaturated fatty acids (93, 18.31%), fatty acid elongation (75, 14.76%), fatty

273   acid biosynthesis (74, 14.57%), ether lipid metabolism (68, 13.39%), sphingolipid metabolism

274   (54, 10.63%), linoleic acid metabolism (44, 8.66%), steroid biosynthesis (42, 8.27%),

275   arachidonic acid metabolism (21, 4.13%), steroid hormone biosynthesis (16, 3.15%), synthesis

276   and degradation of ketone bodies (12, 2.36%), and primary bile acid biosynthesis (5, 0.98%)

277   pathways were involved in these pathways (Fig. 7).

278   **Analysis of DEGs at four developmental stages**

279   Analysis of the fatty acid compositions of *C. majus* seeds and elaiosomes showed that the

280   elaiosome-bearing seeds predominantly contained the mono-unsaturated fatty acid, linoleic acid

281   (C18:2) (77.9%), followed by the mono-unsaturated fatty acid, oleic acid (C18:1) (16%)

282   (Reifenrath, Becker & Poethke, 2012). Furthermore, saturated fatty acids accounted for only six

283   percent of the composition, and most of the saturated fatty acids were identified as palmitic acid

284   (C18:1) (3.5%). However, in only elaiosomes, the most abundant fatty acid also was linoleic acid

285   (37.2%), followed by palmitic acid (34.8%), and oleic acid (24.5%). Previous developmental and

286   histochemical observations (Yang et al., 2015) indicated that the elaiosome meristem of *C. majus*

287   was formed during the flower-bud stage (Ch01), while elaiosome cell growth and enlargement

288   primarily occurred from the flowering stage (Ch02) to the young seed stage (Ch03); the

289   maturation of the seed and elaiosome occurred from the young seed stage to the mature seed

290   stage (Ch04). Therefore, to obtain insight into the regulation of the biosynthesis of saturated fatty

291   acids and unsaturated fatty acids in elaiosomes and seeds, we analyzed the differentially

292   expressed genes (DEGs) of the lipid metabolism pathways in Ch02 compared to Ch01, Ch03

293   compared to Ch02, and Ch04 compared to Ch03.

Identification of unigenes related to fatty acid biosynthesis

From the fatty acids biosynthesis pathway (Path: ko00061) and the biosynthesis pathway of

unsaturated fatty acids (Path: ko01040) at the four developmental stages of *C. majus* elaisomes

and seeds (File S4), we identified 41 significantly DEGs. These DEGs are shown in File S5 and

the corresponging genetic heat map is shown in Figure 8. Four unigenes encoding the enzymes

malonyl CoA-ACP transacylase (*fabG*) (Toomey & Wakil, 1966), acyl-CoA oxidase (*ACOX1*)

(Oaxaca-Castillo et al., 2007), acyl carrier protein desaturase (*DESA1*) (Kachroo et al., 2007),

and β-ketoacyl–acyl carrier protein synthase II (*fabF*) (Magnuson, Carey & Cronan, 1995) were

found to be upregulated in Ch02 compared to Ch01. By contrast, 24 unigenes encoding the

enzymes acetyl-coA carboxylase (Accase) (Bilder et al., 2006), acyl coenzyme A synthetase

(*ACSL*) (Black et al., 1992), *DESA1*, acyl-carrier-protein S-malonyltransferase (*fabD*)

(Magnuson et al., 1992), *fabF*, *fabG*, enoyl-[acyl-carrier-protein] reductase (*fabI*) (Bergler et al.,

1996), β-hydroxyacyl-acyl carrier protein dehydratases (*fabZ*) (Heath & Rock, 1996), acyl-lipid

omega-6 desaturase (Delta-12 desaturase) (*FAD2*) (Kainou et al., 2006), acyl-ACP thioesterases

(*FatA/B*) (Salas & Ohlrogge, 2002), and stearoyl-CoA desaturase (Delta-9 desaturase) (*SCD*)

(Stukey, McDonough & Martin, 1990) were clearly upregulated in Ch03 compared to Ch02.

However, in Ch04 compared to Ch03 identified that out of the 24 previously mentioned

unigenes, 13 genes encoding *Accase*, *ACSL*, *fabD*, *fabF*, *fabG*, *fabI*, and *SCD* were

downregulated, while the 11 remaining genes were not differentially expressed. The above

results indicated that there are more relevant DEGs in Ch02 vs Ch03 compared to Ch01 vs Ch02

and Ch03 vs Ch04. Based on the observations of *C. majus* elaiosome development (Yang et al.,

2015), the remarkable cell growth of the elaiosome also occurred during the flowering stage to

the young seed stage (Ch02 to Ch03). Therefore, we inferred that the biosynthesis of saturated

fatty acids and unsaturated fatty acids in *C. majus* elaisomes might primarily occur during this

period.

Identification of unigenes related to the biosynthesis of saturated fatty acid

In the fatty acid biosynthesis metabolic pathway, a comparison of the DEGs in Ch02 compared

to Ch01 (Table 2, File S6A) showed that most genes that are primarily involved in fatty acid

biosynthesis were not significantly expressed during the flowering stage (Ch02), and only a few

genes were upregulated. By contrast, the gene expression analysis in Ch03 compared to Ch02

(Table 2, File S6B) revealed that the majority of genes were significantly expressed during the

325   young seed stage (Ch03). Most of these expressed genes, which encode Accase, *ACSL*, *DESA1*,

326   *fabD*, *fabF*, *fabG*, *fabI*, *fabZ*, and *FatA/B*, were significantly upregulated in Ch03. On the other

327   hand, a gene expression analysis in Ch04 compared to Ch03 (Table 2, File S6C) revealed that

328   almost all genes were significantly downregulated during the mature seed stage (Ch04). The

329   above results indicated that the biosynthesis of saturated fatty acid in *C. majus* elaisomes and

330   seeds also mainly occurred during the flowering stage to the young seed stage. However, some

331   unigenes that encode *FabG*, exhibited inverse expression patterns. The above-mentioned genes

332   were found to be downregulated in Ch03 compared to Ch02 but were upregulated in Ch02

333   compared to Ch01 and in Ch04 compared to Ch03 (Table 2). These genes are potentially

334   involved in fatty acid synthesis through negative feedback signaling.

335   The enzymes *FatA/B* and *DESA1*, which are the key enzymes involved in palmitic acid

336   synthesis (see File S6), are particularly noteworthy, because palmitic acid is the most abundant

337   saturated fatty acid in *C. majus* elaisomes but is present in very low quantities in the

338   corresponding seeds. *FatA/B* was slightly upregulated in Ch02 compared to Ch01 and

339   significantly upregulated in Ch03 compared to Ch02, but was significantly downregulated in

340   Ch04 compared to Ch03. *DESA1* was significantly upregulated in both Ch02 compared to Ch01

341   and Ch03 compared to Ch02, but significantly downregulated in Ch04 compared to Ch03. The

342   above findings indicate that the biosynthesis of palmitic acid in *C. majus* elaiosomes occurs from

343   the flower bud stage to the flowering stage (Ch01 to Ch02) but primarily occurs from the

344   flowering stage to the young seed stage. Hughes et al. (1994) compared the fatty acid

345   compositions of 12 plant species of elaiosomes and seven orders of insects and showed that

346   palmitic acid levels in elaiosomes were particularly similar to those in insects. However, because

347   *C. majus* was not included in Hughes's study, whether the high palmitic acid content in *C. majus*

348   elaiosomes is similar to that in the ants that disperse the *C. majus* seeds remains unknown and

349   needs further study.

350   **Identification of unigenes related to the biosynthesis of unsaturated fatty acid**

351   Figure 8 and Figure 9 show that significantly upregulated genes were identified in Ch03

352   compared to Ch02 and were involved in three processes in the biosynthesis of unsaturated fatty

353   acid metabolic pathway. The first process is the conversion of palmitic acid (C16:0) to stearic

354   acid (C18:0), which involves genes encoding the enzyme, *fabG*. The second process is the

355   conversion of stearic acid (C18:0) to oleic acid (C18:1, ∆9), which involves genes encoding the

356   enzymes, *SCD* (Delta-9 desaturase) and *DESA1*. The third process is the conversion of oleic acid

357   (C18:1, ∆9) to linoleic acid (C18:2, ∆9∆12), which involves genes encoding the enzyme, *FAD2*

358   (Delta-12 desaturase). The above findings showed that some palmitic acid was converted into

359   stearic acid (saturated fatty acid), and the latter was subsequently desaturated to form oleic acid

360   and linoleic acid.

361   Linoleic acid is a poly-unsaturated fatty acid that cannot be synthesized by insects *de novo*

362   and is thus required to be incorporated in their diet (Dadd, 1973). Previous studies have

363   suggested increasing linoleic acid levels by stimulating the diaspore-carrying behavior of the ants

364   (Gammans, Bullock & Schönrogge, 2005). The high linoleic acid content in both elaisomes and

365   seeds of *C. majus* indicated that the whole diaspores (seed + elaiosome) were probably used to

366   attract ants that contain high levels of (essential) nutrients. In addition, oleic acid was shown to

367   promote ant removal behavior (Marshall, Beattie & Bollenbacher, 1979; Skidmore & Heithaus,

368   1988; Brew, O'Dowd & Rae, 1989; Hughes, Westoby & Jurado, 1994). The higher oleic acid

369   content in *C. majus* elaisomes and the lower oleic acid content in their corresponding seeds

370   suggested that oleic acid in *C. majus* might act as the primary stimulating substance to induce the

371   collection of its diaspores by ants, as previously reported in other plant species.

372   **Validation of unigenes by quantitative real-time PCR (qRT-PCR)**

373   To validate the reliability of the RNA-seq results, four unigenes encoding *ACSL*, *DESA1*, *fabG*,

374   and *FatB* were randomly selected for qRT-PCR validation at the four different developmental

375   stages. These four genes were all related to the fatty acid biosynthesis metabolic pathway and

376   biosynthesis of unsaturated fatty acid metabolic pathway. The qRT-PCR results (File S7) showed

377   that the expression patterns of the four selected genes were very similar to those of RNA

378   sequencing data in Ch03 compared to Ch02 and in Ch04 compared to Ch03 (Fig. 10). In Ch01

379   compared to Ch02, the expression pattern of *DESA1* by qRT-PCR was consistent with that from

380   RNA sequencing (Fig. 10B), and *ACSL*, *fabG*, and *FatB* showed slight differences (Figs. 10A,

381   10C, 10D). It was acceptable that some differences in direct comparison between qRT-PCR and

382   RNA-seq results would occur due to biases in the library preparation for RNA-seq, different

383   normalization approaches, and other technical biases (Li et al., 2010; Wei, Chung & Zhao,

384 2011). qRT-PCR results of the selected genes were generally consistent with the RNA

385 sequencing results, indicating that the RNA-seq results were reliable.

CONCLUSIONS

387 *Chelidonium majus* is a well-known ant-dispersal plant with numerous elaiosome-bearing seeds.

388 In this study, we performed whole-transcriptome profiling of *C. majus* elaiosomes and seeds by

389 Illumina NGS sequencing, and used times series RNA-seq to study the expression patterns of

390 genes associated with fatty acid biosynthesis in them for the first time. A total of 63,064

391 unigenes were generated and 41 significantly DEGs involved in lipid metabolism pathways were

392 identified. By analysis and comparison of the DEGs at the four developmental stages of the

393 elaiosomes and seeds, we confirmed that the fatty acid biosynthesis in the elaiosome is consistent

394 with its cellular growth. The candidate genes involved in the biosynthesis of three primary fatty

395 acids in elaiosomes and seeds were also identified. Our research not only provided important

396 insights into fatty acid biosynthesis in *C. majus* elaiosomes and seeds, but also served as a

397 genomic resource for future evolutionary studies that focus on the convergent evolution of

398 myrmecochorus plants.

ACKNOWLEDGEMENTS

400 The authors would like to acknowledge Beijing Badaling National Forest Park for support with

401 plant materials. We also thank our colleagues for discussions and advice on this study. We are

402 also grateful to two anonymous reviewers for improving our manuscript.

403

REFERENCES

405 **Arora D, Sharma AA. 2013.** Review on phytochemical and pharmacological potential of genus

406   *Chelidonium. Pharmacognosy Journal* **5(4):**184-190.

407 **Beattie AJ. 1985.** The evolutionary ecology of ant-plant mutualisms. *Cambridge University Press*.

408 **Bergler H, Fuchsbichler S, Högenauer G, Turnowsky F. 1996.** The enoyl-[acyl-carrier-protein]

409   reductase (*FabI*) of Escherichia coli, which catalyzes a key regulatory step in fatty acid biosynthesis,

410   accepts *NADH* and *NADPH* as cofactors and is inhibited by palmitoyl-CoA. *European Journal of*

411   *Biochemistry* **242(3):**689-694.

412 **Bilder P, Lightle S, Bainbridge G, Ohren J, Finzel B, Sun F, Holley S, AI-Kassim L, Spessard C,**
413   **MeInick M, Newcomer M, Waldrop GL. 2006.** The structure of the carboxyltransferase
414   component of acetyl-coA carboxylase reveals a zinc-binding motif unique to the bacterial enzyme.
415   *Biochemistry* **45(6):**1712-1722.

416 **Black PN, DiRusso CC, Metzger AK, Heimert TL. 1992.** Cloning, sequencing, and expression of the
417   fadD gene of *Escherichia coli* encoding acyl coenzyme A synthetase. *Journal of Biological*
418   *Chemistry* **267(35)**:25513-25520.

419 **Bono JM, Heithaus ER. 2002.** Sex ratios and the distribution of elaiosomes in colonies of the ant,
420   *Aphaenogaster rudis*. *Insectes Sociaux* **49(4):**320-325.

421 **Boulay R, Coll-Toledano J, Cerdá X. 2006.** Geographic variations in *Helleborus foetidus* elaiosome
422   lipid composition: implications for dispersal by ants. *Chemoecology* **16(1):**1-7.

423 **Brew CR, O'Dowd DJ, Rae ID. 1989.** Seed dispersal by ants: behaviour-releasing compounds in
424   elaiosomes. *Oecologia* **80(4):**490-497.

425 **Chen G, Huang SZ, Chen SC, Chen YH, Liu X, Sun WB. 2016.** Chemical composition of diaspores of
426   the myrmecochorous plant *Stemona tuberosa* Lour. *Biochemical Systematics and Ecology* **64:**31-37.

427 **Ciccarelli D, Andreucci AC, Pagni AM, Garbari F. 2005.** Structure and development of the elaiosome
428   in *Myrtus communis* L.(Myrtaceae) seeds. *Flora* **200(4):**326-331.

429 **Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005.** Blast2GO: a universal tool
430   for annotation, visualization and analysis in functional genomics research. *Bioinformatics*
431   **21(18):**3674-3676.

432 **Dadd RH. 1973.** Insect nutrition: current developments and metabolic implications. *Annual Review of*
433   *Entomology* **18(1):**381-420.

434 **Fischer RC, Ölzant SM, Wanek W, Mayer V. 2005.** The fate of *Corydalis cava* elaiosomes within an
435   ant colony of *Myrmica rubra*: elaiosomes are preferentially fed to larvae. *Insectes Sociaux* **52(1):**55-
436   62.

437 **Fischer RC, Richter A, Hadacek F, Mayer V. 2008.** Chemical differences between seeds and
438   elaiosomes indicate an adaptation to nutritional needs of ants. *Oecologia* **155(3):**539-547.

439 **Gammans N, Bullock JM, Schönrogge K. 2005.** Ant benefits in a seed dispersal mutualism. *Oecologia*
440   **146(1):**43-49.

441 **Giladi I. 2006.** Choosing benefits or partners: a review of the evidence for the evolution of
442   myrmecochory. *Oikos* **112(3):**481-492.

443 **Gorb E, Gorb S. 2000.** Effects of seed aggregation on the removal rates of elaiosome-bearing
444   *Chelidonium majus* and *Viola odourata* seeds carried by *Formica polyctena* ants. *Ecological*
445   *Research* **15(2):**187-192.

446 **Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L,**
447   **Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, Palma F,**
448   **Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011.** Full-length
449   transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*
450   **29:**644-52.

451 **Heath RJ, Rock CO. 1996.** Roles of the *FabA* and *FabZ* β-hydroxyacyl-acyl carrier protein dehydratases
452   in *Escherichia coli* fatty acid biosynthesis. *Journal of Biological Chemistry* **271(44):**27795-27801.

453 **Hughes L, Westoby M. 1990.** Removal rates of seeds adapted for dispersal by ants. *Ecology* **71(1):**138-
454   148.

455 **Hughes L, Westoby M. 1992.** Effect of diaspore characteristics on removal of seeds adapted for dispersal
456   by ants. *Ecology* **73(4):**1300-1312.

457 **Hughes L, Westoby M, Jurado E. 1994.** Convergence of elaiosomes and insect prey: evidence from ant
458   foraging behaviour and fatty acid composition. *Functional Ecology* **8(3):**358-365.

459 **Kachroo A, Shanklin J, Whittle E, Lapchyk L, Hildebrand D, Kachroo P. 2007.** The *Arabidopsis*
460   stearoyl-acyl carrier protein-desaturase family and the contribution of leaf isoforms to oleic acid
461   synthesis. *Plant Molecular Biology* **63(2):**257-271.

462 **Kainou K, Kamisaka Y, Kimura K, Uemura H. 2006.** Isolation of Δ12 and ω3-fatty acid desaturase
463   genes from the yeast Kluyveromyces lactis and their heterologous expression to produce linoleic and
464   α-linolenic acids in *Saccharomyces cerevisiae*. *Yeast* **23(8):**605-612.

465 **Kanehisa M, Goto S. 2000.** KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*
466   **28(1)**:27-30.

467 **Kusmenoglu S, Rockwood LL, Gretz MR. 1989.** Fatty acids and diacylglycerols from elaiosomes of
468   some ant-dispersed seeds. *Phytochemistry* **28(10):**2601-2602.

469 **Lanza J, Schmitt MA, Awad AB. 1992.** Comparative chemistry of elaiosomes of three species of
470   *Trillium*. *Journal of Chemical Ecology* **18(2):**209-221.

471 **Lengyel S, Aaron DG, Andrew ML, Jonathan DM, Robert RD. 2010.** Convergent evolution of seed
472   dispersal by ants, and phylogeny and biogeography in flowering plants: a global survey. *Perspectives*
473   *in Plant Ecology, Evolution and Systematics* **12(1):**43-55.

474 **Li B, Dewey CN. 2011.** RSEM: accurate transcript quantification from RNA-Seq data with or without a
475   reference genome. *BMC Bioinformatics* **12:**323.

476 **Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN. 2010.** RNA-Seq gene expression estimation
477   with read mapping uncertainty. *Bioinformatics* **26:**493.

478 **Lisci M, Bianchini M, Pacini E. 1996.** Structure and function of the elaiosome in some angiosperm
479   species. *Flora* **191(2):**131-141.

480   **Magnuson K, Carey MR, Cronan JE. 1995.** The putative *fabJ* gene of *Escherichia coli* fatty acid
481   synthesis is the *fabF* gene. *Journal of Bacteriology* **177(12):**3593-3595.

482   **Magnuson K, Oh W, Larson TJ, Cronan Jr JE. 1992.** Cloning and nucleotide sequence of the *fabD*
483   gene encoding malonyl coenzyme A-acyl carrier protein transacylase of *Escherichia coli*. *FEBS*
484   *Letters* **299(3):**262-266.

485   **Maji AK, Banerji P. 2015.** *Chelidonium majus* L. (Greater celandine)–a review on its phytochemical and
486   therapeutic perspectives. *International Journal of Herbal Medcine* **3:**10-27.

487   **Mark S, Olesen JM. 1996.** Importance of elaiosome size to removal of ant-dispersed seeds. *Oecologia*
488   **107(1)**:95-101.

489   **Marshall DL, Beattie AJ, Bollenbacher WE. 1979.** Evidence for diglycerides as attractants in an ant-
490   seed interaction. *Journal of Chemical Ecology* **5(3):**335-344.

491   **Morales MA, Heithaus ER. 1998.** Food from seed-dispersal mutualism shifts sex ratios in colonies of
492   the ant *Aphaenogaster rudis*. *Ecology* **79(2):**734-739.

493   **Morrone O, Vega AS, Maier M. 2000.** Elaiosomes in *Urochloa paucispicata* (Poaceae: Panicoideae:
494   Paniceae): anatomy and chemical composition. *Flora* **195(4):**303-310.

495   **Oaxaca-Castillo D, Andreoletti P, Vluggens A, Yu S, Veldhoven PP, Reddy JK, Cherkaoui-Malki**
496   **M. 2007.** Biochemical characterization of two functional human liver acyl-CoA oxidase isoforms 1a
497   and 1b encoded by a single gene. *Biochemical and Biophysical Research Communications*
498   **360(2):**314-319.

499   **Oostermeijer JGB. 1989.** Myrmecochory in *Polygala vulgaris* L., *Luzula campestris* (L.) DC. and *Viola*
500   *curtisii* Forster in a Dutch dune area. *Oecologia* **78(3):**302-311.

501   **Reifenrath K, Becker C, Poethke HJ. 2012.** Diaspore trait preferences of dispersing ants. *Journal of*
502   *Chemical Ecology* **38(9):**1093-1104.

503   **Robinson MD, McCarthy DJ, Smyth GK. 2010.** EdgeR: a Bioconductor package for differential
504   expression analysis of digital gene expression data. *Bioinformatics* **26(1):**139-140.

505   **Salas JJ, Ohlrogge JB. 2002.** Characterization of substrate specificity of plant *FatA* and *FatB* acyl-ACP
506   thioesterases. *Archives of Biochemistry and Biophysics* **403(1):**25-34.

507   **Sernander R. 1906.** Entwurf einer Monographie der europaeischen Myrmekochoren. *Kungl Svenska*
508   *Vetenskapsakadem Handl* **41:**1-407.

509   **Servigne P, Detrain C. 2008.** Ant-seed interactions: combined effects of ant and plant species on seed
510   removal patterns. *Insectes Sociaux* **55(3):**220-230.

511   **Skidmore BA, Heithaus ER. 1988.** Lipid cues for seed-carrying by ants in *Hepatica americana*. *Journal*
512   *of Chemical Ecology* **14(12):**2185-2196.

513   **Soukup VG, Holman RT. 1987.** Fatty acids of seeds of North American pedicillate *Trillium* species.
514   *Phytochemistry* **26(4):**1015-1018.
515   **Stukey JE, McDonough VM, Martin CE. 1990.** The *OLE1* gene of *Saccharomyces cerevisiae* encodes
516   the delta 9 fatty acid desaturase and can be functionally replaced by the rat stearoyl-CoA desaturase
517   gene. *Journal of Biological Chemistry* **265(33):**20144-20149.
518   **Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM,**
519   **Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S,**
520   **Wolf YI, Yin JJ, Natale DA. 2003.** The COG database: an updated version includes eukaryotes.
521   *BMC Bioinformatics* **4:**41.
522   **Toomey RE, Wakil SJ. 1966.** Studies on the mechanism of fatty acid synthesis XV. Preparation and
523   general properties of β-ketoacyl acyl carrier protein reductase from Escherichia coli. *Biochimica et*
524   *Biophysica Acta* **116(2):**189-197.
525   **Unamba CIN, Nag A, Sharma RK. 2015.** Next generation sequencing technologies: the doorway to the
526   unexplored genomics of non-model plants. *Frontiers in Plant Science* **6:**1074.
527   **Van der Pijl L. 1982.** Principles of dispersal. *Springer: Berlin, Germany.*
528   **Wei Z, Chung LM, Zhao H. 2011.** Bias detection and correction in RNA-sequencing data. *BMC*
529   *Bioinformatics* **12:**290.
530   **Werker E. 1997.** Seed anatomy. *Gebruder Borntraeger Verlagsbuchhandlung, Stuttgart, Germany*.
531   **Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li CY, Wei L. 2011.** KOBAS 2.0: a
532   web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids*
533   *Research* **39(Web Server issue):**316-322.
534   **Yang S, Li JX, Xia XF, Zhao LC. 2015.** Elaiosome development in *Chelidonium* L. seeds. *Bulletin of*
535   *Botanical Research* **35(2):**191-199.

536

Table caption:

538

Table 1. Raw data and clean data statistics of RNA sequencing.

540   [1]GC %: The total number of bases G and C as a percentage of the total bases number; [2]Q20%:
541   The percentage of bases whose Phred value is greater than 20 accounts for the total bases;
542   [3]Q30%: The percentage of bases whose Phred value is greater than 30 accounts for the total
543   bases.

544

545   **Table 2.** DEGs in fatty acid biosynthesis metabolic pathway for different developmental stages.

546   [1]UR: Upgraded unigenes number, data in bracket represents significantly DEGs. [2]DR:

547   Downgraded uingenes number, data in bracket represents significantly DEGs.

548

549   **Figure caption:**

550

551   **Figure 1.** *Chelidonium majus* (A-H). **(A)** The naturally growing plants in Beijing Badaling

552   National Forest Park, China. **(B)** Flower buds (Ch01). **(C)** Pollinated flower (Ch02). **(D)** Fruits.

553   **(E)** Young seeds with elaiosomes (Ch03). **(F)** Mature seeds with elaiosomes (Ch04). **(G)**

554   Enlargement of mature seeds with elaiosomes, showing the white elaiosome is contrasted with

555   the glossy black seed. **(H)** Longitudinal section of mature seed and elaiosome, showing the red-

556   stained lipids in broken elaiosome and massive protein granulars in endosperm. **(I)** Diaspore

557   (seed + elaiosome) removal of *C. majus* by ant. Bars: (A) = 10 cm, (B-D) = 1 cm, (E) = 3 mm,

558   (F) = 2 mm, (G, I) = 1 mm, (H) = 100 μm. Abbreviation, EL: elaiosome, EN: endosperm.

559   (Photoed by Liangcheng Zhao)

560

561   **Figure 2.** Statistic of trinity assembly results. N50: Assembled unigenes by length from large to

562   small, and the length of it is half the length of the unigenes to the total (bp). N90: Assembled

563   unigenes by length from large to small, and the length of it is 90% the length of the unigenes to

564   the total (bp).

565

566   **Figure 3.** Characteristics of sequence homology of unigenes against the NR database: **(A)** E-

567   value distribution of BLAST hits for each unigene with an E-value cut-off of 1.0×10^{-5}. **(B)**

568   Similarity distribution of the top BLAST hits for each unigene. **(C)** Species distribution of

569   the top BLAST hits.

570   **Figure 4.** GO classification of assembled unigenes. The abscissa represents the secondary

571   classification terms of GO, the left ordinate represents the percentage of unigenes contained in

572   the secondary classification, the right ordinate represents the number of unigenes. This three

573   colors represent three categories, of which green represents biological process, blue represents

574   cellular component and red represents molecular function.

575

576   **Figure 5.** COG function classification of assembled unigenes. Each column represents the

577   functional classification of COG (Capital letters A~Z are used to indicate the specific meaning,

578   as indicated on the right), and the height of the column is the ordinate of the number of unigenes

579   with this function.

580

581   **Figure 6.** Distribution of unigenes in different KEGG pathways. Genes were divided into 5

582   clusters. A: Metabolism. B: Genetic Information Processing. C: Environmental Information

583   Processing. D: Cellular Process. E: Organismal Systems.

584

585   **Figure 7.** Pathways in 'lipid metabolism' group and number of the unigenes. Each column

586   represents the lipid metabolism pathways of KEGG, and the height of the column is the

587   ordinate of the number of unigenes with this pathway.

588

589   **Figure 8.** DEGs from the fatty acid biosynthesis and biosynthesis of unsaturated fatty acid

590   metabolism diagram. The locus tag and enzyme code shown in fatty acid biosynthesis and

591   biosynthesis of unsaturated fatty acid metabolism are described in a heat map.

592

593   **Figure 9.** The detail RNA degradation database in biosynthesis of unsaturated fatty acid

594   pathway. The red border genes are those that can annotate on these gene products.

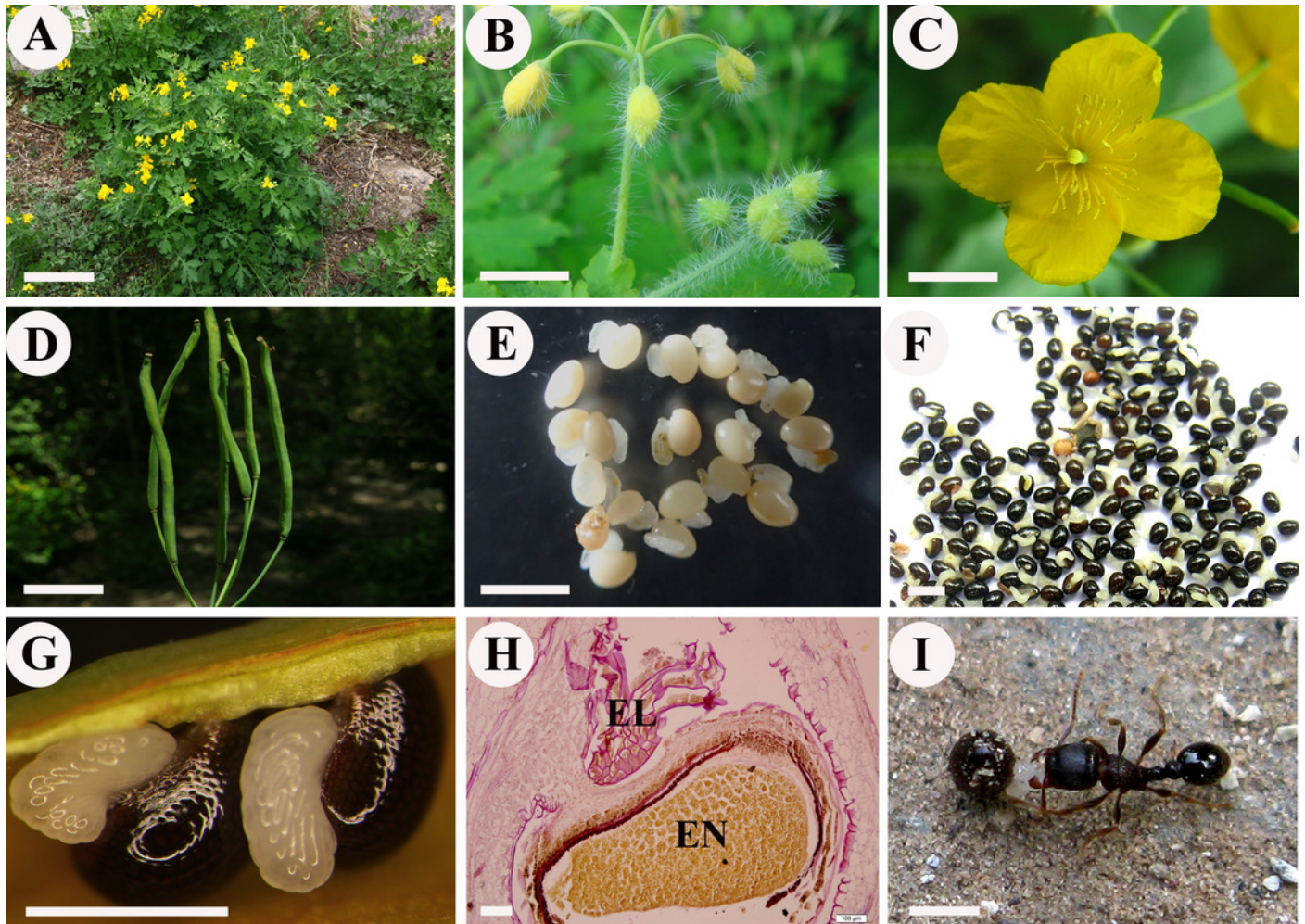
595

596   **Figure 10.** QRT-PCR quantification and RNA-seq results of four selected genes in different
597   developmental stages. (A) *ACSL*. (B) *DESA1*. (C) *FabG*. (D) *FatB*.

598

Figure 1

Chelidonium majus.

(A) The naturally growing plants in Beijing Badaling National Forest Park, China. **(B)** Flower buds (Ch01). **(C)** Pollinated flower (Ch02). **(D)** Fruits. **(E)** Young seeds with elaiosomes (Ch03). **(F)** Mature seeds with elaiosomes (Ch04). **(G)** Enlargement of mature seeds with elaiosomes, showing the white elaiosome is contrasted with the glossy black seed. **(H)** Longitudinal section of mature seed and elaiosome, showing the red-stained lipids in broken elaiosome and massive protein granulars in endosperm. **(I)** Diaspore (seed + elaiosome) removal of *C. majus* by ant. Bars: (A) = 10 cm, (B-D) = 1 cm, (E) = 3 mm, (F) = 2 mm, (G, I) = 1 mm, (H) = 100 μm. Abbreviation, EL: elaiosome, EN: endosperm. (Photoed by Liangcheng Zhao)

Statistic of trinity assembly results.

N50: Assembled unigenes by length from large to small, and the length of it is half the length of the unigenes to the total (bp). N90: Assembled unigenes by length from large to small, and the length of it is 90% the length of the unigenes to the total (bp).

Seq Length Distribution

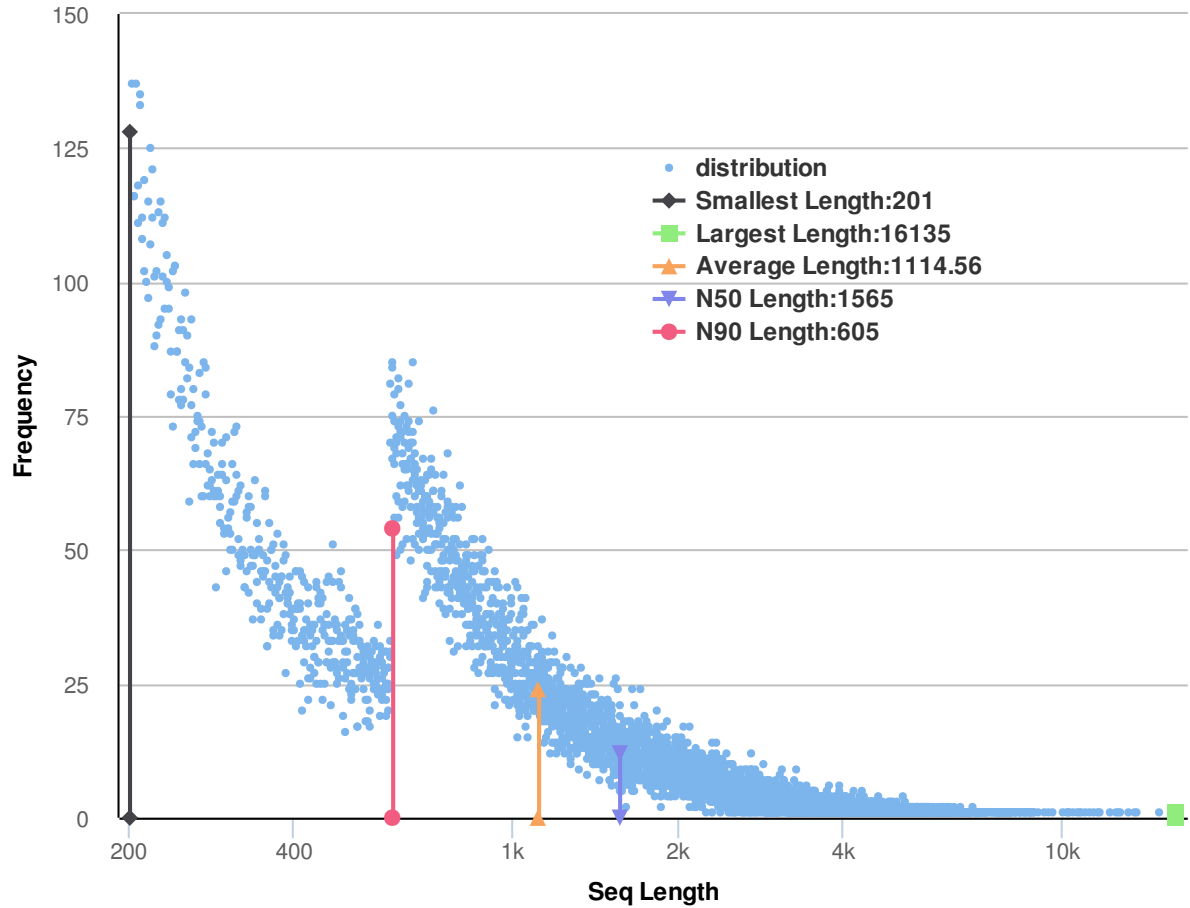
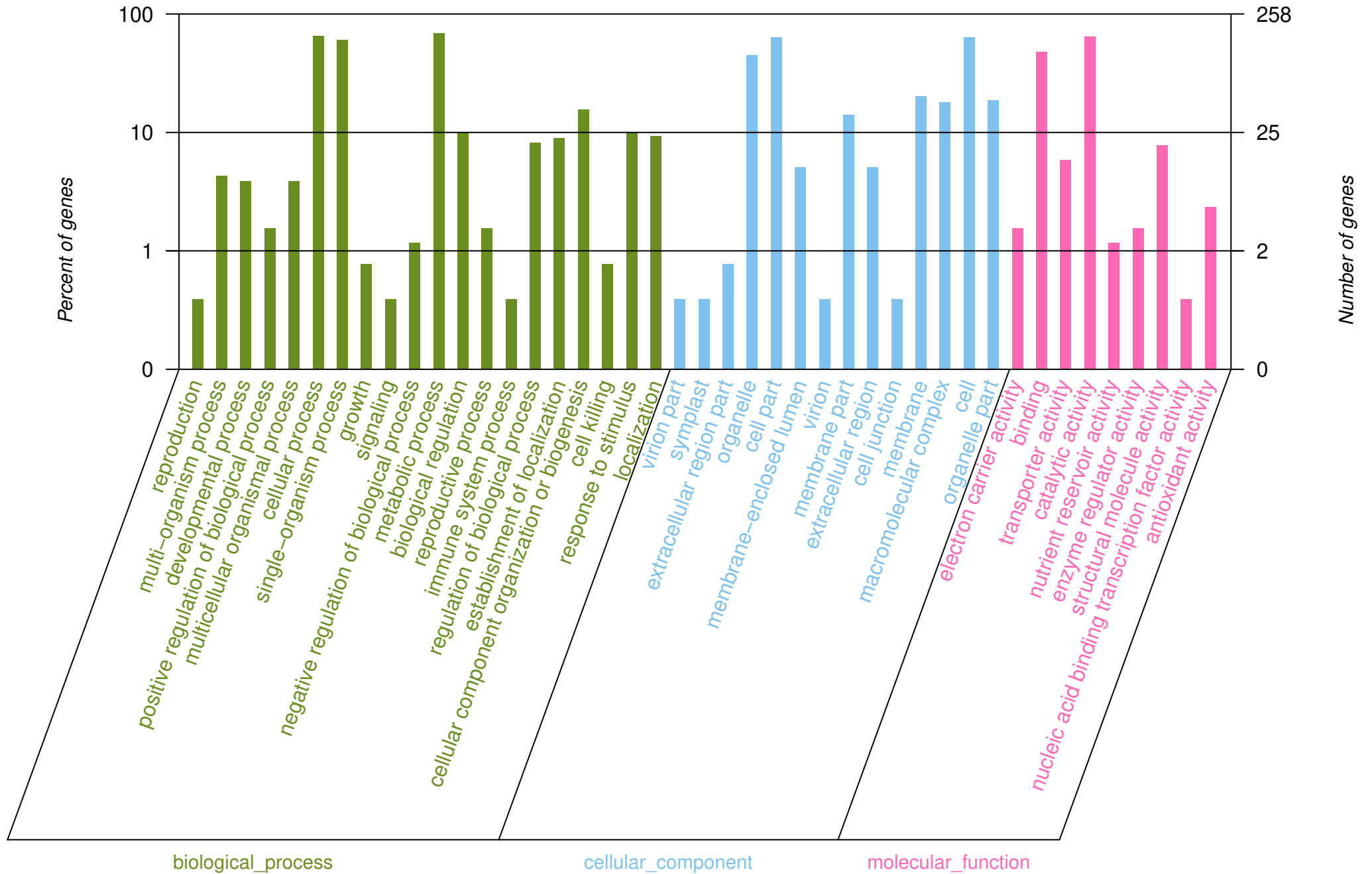



Figure 3

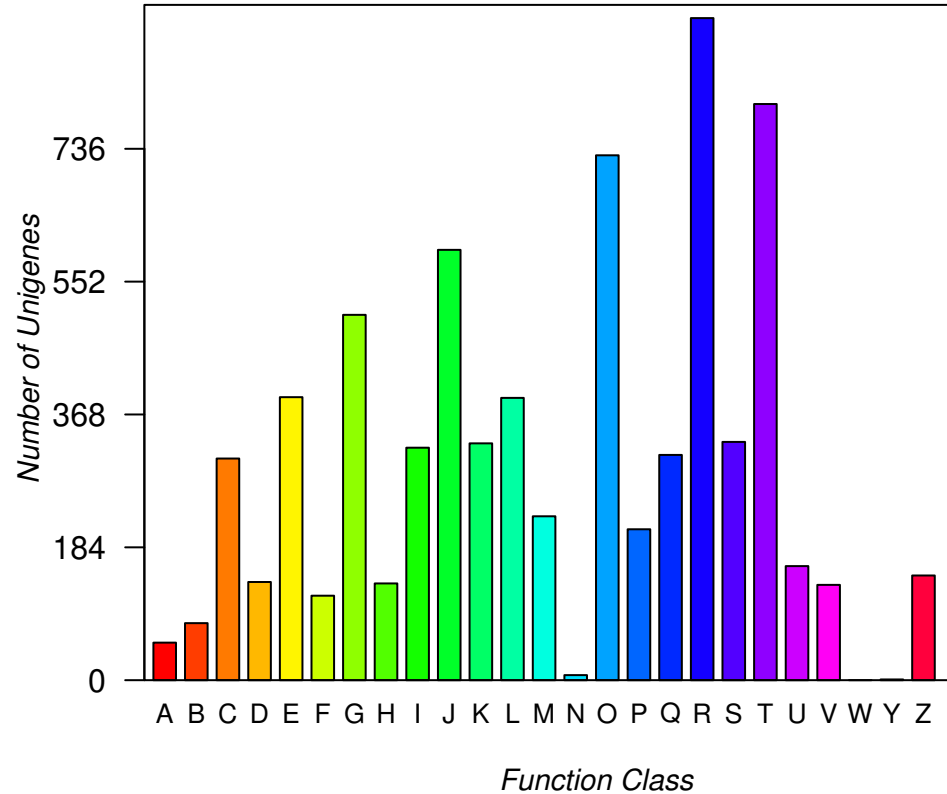
Characteristics of sequence homology of unigenes against NR database.

(A) E-value distribution of BLAST hits for each unigene with an E-value cut-off of 1.0×10^{-5}. (B) Similarity distribution of the top BLAST hits for each unigene. (C) Species distribution of the top BLAST hits.

GO classification of assembled unigenes.

The abscissa represents the secondary classification terms of GO, the left ordinate represents the percentage of unigenes contained in the secondary classification, the right ordinate represents the number of unigenes. This three colors represent three categories, of which green represents biological process, blue represents cellular component and red represents molecular function.

COG function classification of assembled unigenes.

Each column represents the functional classification of COG (Capital letters A~Z are used to indicate the specific meaning, as indicated on the right), and the height of the column is the ordinate of the number of unigenes with this function.

COG Function Classification

A : RNA processing and modification

B : Chromatin structure and dynamics

C : Energy production and conversion

D : Cell cycle control, cell division, chromosome partitioning

E : Amino acid transport and metabolism

F : Nucleotide transport and metabolism

G : Carbohydrate transport and metabolism

H : Coenzyme transport and metabolism

I : Lipid transport and metabolism

J : Translation, ribosomal structure and biogenesis

K : Transcription

L : Replication, recombination and repair

M : Cell wall/membrane/envelope biogenesis

N : Cell motility

O : Posttranslational modification, protein turnover, chaperones

P : Inorganic ion transport and metabolism

Q : Secondary metabolites biosynthesis, transport and catabolism

R : General function prediction only

S : Function unknown

T : Signal transduction mechanisms

U : Intracellular trafficking, secretion, and vesicular transport

V : Defense mechanisms

W : Extracellular structures

Y : Nuclear structure

Z : Cytoskeleton

Distribution of unigenes in different KEGG pathways.

Genes were divided into 5 clusters. A: Metabolism. B: Genetic Information Processing. C: Environmental Information Processing. D: Cellular Process. E: Organismal Systems.

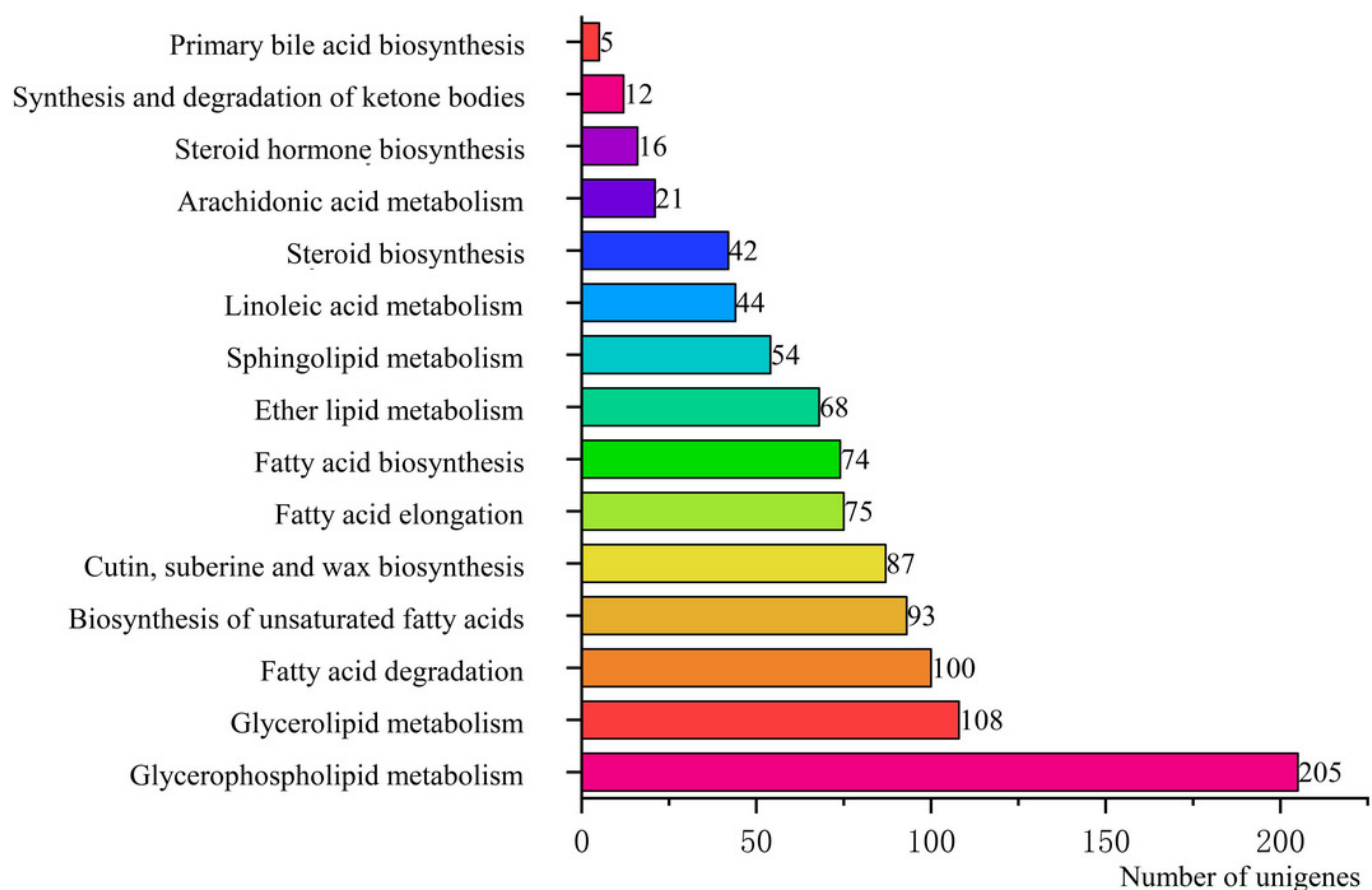
Category	Value
Sensory system	145
Nervous system	190
Immune system	415
Excretory system	85
Environmental adaptation	350
Endocrine system	453
Digestive system	210
Development	78
Circulatory system	100
Transport and catabolism	531
Cell motility	45
Cell growth and death	451
Cell communication	145
Signaling molecules and interaction	288
Signal transduction	878
Membrane transport	223
Translation	1099
Transcription	107
Replication and repair	448
Folding, sorting and degradation	675
Xenobiotics biodegradation and metabolism	117
Nucleotide metabolism	325
Metabolism of terpenoids and polyketides	275
Metabolism of other amino acids	289
Metabolism of cofactors and vitamins	311
Lipid metabolism	508
Glycan biosynthesis and metabolism	254
Global and overview maps	2736
Energy metabolism	590
Chemical structure transformation maps	33
Carbohydrate metabolism	1022
Biosynthesis of other secondary metabolites	331
Amino acid metabolism	800

Percent of Isoform(%)

Groups: E, D, C, B, A

Figure 7

Pathways in 'lipid metabolism' group and number of their unigenes.

Each column represents the lipid metabolism pathways of KEGG, and the height of the column is the ordinate of the number of unigenes with this pathway.

DEGs from the fatty acid biosynthesis pathway and biosynthesis of unsaturated fatty acid pathway diagram.

The locus tag and enzyme code shown in fatty acid biosynthesis pathway and biosynthesis of unsaturated fatty acid pathway are described in a heat map.

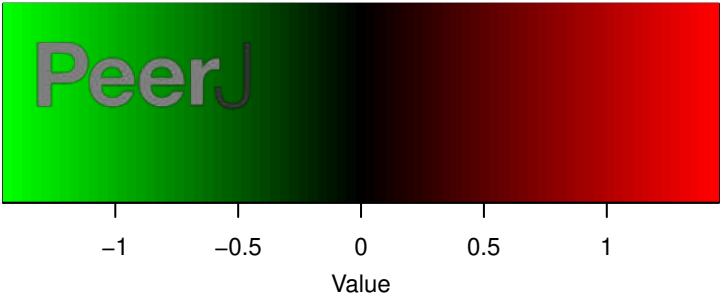
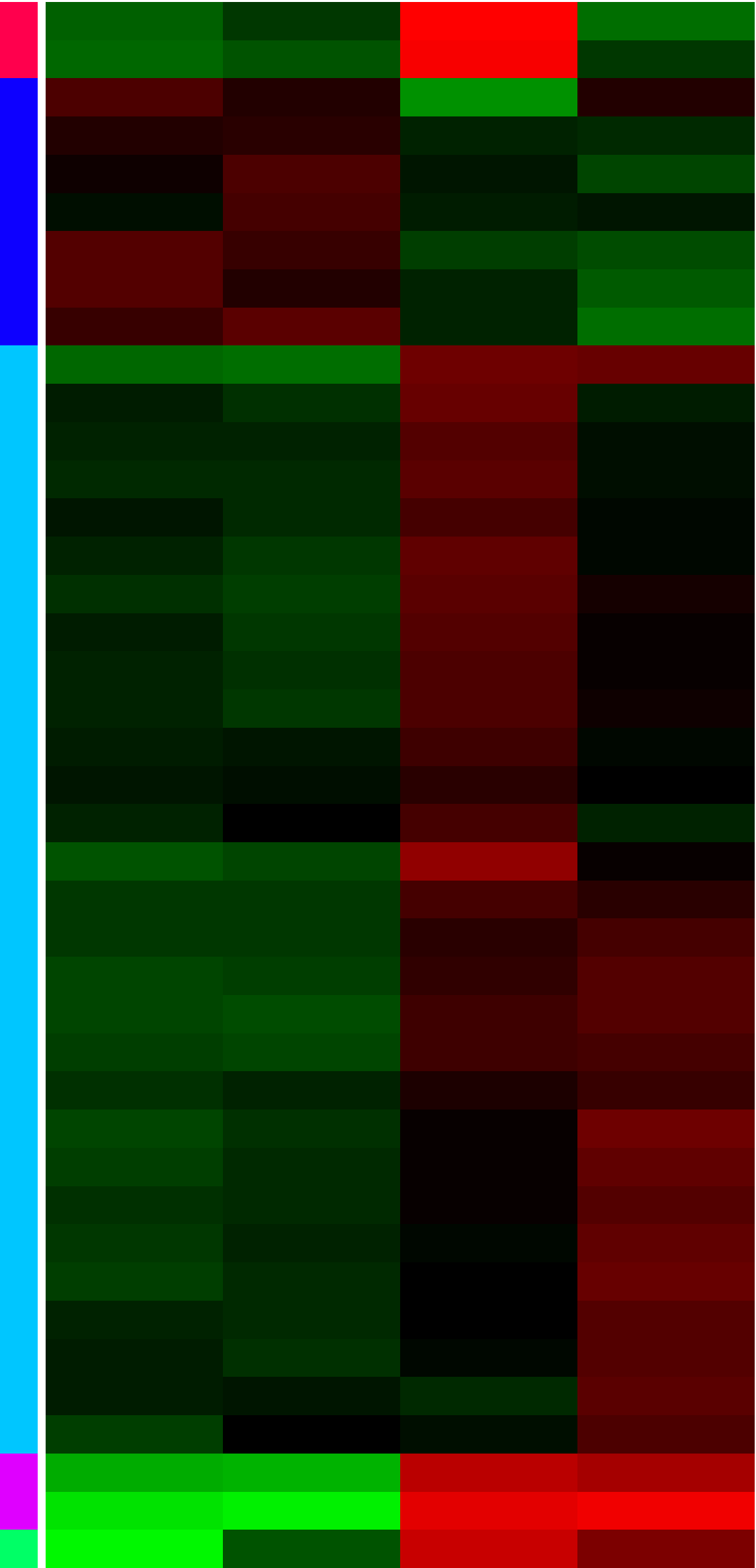
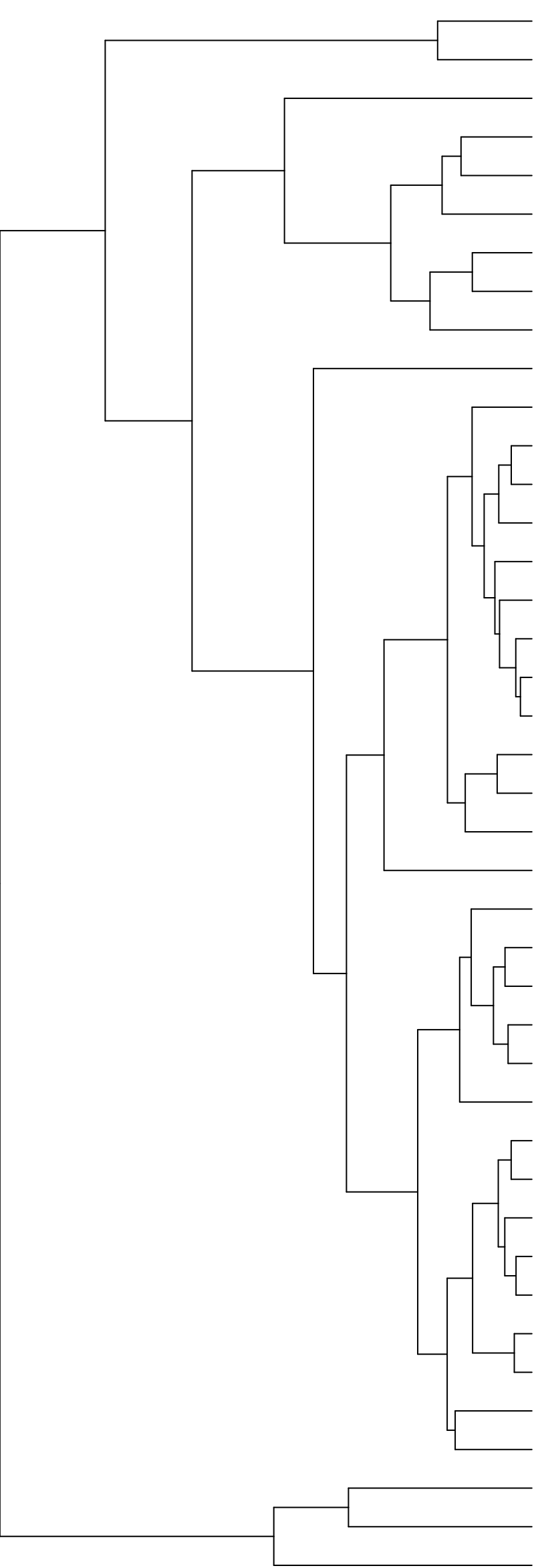
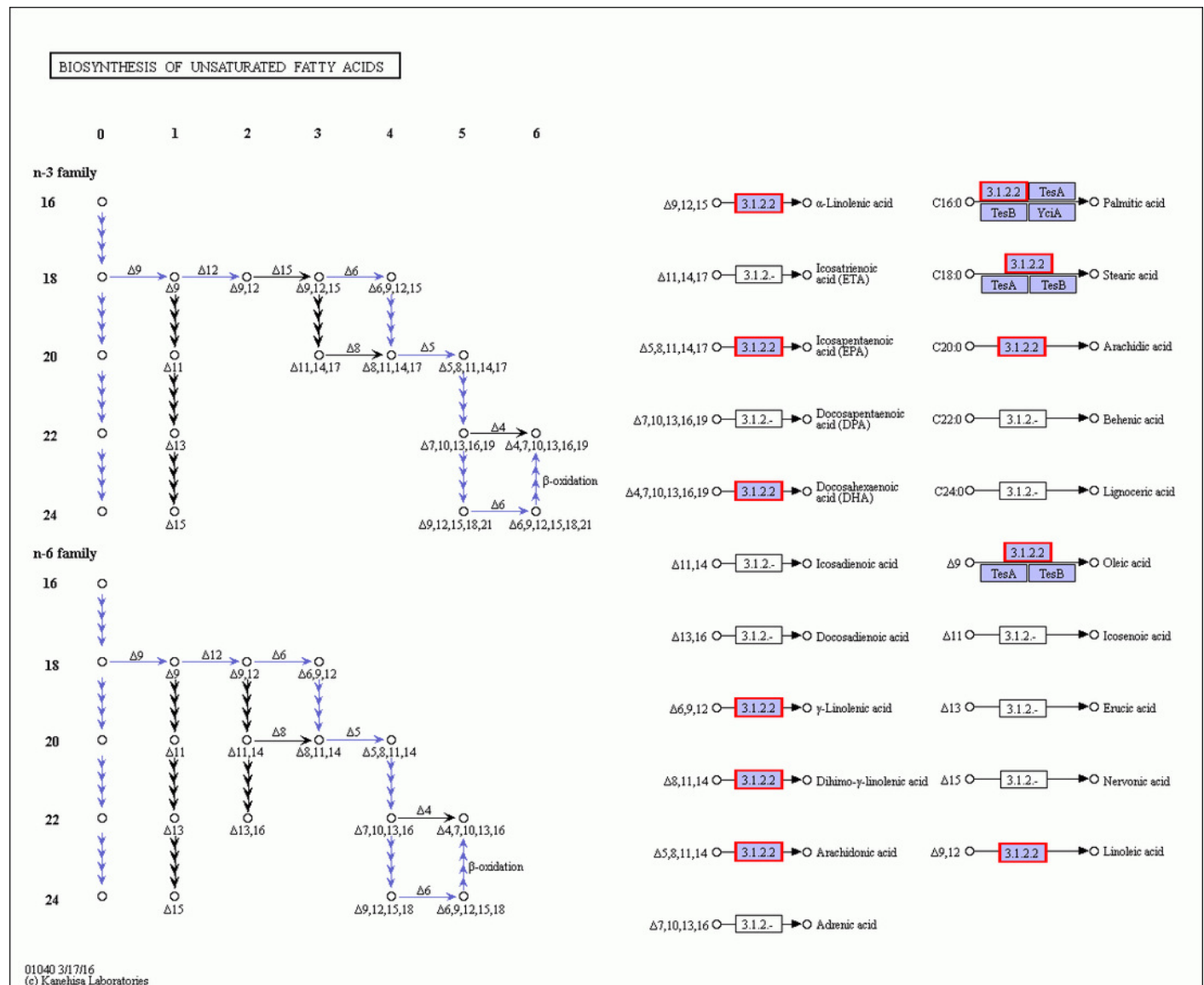
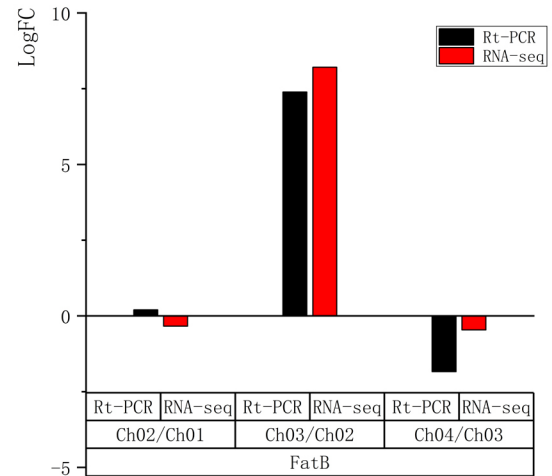
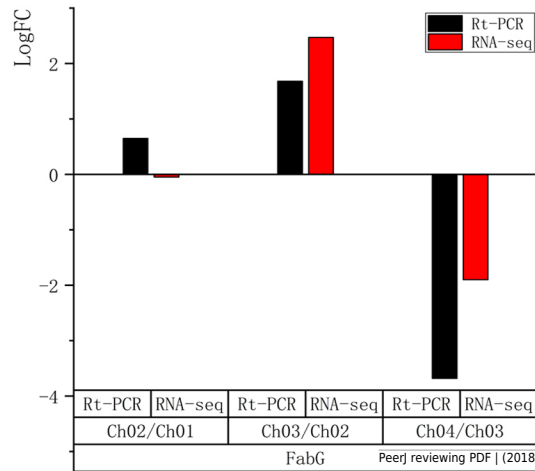
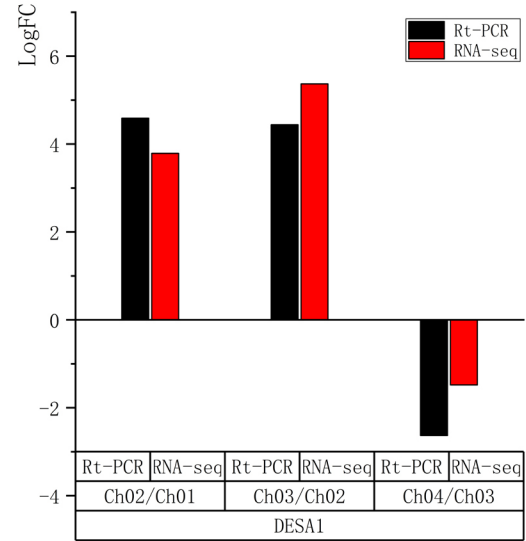
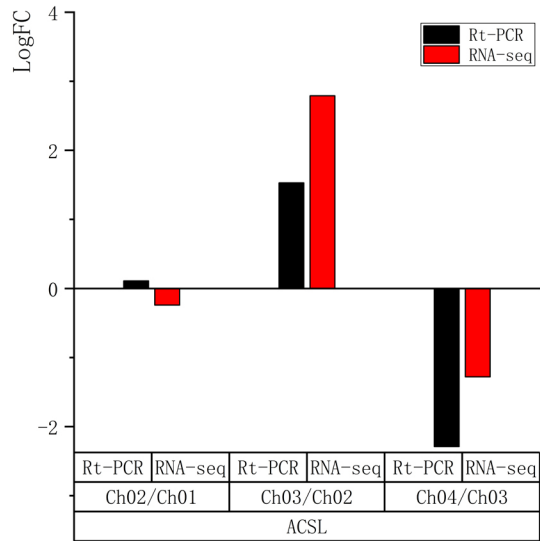



Figure 9

The detail RNA degradation database in biosynthesis of unsaturated fatty acid pathway.

The red border are those genes that can annotate on these gene products.

qRT-PCR quantification and RNA-seq results of four selected genes in different developmental stages of *C. majus* .

(A) *ACSL*. (B) *DESA1*. (C) *FabG*. (D) *FatB*.

Raw data and valid data statistics of RNA sequencing.

[1]GC %: The total number of bases G and C as a percentage of the total bases number;

[2]Q20%: The percentage of bases whose Phred value is greater than 20 accounts for the total

bases; [3]Q30%: The percentage of bases whose Phred value is greater than 30 accounts for

the total bases.

Table 1. Raw data and clean data statistics of RNA sequencing.

Sample_ID	Raw_Reads	Clean_Reads	Clean_Bases	Error%	Q20%[2]	Q30%[3]	GC%[1]
Ch01a	46,184,952	44,177,128	6,461,240,190	0.0138	97.66	93.3	45.02
Ch01b	56,008,592	54,268,742	7,954,613,129	0.0135	97.8	93.66	42.33
Ch01c	75,452,800	72,811,640	10,706,410,254	0.0136	97.78	93.53	42.81
Ch01(total)	**177,646,344**	**171,257,510**	**25,122,263,573**	**0.0136**	**97.75**	**93.5**	**43.39**
Ch02a	49,629,792	47,787,888	7,035,686,679	0.0135	97.81	93.63	43.52
Ch02b	48,985,314	46,706,162	6,845,643,793	0.014	97.58	93.13	44.03
Ch02c	48,604,760	46,676,484	6,849,388,718	0.0137	97.72	93.43	44.32
Ch02(total)	**147,219,866**	**141,170,534**	**20,730,719,190**	**0.0137**	**90.7**	**93.4**	**43.95**
Ch03a	46,603,464	45,034,614	6,637,326,310	0.0137	97.76	93.5	43.75
Ch03b	51,901,360	50,226,988	7,404,348,166	0.0135	97.83	93.67	43.56
Ch03c	53,968,514	52,416,848	7,726,545,508	0.0133	97.92	93.88	43.54
Ch03(total)	**152,473,338**	**147,678,450**	**21,768,219,984**	**0.0135**	**97.84**	**93.68**	**43.62**
Ch04a	52,126,852	50,435,348	7,426,968,007	0.0135	97.84	93.67	43.67
Ch04b	53,366,564	51,854,584	7,647,296,324	0.0133	97.92	93.86	43.7
Ch04c	52,973,818	51,264,748	7,548,125,708	0.0135	97.84	93.72	43.48
Ch04(total)	**158,467,234**	**153,554,680**	**22,622,390,039**	**0.0134**	**97.87**	**93.75**	**43.62**

2   [1]GC %: The total number of bases G and C as a percentage of the total bases number; [2]Q20%: The percentage of bases whose
3   Phred value is greater than 20 accounts for the total bases; [3]Q30%: The percentage of bases whose Phred value is greater than 30
4   accounts for the total bases.

DEGs in fatty acid biosynthesis metabolic pathway for different developmental stages.

[1]UR: Upgraded unigenes number, data in bracket represents significantly DEGs. [2]DR: Downgraded uingenes number, data in bracket represents significantly DEGs.

Table 2. DEGs in fatty acid biosynthesis metabolic pathway for different developmental stages.

KEGG pathway	Gene	Gene ID	Gene name	Enzyme	Ch01 vs Ch02 DEGs number		Ch02 vs Ch03 DEGs number		Ch03 vs Ch04 DEGs number	
					UR[1]	DR[2]	UR	DR	UR	DR
path:ko00061	Accase (accA)	K01962	acetyl-CoA carboxylase	EC 6.4.1.2	0	2(0)	1(1)	1(0)	1(0)	1(1)
path:ko00061	Accase (accB)	K02160	acetyl-CoA carboxylase	EC 6.4.1.2	2(0)	1(0)	3(3)	0	2(0)	1(1)
path:ko00061	Accase (accC)	K01961	acetyl-CoA carboxylase	EC 6.4.1.2	0	1(0)	1(1)	0	0	1(1)
path:ko00061	Accase (accD)	K01963	acetyl-CoA carboxylase	EC 6.4.1.2	1(0)	0	1(1)	0	0	1(1)
path:ko00061	Accase (ACACA)	K11262	acetyl-CoA carboxylase	EC 6.4.1.2	1(0)	0	0	1(0)	1(0)	0
path:ko00061	fabD	K00645	[acyl-carrier-protein] S-malonyltransferase	EC 2.3.1.39	0	1(0)	1(1)	0	0	1(1)
path:ko00061	fabF	K09458	3-oxoacyl-[acyl-carrier-protein] synthase II	EC 2.3.1.179	1(1)	4(0)	5(4)	0	1(0)	4(3)
path:ko00061	fabG	K00059	beta-ketoacyl reductase	EC 1.1.1.100	1(1)	2(0)	2(1)	1(1)	2(1)	1(1)
path:ko00061	DESA1	K03921	acyl carrier protein desaturase	EC 1.14.19.2	1(1)	1(0)	2(2)	0	0	2(0)
path:ko00061	fabZ	K02372	β-hydroxyacyl-acyl carrier protein dehydratases	EC 4.2.1.59	1(0)	0	1(1)	0	0	1(0)
path:ko00061	fabI	K00208	enoyl-[acyl-carrier-protein] reductase	EC 1.3.1.10	0	1(0)	1(1)	0	0	1(1)
path:ko00061	FATA	K10782	acyl-ACP thioesterase A	EC 3.1.2.14	0	1(0)	1(1)	0	0	1(0)
path:ko00061	FATB	K10781	acyl-ACP thioesterase B	EC 3.1.2.14	1(0)	2(0)	2(1)	0	2(1)	1(0)
path:ko00061	ACSL	K01897	acyl coenzyme A synthetase	EC 6.2.1.3	4(0)	4(0)	7(1)	1(0)	5(2)	3(2)

2   [1]UR: Upgraded unigenes number, data in bracket represents significantly DEGs. [2]DR: Downgraded uingenes number, data in bracket represents significantly DEGs.