農研機構
NARO 国立研究開発法人 農業・食品産業技術総合研究機構

Characterization of a gene regulatory network underlying astringency loss in persimmon fruit

メタデータ	言語: English 出版者: 公開日: 2019-09-27 キーワード (Ja): キーワード (En): Abiotic stress, Coexpression network analysis, Diospyros kaki, Fruit transcriptome, Proanthocyanidins 作成者: 尾上, 典之, 河野, 淳, 佐藤, 明彦 メールアドレス: 所属:
URL	https://repository.naro.go.jp/records/2679

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 International License.

Characterization of a gene regulatory network underlying astringency loss in persimmon fruit

Soichiro Nishiyama[1], Noriyuki Onoue[2], Atsushi Kono[2], Akihiko Sato[2], Keizo Yonemori[3], Ryutaro Tao[1]

[1]Graduate School of Agriculture, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan
[2]Institute of Fruit Tree and Tea Science, National Agriculture and Food Research Organization (NARO), Akitsu, Higashihiroshima, Hiroshima 739-2494, Japan
[3]Faculty of Agriculture, Ryukoku University, Seta Oe-cho, Otsu 520-2194, Japan

Corresponding author: Ryutaro Tao (rtao@kais.kyoto-u.ac.jp)

Abstract

Main conclusion **Transcriptome analysis of a persimmon population segregating for an astringency trait in fruit suggested central roles for a limited number of transcriptional regulators in the loss of proanthocyanidin accumulation.**

Persimmon (*Diospyros kaki*; 2n = 6x = 90) accumulates a large amount of proanthocyanidins (PAs) in its fruit, resulting in an astringent taste. Persimmon cultivars are classified into four types based on the nature of astringency loss and the amount of PAs at maturity. Pollination-constant and non-astringent (PCNA)-type cultivars stop accumulating PAs in the early stages of fruit development and their fruit can be consumed when still firm without the need for artificial deastringency treatments. While the PCNA trait has been shown to be conferred by a recessive allele at a single locus (*ASTRINGENCY*; *AST*), the exact genetic determinant remains unidentified. Here, we conducted transcriptome analyses to elucidate the regulatory mechanism underlying this trait using developing fruits of a F_1 population segregating for the PCNA trait. Comparisons of the transcriptomes of PCNA and non-PCNA individuals and hierarchical clustering revealed that genes related to the flavonoid pathway and to abiotic stress responses involving light stimulation were expressed coordinately with PA accumulation. Furthermore, coexpression network analyses suggested that three putative transcription factors were central to the PA regulatory network and that at least *DkMYB4* and/or *DkMYC1*, which have been reported to form a protein complex with each other for PA regulation, may have a central role in the differential expression of PA biosynthetic pathway genes between PCNA and non-PCNA.

Keywords Abiotic stress / Coexpression network analysis / *Diospyros kaki* / Fruit transcriptome / Proanthocyanidins

Acknowledgments We thank Dr. Takashi Akagi (Kyoto University) for critical advice and discussions on data interpretation. We also appreciate bioinformatics support from Dr. Luca Comai and Dr. Isabelle M Henry (UC Davis).

Abbreviations

CIPK CBL-interacting protein kinase
GO gene ontology
~~MBW MYB-bHLH-WD40~~
PA proanthocyanidin
PCNA pollination-constant and non-astringent
RPKM reads per kilobase of exon per million reads
TOM topological overlap measure

Introduction

Flavonoids are secondary metabolites that are extensively distributed throughout the plant kingdom and generally have protective roles against biotic and abiotic stresses. Proanthocyanidins (PAs), also known as condensed tannins, are one group of end-products derived from the flavonoid biosynthetic pathway. PAs are produced by polymerization of flavan-3-ols and are known to act as defense compounds against microbial pathogens, insect pests, and larger herbivores. PAs also act as free radical scavengers in plants to confer protective functions against biotic and abiotic stresses; these metabolites could, therefore, add human health benefits to fruit and vegetables (Dixon et al. 2005; Bagchi et al. 2014). PAs are also known to be important determinants of astringency and bitterness in fruit and vegetables (Dixon et al. 2005; Aron and Kennedy 2008).

Extensive studies have attempted to elucidate the PA biosynthetic pathway, particularly in model plants. From these studies, it has been determined that the biosynthetic pathway leading to the production of PA is shared with that for other flavonoids until just before the last step. Furthermore, genes encoding enzymes involved in the PA biosynthetic pathway and PA compartmentation into the vacuole, along with their transcriptional regulators such as the MYB-~~basic-helix-loop-helix (~~bHLH~~)~~-WD40~~ (MBW)~~ complex, have been identified and characterized (Lepiniec et al. 2006; Saito et al. 2013; Xu et al. 2015). However, less is known regarding the factors upstream of the ~~MBW~~ MYB-bHLH-WD40 complex or the gene regulatory mechanisms underlying the effects of environmental changes on the PA biosynthetic pathway (Li 2014). The latter is of particular

significance given that flavonoid accumulation is responsive to environmental conditions such as abiotic stress (Jaakola and Hohtola 2010; Fini et al. 2011; Xu et al. 2015).

Oriental persimmon (*Diospyros kaki* Thunb), which is an important fruit crop in East Asia, accumulates a large amount of PA in its fruit. PAs in persimmon fruit are highly polymerized and the accumulation continues until the middle of the fruit development, resulting in a strong astringent taste even at maturity (Akagi et al. 2011). Depending on their genotype, persimmon fruit are either astringent (inedible) or non-astringent (edible) at commercial harvest time when the fruit are still firm. However, regardless of the genotype, persimmon fruit becomes edible when it is over-ripe and becomes jelly soft. The commercial significance of astringency has led to soluble PA content at commercial maturity being used for persimmon cultivar classification (Yonemori et al. 2000). Persimmon cultivars are usually classified into four groups depending on the astringency loss at commercial maturity and pollination (seed formation): pollination constant and non-astringent (PCNA); pollination variant and non-astringent (PVNA); pollination variant and astringent (PVA); and pollination constant and astringent (PCA).

Among the four classes, PCNA-type cultivars stop accumulating PAs in the early stages of fruit development and the fruit becomes edible without any artificial de-astringency treatment (Yonemori and Matsushima 1985). There are two different types of PCNA traits that have different genetic controls, Japanese-type PCNA (JPCNA; represent as PCNA below) and Chinese-type PCNA (CPCNA). Both types are targeted by breeders because the fruit can be consumed without labor intensive and costly artificial de-astringency treatments (Sato and Yamada 2016). The PCNA trait has been shown to be conferred by a recessive allele at a single locus referred to as *ASTRINGENCY* (*AST*) (Ikeda et al. 1985; Yamada and Sato 2002; Akagi et al. 2012b). The exact determinant for JPCNA, however, remains to be identified.

Previous mRNA profiling experiments demonstrated that expression of the genes involved in the PA and shikimate pathways were substantially reduced in PCNA fruit (Ikegami et al. 2007; Akagi et al. 2009a). Furthermore, putative components of the MBW transcription complex, which is involved in the regulation of PA and shikimate pathway genes, such as the MYB transcription factor *DkMYB4* and bHLH transcription factor *DkMYC1* were shown to be specifically down-regulated in PCNA fruit at the early fruit development stage (Akagi et al. 2009b~~; Akagi et al.,~~ 2010; Su et al. 2012; Naval et al. 2016). It has been also reported that ABA signaling may be involved in PA biosynthesis in persimmon fruit via *DkMYB4* activation (Akagi et al.~~,~~ 2012a). Although these reports collectively indicated important roles for *DkMYB4* and *DkMYC1* in the loss of PA accumulation found in PCNA fruit (Akagi et al. 2009b; Su et al. 2012), further experimentation is required to confirm this hypothesis.

This study was aimed to characterize the gene regulatory network underlying PA accumulation and astringency loss in persimmon fruit via differential gene expression, clustering,

and coexpression network analyses using fruit transcriptomes obtained from individuals segregating for the PCNA trait. We discuss the potential role of *DkMYB4* and *DkMYC1* in the differential PA accumulation observed in PCNA and non-PCNA fruit.

Materials and methods

Plant materials

Twelve offspring (6 PCNAs and 6 non-PCNAs) of a cross 'Taiten' (non-PCNA) x 'Kanshu' (PCNA) and their parents, 'Taiten' and 'Kanshu', were used in this study. They were grown in the orchard of the Grape and Persimmon Research Division, NARO Institute of Fruit Tree and Tea Science, Japan. We determined the presence of the *AST* and *ast* alleles in each individual using the AST-F/PCNA-F/5R3R genetic marker for the *AST* and *ast* alleles (Kanzaki et al. 2010) (Supplementary Fig. S1). Three to six fruit were sampled from each individual on June 22, July 14, and August 19, 2015. The sampled fruit were weighed and peeled before the fruit flesh was diced, frozen in liquid nitrogen, and stored at -80°C until use. Frozen samples were used directly for RNA extraction or lyophilized for PA extraction

PA content

Soluble PAs were extracted from 10 mg of lyophilized flesh in 2 ml of 80% methanol at room temperature as previously described (Akagi et al. 2009b). The PA content of each sample was determined using 4-dimethylaminocinnamaldehyde (Li et al. 1996) with three technical replications and described as (+)-catechin equivalent.

Library construction and sequencing

Total RNA was extracted using the hot-borate method (Wan and Wilkins 1994) from the samples obtained on June 22 and July 14 and quantified using a Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA). An Illumina sequence library for each sample was constructed as described by Akagi et al. (2014) from 10 μg of total RNA. All 28 libraries were multiplexed and sequenced using Illumina Hiseq4000 (100-bp paired-end reads). All sequence data achieved in this study has been deposited in the DDBJ database (ID DRA006042).

Sequence processing, assembly, and functional annotation

All sequences were pre-processed using a custom python script (http://comailab.genomecenter.ucdavis.edu/index.php/Barcoded_data_preparation_tools). Sequences with base quality Phred scores of lower than 20 were trimmed and reads shorter than 35-bp were

145 discarded.

Pooled reads from the June 22 and July 14 samples of 'Taiten' were subjected to *de novo* assembly by CLC Genomics Workbench version 7.5.2 (CLC bio, Aarhus, Denmark) with the following parameters; 30-bp word size, automatic detection of bubble size, auto-detection of paired distance, discarding contigs shorter than 200-bp, scaffolding performed, and contigs updated by
150 re-mapping.

The assembled contigs were annotated with Blast2GO version 4.0.2 using Blastx against the NCBI nr protein database with an E-value cutoff < 1e−5; up to 20 hits were retained. InterProScan with default parameters was also used for the annotation. The "Mapping and Annotation" function with default parameters was performed with the Blastx and InterProScan
155 results to assign gene ontology (GO) terms, EC numbers, and possible descriptions to each contig.

Alignment and differential expression analyses

The reads from each library were aligned to the assembled contigs using bwa-mem program version 0.7.12 (Li and Durbin 2009) with default parameters. The number of mapped reads
160 for each contig was counted from the sam files; when a read was mapped to multiple contigs, multiple counts were made for the read. The expression level of each contig was calculated as reads per kilobase of exon per million sequence reads (RPKM) value and contigs with < 1 of average RPKM for all libraries were discarded.

The DESeq2 package (Love et al. 2014) was used with the count data before RPKM
165 normalization to determine differentially expressed genes between PCNA and non-PCNA samples. Data from each time point (June and July) were compared separately. A threshold value of false discovery rate (FDR) for differentially expressed genes was set to 0.1. The differentially expressed genes were hierarchically clustered by GENE-E software (https://software.broadinstitute.org/GENE-E/index.html) with the average-linkage method, using a
170 matrix of mean RPKM at each time point (June and July) and genotype (PCNA and non-PCNA).

Coexpression network analysis

Genes with low RPKM or low coefficient of variation scores (mean RPKM < 1 or CV < 0.5 for all samples) were removed before network construction. The WGCNA package (Langfelder
175 and Horvath 2008) was used to characterize the coexpression network based on RPKM value. Briefly, coexpression relationships between genes were defined as the adjacency calculated by raising Pearson's correlation coefficient to soft threshold power $\beta = 7$. The soft threshold was determined as a value that R^2 of the scale-free topology criteria amounted to 0.85. Coexpression modules were determined by hierarchical clustering based on the dissimilarity of topological overlap
180 measure (TOM) calculated from the adjacency matrix using TOM similarity function. Here, TOM

was defined as a coexpression relationship between two genes. The minimum module size was set as 50 and modules with less than 0.25 of the tree height were merged into their closest modules. Gene expression profiles for the modules were summarized as module eigengene by calculating the first principal component of the expression profile of genes assigned to the module. Correlations between module eigengene and PA traits were used for the quantification of relationships between the modules and PA accumulation.

The VisANT program (Hu et al. 2004) was used to visualize the network of a given module. A TOM cutoff for coexpression in the visualization was determined based on the strength of correlation of node degree distribution and set as 0.27 for the magenta module.

GO enrichment analysis

GO enrichment analysis was performed using the goseq package (Young et al. 2010) in R with default parameters. The annotations of genes fulfilling the RPKM cutoff were used as a reference for GO analysis. A threshold for the significance of enriched GO terms was set as FDR < 0.1 and the enriched GO terms with five or more differentially expressed genes were employed in this study.

Results

PA accumulation in a population segregating for the PCNA trait

Regardless of the astringency phenotype, all individuals in the segregating population, except for no. 118, showed almost the same level of PA concentration in fruit on June 22 (Fig. 1a); No. 118 showed less PA concentration than the others. On July 14 and on August 19, the PA content of all the progeny indicated genotype-dependent segregation. Non-PCNA individuals accumulated PA in fruit throughout the experimental period, while PA accumulation in the fruit of PCNA individuals slowed during the period of June to July and stopped completely in July to August. (Fig. 1b)

***De novo* transcriptome assembly and gene annotation**

Assembly of pooled clean reads of 'Taiten' obtained from the June (14,283,428 reads) and July samples (35,057,646 reads) resulted in 103,306 contigs with an average length of 580 bp. All the contigs were annotated with various sequence databases using Blast2GO. The summary of the assembly and functional annotation is shown in Table 1. Among the 103,306 contigs, 38,745 contigs (37.5%) had at least one hit in the nr database and 32,591 (31.5%) contigs among them were assigned at least one GO term. However, 64,561 contigs (62.5 %) showed no blast hits, probably

because of the short length of the contig sequences obtained.

Differential expression analysis and clustering

A total of 1,818 unique differentially expressed genes (FDR < 0.1) were identified in comparisons between PCNA and non-PCNA individuals at two stages. When June and July samples were compared separately, 530 genes were differentially expressed in June samples (192 were up- and 338 were down-regulated in PCNA) and 1639 genes were differentially expressed in July samples (703 were up- and 936 were down-regulated in PCNA).

The 1,818 genes were hierarchically clustered into eight clusters based on the mean RPKM values for the two genotypes (PCNA vs non-PCNA) and stages (June vs July) (Fig. 2). Among the eight clusters generated, cluster 2 included genes with higher expression in non-PCNA individuals in June than July and generally lower expression in PCNA than non-PCNA individuals, especially in July; this is consistent with the PA accumulation pattern. Forty-one GO terms were significantly enriched in cluster 2 (Table 2). Genes related to flavonoid biosynthesis, transportation, and photosynthesis were significantly enriched in cluster 2. Significantly enriched GO terms were also found in cluster 4 and cluster 6. Eleven GO terms, such as "glutamate biosynthetic process" and "iron-sulfur cluster binding" were enriched in cluster 4 (Supplementary Table S1) while 15 GO terms including "production of siRNA involved in RNA inference" and "RISC complex" were enriched in cluster 6 (Supplementary Table S2). No significantly enriched GO terms were detected in the other clusters.

As ABA signaling has been suggested to be involved in PA biosynthesis in persimmon fruit via *DkMYB4* activation (Akagi et al., 2012a), we focused on ABA-related genes and factors potentially related to abiotic stress, such as ABA receptor genes, WRKY, zinc finger, and calcineurin B-like proteins (CBL)-interacting protein kinase (CIPK)-like genes in cluster 2 (Supplementary Fig. S2). These genes were significantly down-regulated in PCNA fruit in July.

Coexpression network analysis

The genes (25,543) that met our defined cutoff values (RPKM < 1 and CV < 0.5) were classified into 25 modules using the WGCNA package (Fig. 3a). Among these modules, the magenta module showed the strongest positive correlation with both PA accumulation rate and concentration (Fig. 3b). GO enrichment analysis for the magenta module indicated that GO terms related to the flavonoid pathway and photosynthesis were significantly enriched, consistent with the general characteristics of cluster 2 (Supplementary Table S3). GO analysis for the other modules significantly correlated ($P < 0.05$) with PA concentration. Modules with negative correlation showed enrichment of terms related to stress or defense response; the "chitin catabolic process" term was enriched in the darkred module, the "pectin esterase inhibitor activity" term was enriched in the

thistle1 module, and the "protein-disulfide reductase activity" term was enriched in the greenyellow module (Supplementary Table S4). There was no enriched GO term in the lightcyan module. When
255 examining the modules showing a positive correlation with PA concentration, the term photosynthesis was enriched in the grey60 module, the "microtubule binding" and "oxidation-reduction process" terms were enriched in the blue module, and the RNA interference-related terms were enriched in the cyan module (Supplementary Table S5).

We further focused on the magenta module because it contained the majority of the
260 enzymes associated with the PA pathway and showed a strong correlation with PA accumulation. Network visualization of the magenta module based on TOM values indicated that genes for several enzymes had a number of interactions with genes in this module, while transcription factors or CIPKs had fewer interactions with other genes (Fig. 4, Supplementary Table S6). With network centrality as determined by the sum of TOM values, four contigs annotated as transcription factors
265 were found in the top 100 list in the magenta module (Supplementary Table S7). As Contig20409 and Contig12413 were both identical to a part of the *DkMYC1* sequence (data not shown), the number of putative transcription factors with high centrality in this coexpression module appeared to be only three; *DkMYB4*, *DkMYC1,* and *LSD ONE LIKE 1* (*LOL1*).

270 **Discussion**

High levels of PA accumulation in fruit is a typical characteristic of persimmon, with the amount of soluble PAs at commercial maturity one of the most commercially important traits in this species. Studies on PA in persimmon fruit have, therefore, been conducted from various viewpoints.
275 Analysis of the PA biosynthetic pathway has shown that PA in persimmon fruit is thought to be synthesized through the shikimate pathway, the flavonoid pathway, and finally the PA-specific pathway in a same manner as other species (Akagi et al. 2011). Furthermore, PCNA-specific down-regulation has been reported for expression of the genes encoding enzymes in these pathways such as 3-dehydroquinate dehydratase/shikimate 5-dehydrogenase and
280 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase in the shikimate pathway, flavonoid 3'5'-hydroxylase and phenylalanine ammonia lyase in the flavonoid pathway, and anthocyanidin reductase in the PA-specific pathway (Ikegami et al. 2007; Akagi et al. 2009a). mRNAseq and GO analyses in this study indicated that the expression of the genes encoding these enzymes was already altered in PCNA in late June when the first quarter of fruit development had been completed under
285 the local climate conditions. This suggests that *AST* has a function in the early stages of fruit development; this is fully consistent with previous studies (Ikegami et al. 2007; Akagi et al. 2009a).

This study showed that down-regulation of the shikimate pathway, upstream of the flavonoid pathway, could accompany the PCNA trait. In the shikimate pathway,

phosphoenolpyruvate and erythrose 4-phosphate are converted to chorismate, a precursor of the
290 aromatic amino acids and many aromatic secondary metabolites (Herrmann and Weaver 1999) including precursors of PAs. Hierarchical clustering of genes differentially expressed between PCNA and non-PCNA individuals in this study showed that "3-deoxy-7-phosphoheptulonate synthase activity" and "shikimate 3-dehydrogenase (NADP+) activity", both of which represent key enzymatic activities in the shikimate pathway, were significantly enriched in cluster 2, indicating
295 that the expression of these genes was correlated with PA accumulation. Given that PAs account for approximately 1% of the flesh weight of non-PCNA fruit at maturity (Taira et al. 1998), coordinated changes in the expression of genes associated with the shikimate and flavonoid pathways in persimmon may be attributed to a substantial demand for carbon flow to these pathways in developing non-PCNA fruit.

300 GO analysis of the genes assigned to cluster 2 and the magenta module indicated that their global expression profiles were quite similar to those observed with overexpression of *VvMYBPA1* and *VvMYBPA2* in grapevine, except for differences in the GO photosynthesis/chloroplast term (Terrier et al. 2008). Association of the genes related to photosynthesis and PA accumulation is a unique finding in cluster 2 and the magenta module (Table 2, Supplementary Table S3). Interestingly,
305 the associated genes seemed to be specific to light harvesting in photosystem I. This observation could be related to the light sensitivity of both flavonoid biosynthesis and photosynthesis genes, though further work is required to confirm this (Berry et al. 2013; Zoratti et al. 2014). Alternatively, it may that the enrichment of photosynthesis genes was a pleiotropic effect of *AST* expression or a gene(s) closely linked to *AST*.

310 In *VvMYBPA*-overexpressing grapevine hairy roots, both enzymes in the flavonoid pathway and also CIPKs were activated (Terrier et al. 2008), consistent with our results. CBLs and their target CIPKs are known to act as Ca^{2+} sensor-kinase modules (Shi et al. 1999), and it has previously been proposed that Ca^{2+} and ABA signaling form tight signaling networks in the context of plant stress signaling (Weinl and Kudla 2009; Edel and Kudla 2016). There are two possible
315 explanations for the *AST*-mediated role of CIPKs. One possibility is that the unknown master regulator has a function in abiotic stress-related signaling induction and affects the ABA-Ca^{2+} signaling complex, resulting in PA accumulation via ABA signal transduction. This possibility is supported by Akagi et al. (2012a) who showed that DkbZIP5 recognized an ABA-responsive element in the promoter region of *DkMyb4* and acted as its transcriptional regulator in an
320 ABA-dependent manner. Another possibility is that CIPKs were secondarily induced by stress derived from the vast amounts of PAs in non-PCNA individuals. This possibility is also discussed for *VvMYBPAs*-overexpressing grapevine (Terrier et al. 2008). The fact that exogenous ABA treatment sometimes had no clear effects on PCNA fruit (Nishiyama et al. 2014), as opposed to the results observed by Akagi et al (2012a), may support this possibility. Interestingly, the contig

325 annotated as PYL, an ABA receptor, was strongly down-regulated in PCNA fruit (Supplementary Fig. S2). Further analysis of this gene could give provide clues about the regulatory relationship between ABA signaling and PA accumulation.

In the coexpression network analysis, several genes encoding enzymes in the PA pathway such as *phenylalanine ammonia-lyase* and *cinnamate-4-hydroxylase*, had a great number of 330 relationships within the magenta module. This indicates that the expression of genes encoding enzymes in the PA pathway was strongly associated with each other under the control of *AST*. As for the transcriptional regulators, three putative transcription factor-encoding genes, *DkMYB4*, *DkMYC1,* and *LOL1*-like, were identified as having high connectivity within the magenta module (Fig. 4, Supplementary Table S7). This suggests that *AST*-mediated PA regulation is controlled by only a 335 few transcriptional regulators. Among these three factors, *DkMYB4* has been shown to have a PA regulatory function in persimmon and kiwifruit (*Actinidia deliciosa*) callus (Akagi et al. 2009b). Although the function of DkMYC1 has yet to be fully elucidated, it has previously been shown that *DkMYC1* expression was associated with PA accumulation and that DkMYC1 bound to DkMYB4 (Su et al. 2012; Naval et al. 2016). LOL1 is a zinc finger protein that was proposed function as a 340 transcriptional regulator or scaffold for cell death execution genes in response to oxidative stress (Epple et al. 2003). Although there is no report on the hierarchical relationship between the expression of *LOL1* and PA accumulation, the high connectivity of the *LOL1*-like gene in the magenta module suggests that there would be a substantial amount of abiotic stress in developing non-PCNA fruit. If *LOL1*-like was found to be upregulated by the secondary consequences of PA 345 accumulation, only *DkMYB4* and/or *DkMYC1* would have central roles in the differential expression of genes in the PA pathway between PCNA and non-PCNA fruit. As the numbers of connections with *DkMYB4* and *DkMYC1* in the coexpression network were relatively few (Fig. 4) and the functional annotation in this study remains incomplete, there is still a possibility that other regulators are also involved in the differential expression of PA pathway genes between PCNA and non-PCNA 350 fruit.

In the modules negatively correlated with PA concentration, the enrichment of genes possibly involved in defense response, such as chitinase activity (Collinge et al. 1993) and pectin metabolism (Pelloux et al. 2007), were detected (Supplementary Table S4). This could be related to an essential role of PAs as defense compounds. Interestingly, although PA accumulation in PCNA 355 was terminated in the early stages of fruit development, we could detect substantial amounts of PAs in PCNA fruit during our experimental term (Fig. 1). Taken together, the induction of defense responsive genes in PCNA fruit implies that developing persimmon fruit undergo substantial attacks from biotic stressors and that the vast amount of PAs in non-PCNA fruit is necessary for protection from these stresses.

360 In this study, we proposed that the differential gene expression caused by mutation of *AST*

was regulated by only a few transcriptional regulators in developing fruit, with *DkMYB4* and/or *DkMYC1* having a central role in the differential PA accumulation in PCNA vs non-PCNA fruit. Clustering and coexpression analyses suggested that most of the genes positively correlated with PA accumulation encode enzymes in the PA pathway, indicating that *AST* has limited pleiotropic effects. However, several functional gene sets such as those related to ABA signaling, CIPKs, and photosystem I were identified in the same cluster/module that was related to PA accumulation. Since these genes could be potential targets of PA regulators and thus *AST*; however, none of these genes were located on the diploid *D. lotus* genomic region syntenic to *D. kaki AST* locus and thus they were unlikely to be *AST* (S. Nishiyama, N. Onoue, A. Kono, A. Sato, K. Ushijima, H. Yamane, R. Tao and K. Yonemori in preparation). Further analysis of the genes and gene functions highlighted from the transcriptome analyses conducted in this study will lead to further understanding of the PA accumulation-related characteristics of persimmon fruit. Notably, defense systems appeared to be preferentially activated in PCNA fruit with less PAs than non-PCNA fruit. The PCNA trait is, therefore, a beneficial example to understand the role of PAs in plant-environment interactions and stress physiology.

Author contribution statement SN, AS, KY, and RT conceived and designed this study. SN conducted the experiments. KY and RT supervised the experiments. NO, AK, and AS contributed to the construction and management of the experimental persimmon trees. SN drafted the manuscript. All authors participated data interpretation and approved the manuscript.

Compliance with ethical standards

Funding: This work was supported by Grant-in-Aid for JSPS Research Fellow to SN (Grant Number JP16J10408), and for Scientific Research (B) to KY (Grant Number JP16H04876) from Japan Society for the Promotion of Science.

Conflict of interest: The authors declare that they have no conflicts of interest.

References

Akagi T, Ikegami A, Suzuki Y, Yoshida J, Yamada M, Sato A, Yonemori K (2009a) Expression balances of structural genes in shikimate and flavonoid biosynthesis cause a difference in proanthocyanidin accumulation in persimmon (*Diospyros kaki* Thunb.) fruit. Planta 230:899–

915. doi: 10.1007/s00425-009-0991-6

Akagi T, Ikegami A, Tsujimoto T, Kobayashi S, Sato A, Kono A, Yonemori K (2009b) DkMyb4 is a Myb transcription factor involved in proanthocyanidin biosynthesis in persimmon fruit. Plant Physiol 151:2028–2045. doi: 10.1104/pp.109.146985

Akagi T, Ikegami A, Yonemori K (2010) DkMyb2 wound-induced transcription factor of persimmon (*Diospyros kaki* Thunb.), contributes to proanthocyanidin regulation. Planta 232:1045–1059. doi: 10.1007/s00425-010-1241-7

Akagi T, Katayama-Ikegami A, Yonemori K (2011) Proanthocyanidin biosynthesis of persimmon (*Diospyros kaki* Thunb.) fruit. Sci Hortic 130:373–380. doi: 10.1016/j.scienta.2011.07.021

Akagi T, Katayama-Ikegami A, Kobayashi S, Sato A, Kono A, Yonemori K (2012a) Seasonal abscisic acid signal and a basic leucine zipper transcription factor, DkbZIP5, regulate proanthocyanidin biosynthesis in persimmon fruit. Plant Physiol 158:1089–1102. doi: 10.1104/pp.111.191205

Akagi T, Tao R, Tsujimoto T, Kono A, Yonemori K (2012b) Fine genotyping of a highly polymorphic *ASTRINGENCY*-linked locus reveals variable hexasomic inheritance in persimmon (*Diospyros kaki* Thunb.) cultivars. Tree Genet Genomes 8:195–204. doi: 10.1007/s11295-011-0432-0

Akagi T, Henry IM, Tao R, Comai L (2014) A Y-chromosome-encoded small RNA acts as a sex determinant in persimmons. Science 346:646–650. doi: 10.1126/science.1257225

Aron PM, Kennedy JA (2008) Flavan-3-ols: Nature, occurrence and biological activity. Mol Nutr Food Res 52:79–104. doi: 10.1002/mnfr.200700137

Bagchi D, Swaroop A, Preuss HG, Bagchi M (2014) Free radical scavenging, antioxidant and cancer chemoprevention by grape seed proanthocyanidin: an overview. Mutat Res 768:69–73. doi: 10.1016/j.mrfmmm.2014.04.004

Berry JO, Yerramsetty P, Zielinski AM, Mure CM (2013) Photosynthetic gene expression in higher plants. Photosynth Res 117:91–120. doi: 10.1007/s11120-013-9880-8

Collinge DB, Kragh KM, Mikkelsen JD, Nielsen KK, Rasmussen U, Vad K (1993) Plant chitinases. Plant J 3:31–40

Dixon RA, Xie DY, Sharma SB (2005) Proanthocyanidins - A final frontier in flavonoid research? New Phytol 165:9–28. doi: 10.1111/j.1469-8137.2004.01217.x

Edel KH, Kudla J (2016) Integration of calcium and ABA signaling. Curr Opin Plant Biol 33:83–91. doi: 10.1016/j.pbi.2016.06.010

Epple P, Mack AA, Morris VR, Dangl JL (2003) Antagonistic control of oxidative stress-induced cell death in *Arabidopsis* by two related, plant-specific zinc finger proteins. Proc Natl Acad Sci USA 100:6831–6836. doi: 10.1073/pnas.1130421100

Fini A, Brunetti C, Di Ferdinando M, Ferrini F, Tattini M (2011) Stress-induced flavonoid

biosynthesis and the antioxidant machinery of plants. Plant Signal Behav 6:709–11. doi: 10.4161/psb.6.5.15069

Herrmann KM, Weaver LM (1999) The shikimate pathway. Annu Rev Plant Physiol Plant Mol Biol 50:473–503. doi: 10.1146/annurev.arplant.50.1.473

Hu Z, Mellor J, Wu J, DeLisi C (2004) VisANT: an online visualization and analysis tool for biological interaction data. BMC Bioinformatics 5:17. doi: 10.1186/1471-2105-5-17

Ikeda I, Yamada M, Kurihara A, Nishida T (1985) Inheritance of astringency in Japanese persimmon. J Japanese Soc Hortic Sci 54:39–45

Ikegami A, Eguchi S, Kitajima A, Inoue K, Yonemori K (2007) Identification of genes involved in proanthocyanidin biosynthesis of persimmon (*Diospyros kaki*) fruit. Plant Sci 172:1037–1047. doi: 10.1016/j.plantsci.2007.02.010

Jaakola L, Hohtola A (2010) Effect of latitude on flavonoid biosynthesis in plants. Plant, Cell Environ 33:1239–1247. doi: 10.1111/j.1365-3040.2010.02154.x

Kanzaki S, Akagi T, Masuko T, Kimura M, Yamada M, Sato A, Mitani N, Ustunomiya N, Yonemori K (2010) SCAR markers for practical application of marker-assisted selection in persimmon (*Diospyros kaki* Thunb.) breeding. J Japanese Soc Hortic Sci 79:150–155. doi: 10.2503/jjshs1.79.150

Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9:559. doi: 10.1186/1471-2105-9-559

Lepiniec L, Debeaujon I, Routaboul J-M, Baudry A, Pourcel L, Nesi N, Caboche M (2006) Genetics and biochemistry of seed flavonoids. Annu Rev Plant Biol 57:405–430. doi: 10.1146/annurev.arplant.57.032905.105252

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. doi: 10.1093/bioinformatics/btp324

Li S (2014) Transcriptional control of flavonoid biosynthesis: Fine-tuning of the MYB-bHLH-WD40 (MBW) complex. Plant Signal Behav 8:1–7. doi: 10.4161/psb.27522

Li Y-G, Tanner G, Larkin P (1996) The DMACA–HCl protocol and the threshold proanthocyanidin content for bloat safety in forage legumes. J Sci Food Agric 70:89–101. doi: 10.1002/(SICI)1097-0010(199601)70:1<89::AID-JSFA470>3.0.CO;2-N

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. doi: 10.1186/s13059-014-0550-8

Naval M del M, Gil-Muñoz F, Lloret A, Besada C, Salvador A, Badenes ML, Ríos G (2016) A WD40-repeat protein from persimmon interacts with the regulators of proanthocyanidin biosynthesis DkMYB2 and DkMYB4. Tree Genet Genomes 12:1–11. doi: 10.1007/s11295-016-0969-z

Nishiyama S, Yonemori K, Ueda J (2014) Effect of ABA treatment on tannin accumulation at an

early stage of fruit development in persimmon. Acta Hortic 1042:217–222.

470 Pelloux J, Rustérucci C, Mellerowicz EJ (2007) New insights into pectin methylesterase structure and function. Trends Plant Sci 12:267–277. doi: 10.1016/j.tplants.2007.04.001

Saito K, Yonekura-Sakakibara K, Nakabayashi R, Higashi Y, Yamazaki M, Tohge T, Fernie AR (2013) The flavonoid biosynthetic pathway in *Arabidopsis*: Structural and genetic diversity. Plant Physiol Biochem 72:21–34. doi: 10.1016/j.plaphy.2013.02.001

475 Sato A, Yamada M (2016) Persimmon breeding in Japan for pollination-constant non-astringent (PCNA) type with marker-assisted selection. Breed Sci 66:60–68. doi: 10.1270/jsbbs.66.60

Shi J, Kim KN, Ritz O, Albrecht V, Gupta R, Harter K, Luan S, Kudla J (1999) Novel protein kinases associated with calcineurin B-like calcium sensors in *Arabidopsis*. Plant Cell 11:2393–2405. doi: 10.1105/tpc.11.12.2393

480 Su F, Hu J, Zhang Q, Luo Z (2012) Isolation and characterization of a basic Helix-Loop-Helix transcription factor gene potentially involved in proanthocyanidin biosynthesis regulation in persimmon (*Diospyros kaki* Thunb.). Sci Hortic 136:115–121. doi: 10.1016/j.scienta.2012.01.013

Taira S, Matsumto N, Ono M (1998) Accumulation of soluble and insoluble tannins during fruit
485 development in nonastringent and astringent persimmon. J Japanese Soc Hortic Sci 67:572–576. doi: 10.1248/cpb.37.3229

Terrier N, Torregrosa L, Ageorges A, Vialet S, Verries C, Cheynier V, Romieu C (2008) Ectopic Expression of VvMybPA2 promotes proanthocyanidin biosynthesis in grapevine and suggest additional targets in the pathway. Plant Physiol 149:1028–1041. doi: 10.1104/pp.108.131862

490 Wan CY, Wilkins T a (1994) A modified hot borate method significantly enhances the yield of high-quality RNA from cotton (*Gossypium hirsutum* L.). Anal. Biochem. 223:7–12.

Weinl S, Kudla J (2009) The CBL–CIPK Ca^{2+}-decoding signaling network: function and perspectives. New Phytol 184:517–528. doi: 10.1111/j.1469-8137.2009.02938.x

Xu W, Dubos C, Lepiniec L (2015) Transcriptional control of flavonoid biosynthesis by
495 MYB-bHLH-WDR complexes. Trends Plant Sci 20:176–185. doi: 10.1016/j.tplants.2014.12.001

Yamada M, Sato A (2002) Segregation for fruit astringency type in progenies derived from crosses of 'Nishimurawase' x pollination constant non-astringent genotypes in oriental persimmon (*Diospyros kaki* Thunb.). Sci Hortic 92:107–111. doi: 10.1016/S0304-4238(01)00285-0

500 Yonemori K, Matsushima J (1985) Property of development of tannin cells in non-astringent type fruits of Japanese persimmon (*Diospyros kaki*) and its relationship to natural deastringency. J Japanese Soc Hortic Sci 54:201–208

Yonemori K, Sugiura A, Yamada M (2000) Persimmon genetics and breeding. Plant Breed Rev 19:191-225

505 Young MD, Wakefield MJ, Smyth GK, Oshlack A (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol 11:R14. doi: 10.1186/gb-2010-11-2-r14

Zoratti L, Karppinen K, Luengo Escobar A, Häggman H, Jaakola L (2014) Light-controlled flavonoid biosynthesis in fruits. Front Plant Sci 5:534. doi: 10.3389/fpls.2014.00534

510

Figure legends

Fig. 1 Proanthocyanidin (PA) content in a population segregating for the pollination-constant and non-astringent (PCNA) trait. Twelve offspring (6 PCNAs and 6 non-PCNAs) of a cross 'Taiten' (non-PCNA) x 'Kanshu' (PCNA) and their parents, 'Taiten' (Tt) and 'Kanshu' (Ks), were
515 investigated. Line colors indicate the *ASTRINGENCY* (*AST*) (Green) and *ast* (red) genotype of each individual. **a** Patterns of PA concentration in fruit. **b** PA accumulation per day. Error bar indicates standard deviation of three technical replicates.

Fig. 2 Hierarchical clustering of genes differentially expressed between PCNA and non-PCNA
520 individuals. **a** Heat map of clustered differentially expressed genes. The dendrogram was constructed using the average-linkage method from a data matrix of mean reads per kilobase of exon per million reads (RPKM) values of genotypes (PCNA:P/non-PCNA:NP) and stages (June/July). Genes were partitioned into clusters at the red line in the tree. **b** Expression patterns of genes assigned to each cluster. Z-scores were calculated from the data matrix of mean RPKM

525

Fig. 3 Module detection using the WGCNA package. **a** The number of genes in 25 coexpression modules. **b** Heat map of the correlations between module eigengene of detected modules and PA content. Values in the upper and lower side of each cell indicate *R* and *p* values, respectively

530 **Fig. 4** Coexpression network of genes assigned to the magenta module. A topological overlap measure (TOM) threshold of coexpression was set at 0.27 according to the distribution of the number of edges for each node. Genes with more than 50 edges, transcription factors, and calcineurin B-like proteins (CBL)-interacting protein kinase (CIPKs) are shown with their annotations and the number of edges. The color of each square indicates functional classification
535 (shown in Table S5); blue, CIPKs; red, transcription factors; yellow, enzymes/transporters in the flavonoid pathway; green, other chloroplastic genes; orange, other oxidoreductases; light blue, other transferases; and black, functionally unknown/others. The edges connected with transcription factors are shown with red line

540 **Table 1** Summary of the assembly and functional annotation of 'Taiten' persimmon flesh

transcriptome.

Table 2 Gene ontology (GO) enrichment analysis of differentially expressed genes assigned to cluster 2

Fig. S1 PCR amplification of *AST* marker in the siblings segregating for the non-astringent trait. A multiplexed primer set AST-F/PCNA-F/5R3R was used to genotype *AST* (Kanzaki et al, 2010). A black arrow indicates non-PCNA specific amplification.

Fig. S2 Box plots indicating expression levels of putative regulatory genes assigned to cluster 2 for each time point and genotype. The ends of the whiskers indicate maximum and minimum values of RPKM. Top of the box, the band inside the box, and the bottom of the box indicate the first, second, and third quantiles, respectively. NP: non-PCNA. P: PCNA

Table S1 GO analysis of the differentially expressed genes assigned to cluster 4

Table S2 GO analysis of the differentially expressed genes assigned to cluster 6

Table S3 GO analysis of genes assigned to the magenta module

Table S4 GO analysis of genes assigned to modules negatively correlated with PA traits

Table S5 GO analysis if genes assigned to modules positively correlated with PA traits

Table S6 Functional classification and edge counts of genes for network visualization of the magenta module

Table S7 Top 100 genes for connectivity in the magenta module

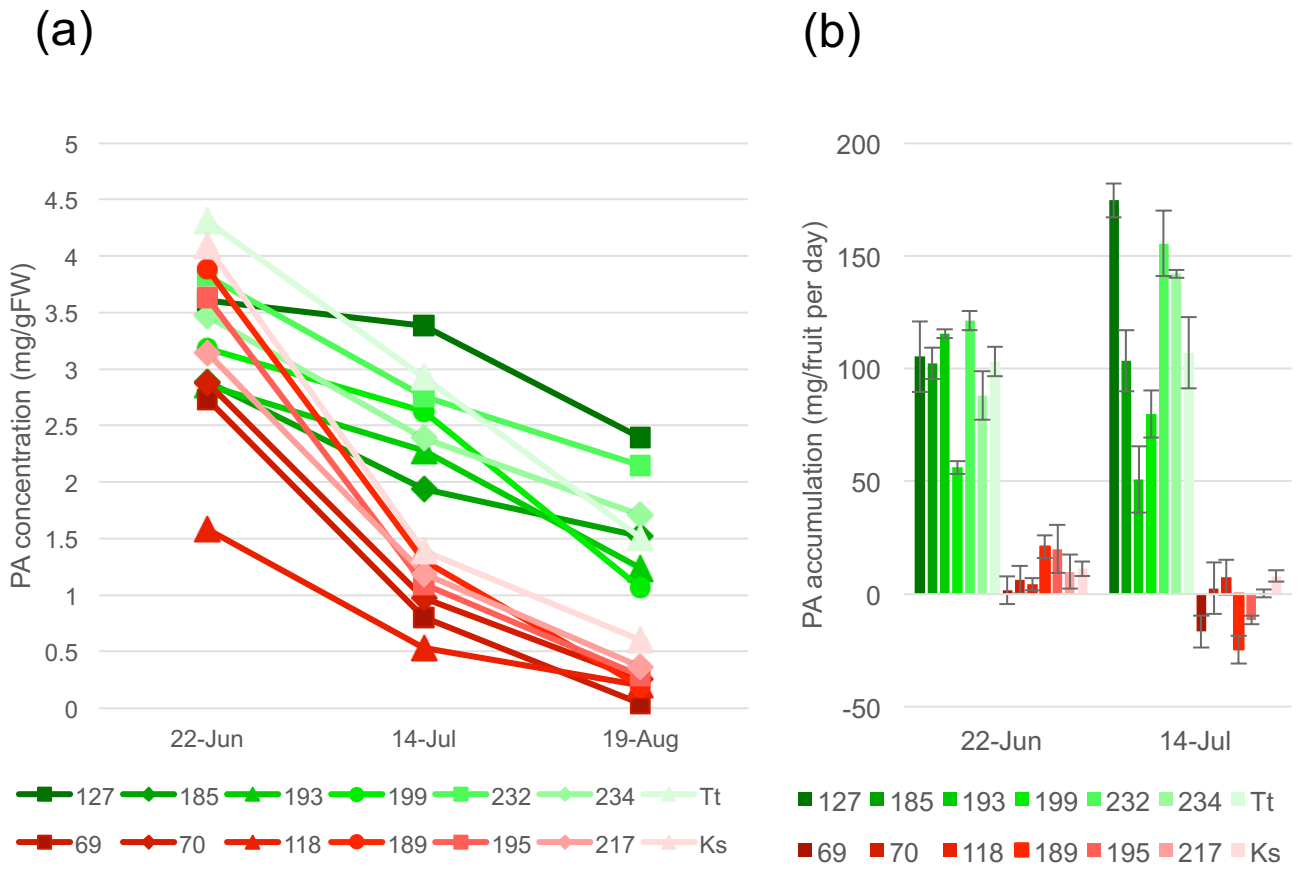

Fig. 1 Proanthocyanidin (PA) content in a population segregating for the pollination-constant and non-astringent (PCNA) trait. Twelve offspring (6 PCNAs and 6 non-PCNAs) of a cross 'Taiten' (non-PCNA) x 'Kanshu' (PCNA) and their parents, 'Taiten' (Tt) and 'Kanshu' (Ks), were investigated. Line colors indicate the *ASTRINGENCY* (*AST*) (Green) and *ast* (red) genotype of each individual. (a) Patterns of PA concentration in fruit. (b) PA accumulation per day. Error bar indicates standard deviation of three technical replicates.

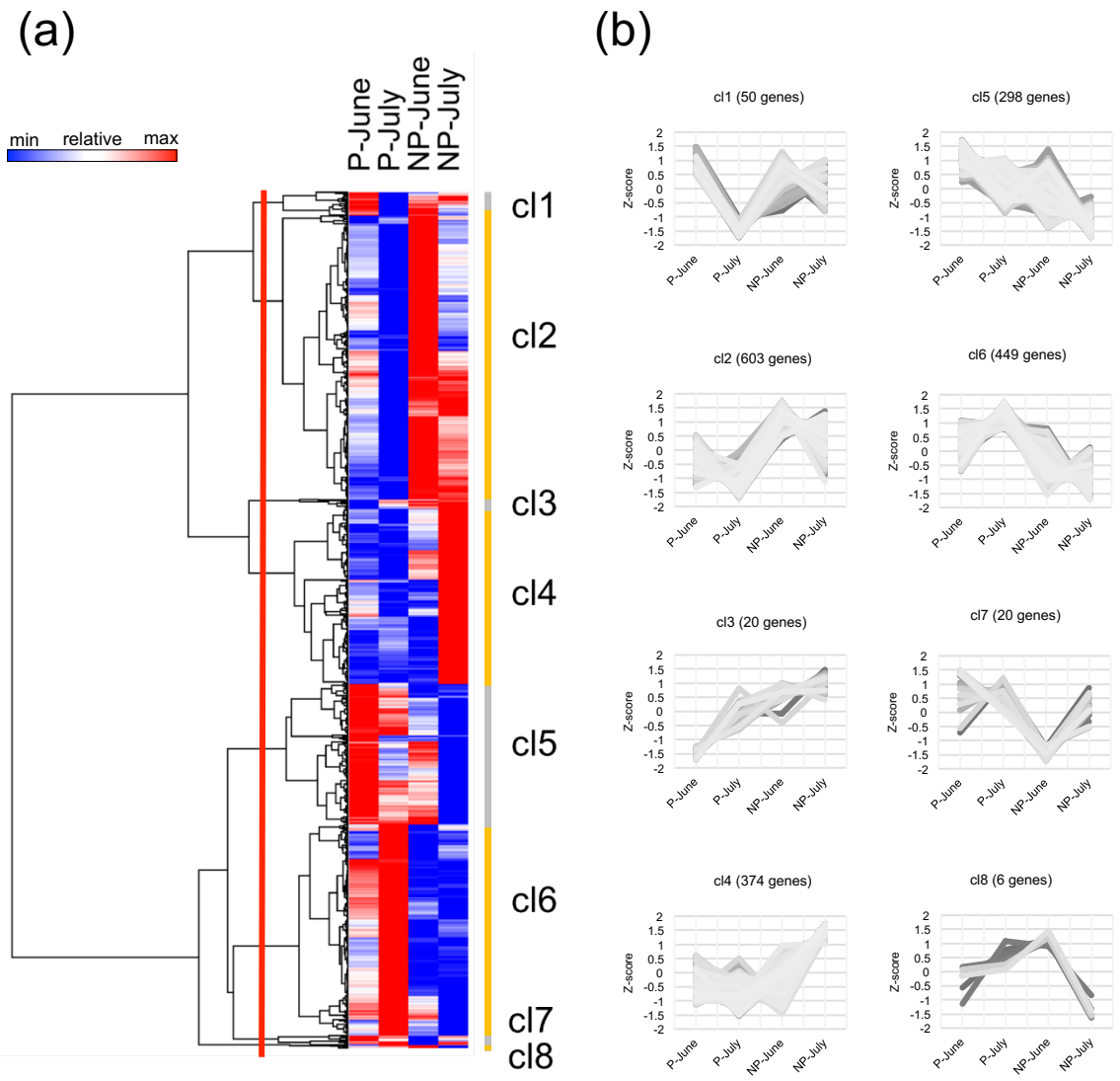

Fig. 2 Hierarchical clustering of genes differentially expressed between PCNA and non-PCNA individuals. (a) Heat map of clustered differentially expressed genes. The dendrogram was constructed using the average-linkage method from a data matrix of mean reads per kilobase of exon per million reads (RPKM) values of genotypes (PCNA:P/non-PCNA:NP) and stages (June/July). Genes were partitioned into clusters at the red line in the tree. (b) Expression patterns of genes assigned to each cluster. Z-scores were calculated from the data matrix of mean RPKM.

(a)

module	numGenes[a]
bisque4	235
black	1071
blue	6943
brown	2375
brown4	97
cyan	514
darkgreen	326
darkgrey	581
darkolivegreen	1185
darkorange	4174
darkorange2	515
darkred	332
greenyellow	673
grey	85
grey60	973
lightcyan	712
magenta	842
midnightblue	760
plum1	146
plum2	91
purple	709
red	1303
sienna3	200
tan	628
thistle1	73

[a] Number of genes assigned to the module.

(b)

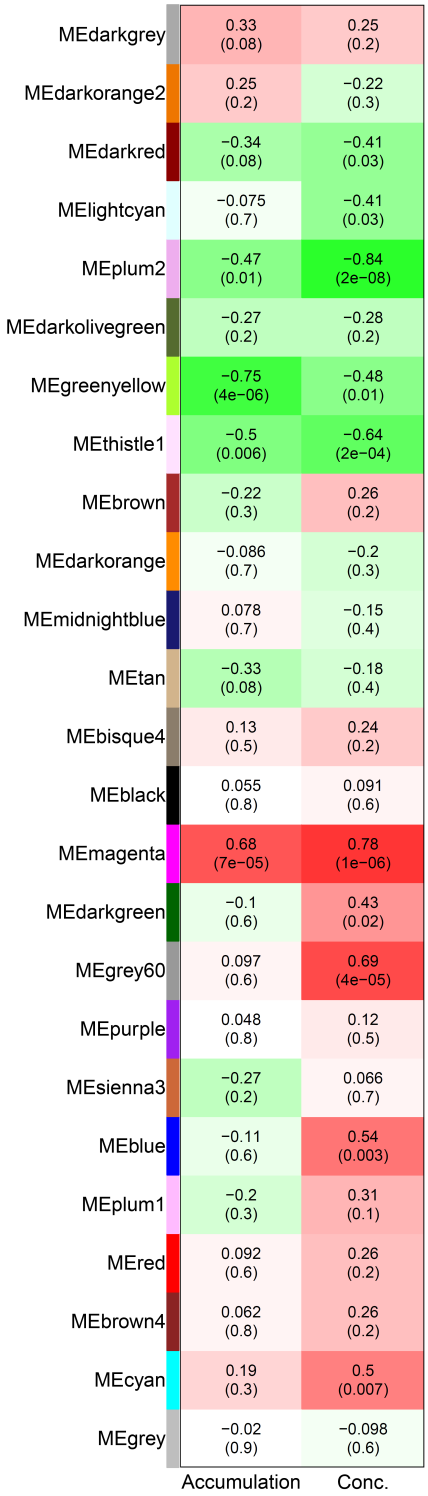

Fig. 3 Module detection using the WGCNA package. (a) The number of genes in 25 coexpression modules. (b) Heat map of the correlations between module eigengene of detected modules and PA content. Values in the upper and lower side of each cell indicate R and p values, respectively.

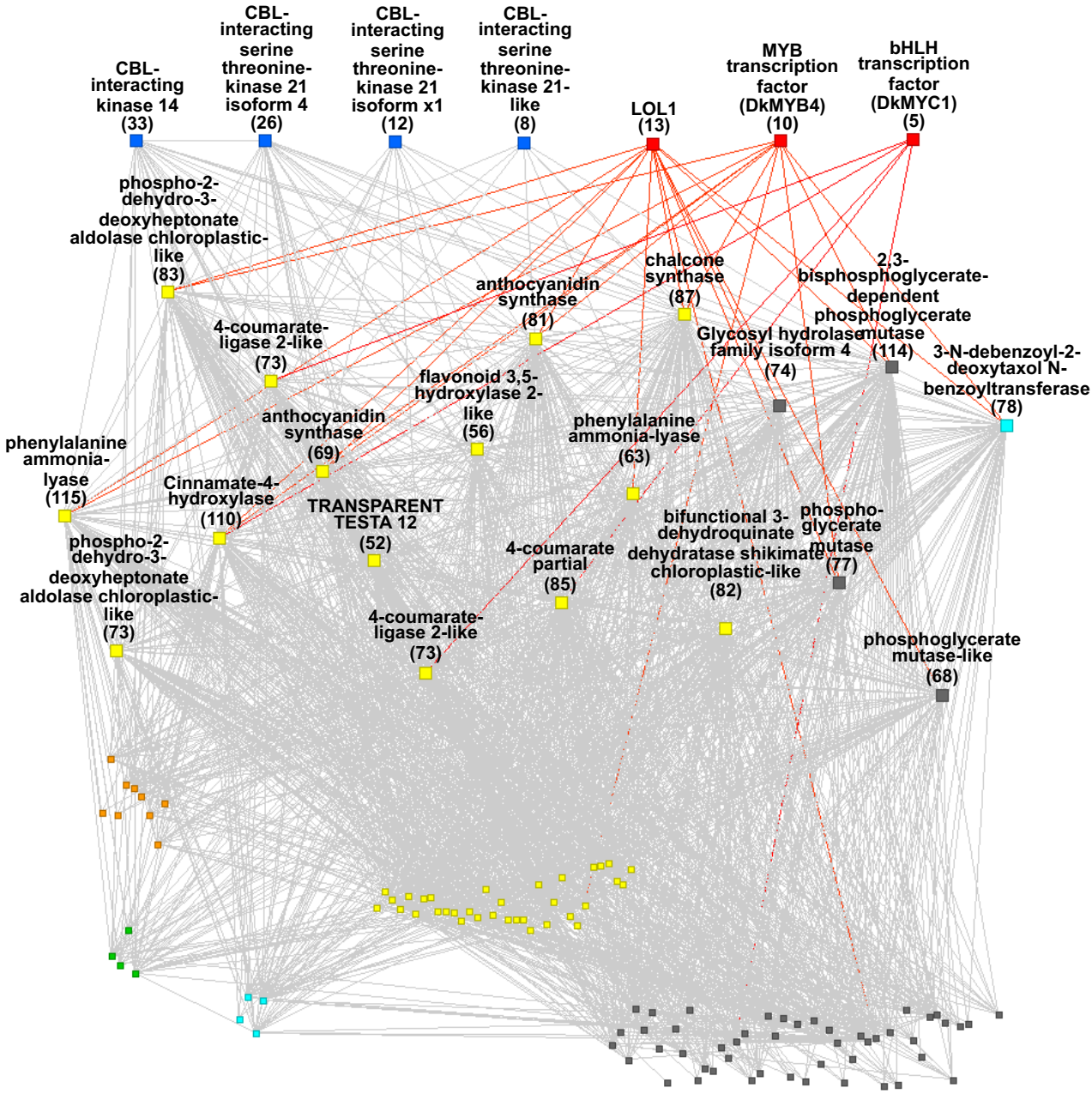

Fig. 4 Coexpression network of genes assigned to the magenta module. A topological overlap measure (TOM) threshold of coexpression was set at 0.27 according to the distribution of the number of edges for each node. Genes with more than 50 edges, transcription factors, and calcineurin B-like proteins (CBL)-interacting protein kinase (CIPKs) are shown with their annotations and the number of edges. The color of each square indicates functional classification (shown in Table S5); blue, CIPKs; red, transcription factors; yellow, enzymes/ transporters in the flavonoid pathway; green, other chloroplastic genes; orange, other oxidoreductases; light blue, other transferases; and black, functionally unknown/others. The edges connected with transcription factors are shown with red line.

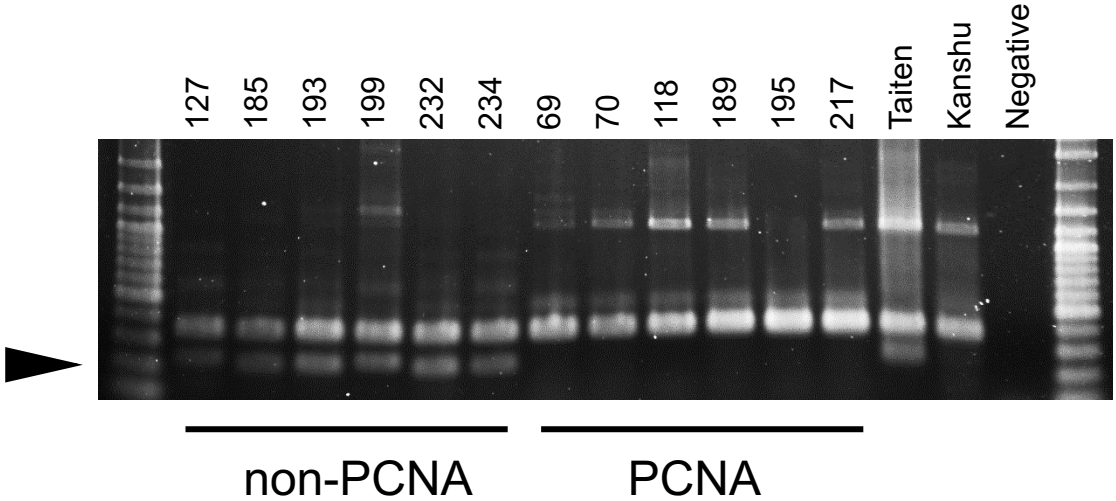

Fig. S1 PCR amplification of *AST* marker in the siblings segregating for the non-astringent trait. A multiplexed primer set AST-F/PCNA-F/5R3R was used to genotype *AST* (Kanzaki et al, 2010). A black arrow indicates non-PCNA specific amplification.

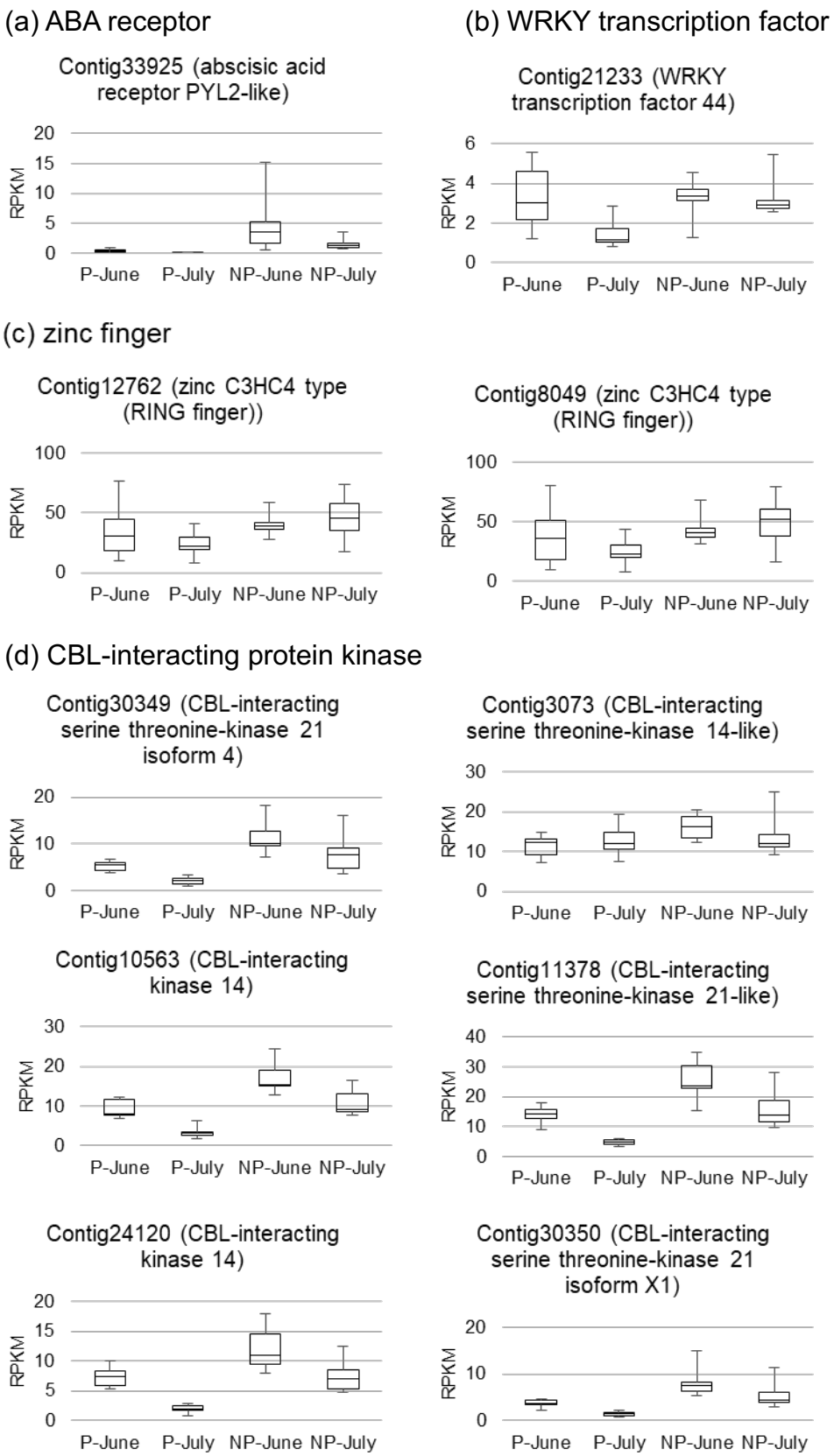

Fig. S2 Box plots indicating expression levels of putative regulatory genes assigned to cluster 2 for each time point and genotype. The ends of the whiskers indicate maximum and minimum values of RPKM. Top of the box, the band inside the box, and the bottom of the box indicate the first, second, and third quantiles, respectively. NP: non-PCNA. P: PCNA.

Table 1 Summary of the assembly and functional annotation of ‘Taiten’ persimmon flesh transcriptome.

assembly	
maximum	16620
minimum	67
average	580
N25	380
N50	726
N75	1526
count	103306
annotation	
with GO term	32591
blast hits without GO term	6154
without blast hits	64561

Table 2 Gene ontology (GO) enrichment analysis of differentially expressed genes assigned to cluster 2

	category	term	ontology[a]	numDEInCat[b]	numInCat[c]	pvalue	FDR
flavonoids pathway	GO:0009800	cinnamic acid biosynthetic process	BP	8	8	0	0
	GO:0045548	phenylalanine ammonia-lyase activity	MF	8	8	0	0
	GO:0009073	aromatic amino acid family biosynthetic process	BP	17	29	5.6.E-26	9.7.E-23
	GO:0003849	3-deoxy-7-phosphoheptulonate synthase activity	MF	9	12	3.7.E-16	2.1.E-13
	GO:0006559	L-phenylalanine catabolic process	BP	8	9	7.8.E-15	4.0.E-12
	GO:0009423	chorismate biosynthetic process	BP	8	10	8.9.E-15	4.2.E-12
	GO:0009718	anthocyanin-containing compound biosynthetic process	BP	7	11	1.4.E-11	5.0.E-09
	GO:0004764	shikimate 3-dehydrogenase (NADP+) activity	MF	6	14	4.3.E-09	1.2.E-06
	GO:0003855	3-dehydroquinate dehydratase activity	MF	6	14	4.3.E-09	1.2.E-06
	GO:0019632	shikimate metabolic process	BP	6	14	6.3.E-09	1.5.E-06
	GO:0045552	dihydrokaempferol 4-reductase activity	MF	5	14	3.1.E-07	5.8.E-05
	GO:0009813	flavonoid biosynthetic process	BP	8	95	2.8.E-05	3.4.E-03
	GO:0044550	secondary metabolite biosynthetic process	BP	6	63	1.5.E-04	1.5.E-02
	GO:0080043	quercetin 3-O-glucosyltransferase activity	MF	6	80	5.3.E-04	4.1.E-02
transport	GO:0016747	transferase activity, transferring acyl groups other than amino-acyl groups	MF	6	39	9.2.E-06	1.3.E-03
	GO:0015238	drug transmembrane transporter activity	MF	6	41	1.1.E-05	1.5.E-03
	GO:0015297	antiporter activity	MF	6	59	7.9.E-05	8.5.E-03
	GO:0006855	drug transmembrane transport	BP	6	70	2.6.E-04	2.4.E-02
photosynthesis/chloroplast	GO:0016168	chlorophyll binding	MF	16	28	4.1.E-24	5.3.E-21
	GO:0018298	protein-chromophore linkage	BP	15	31	7.2.E-21	7.4.E-18
	GO:0009522	photosystem I	CC	14	29	8.3.E-20	7.2.E-17
	GO:0009768	photosynthesis, light harvesting in photosystem I	BP	11	15	9.0.E-19	6.7.E-16
	GO:0031409	pigment binding	MF	10	14	5.0.E-17	3.2.E-14
	GO:0010287	plastoglobule	CC	12	47	7.2.E-13	2.9.E-10
	GO:0009535	chloroplast thylakoid membrane	CC	18	187	3.1.E-11	1.1.E-08
	GO:0009416	response to light stimulus	BP	10	42	5.7.E-11	1.8.E-08
	GO:0009941	chloroplast envelope	CC	13	322	3.2.E-04	2.9.E-02
	GO:0015979	photosynthesis	BP	6	90	7.6.E-04	5.8.E-02
others	GO:0055114	oxidation-reduction process	BP	58	1640	3.2.E-13	1.4.E-10
	GO:0016829	lyase activity	MF	11	81	3.1.E-09	9.3.E-07
	GO:0050662	coenzyme binding	MF	10	73	2.2.E-08	5.2.E-06
	GO:0008422	beta-glucosidase activity	MF	6	31	2.3.E-06	3.6.E-04
	GO:0006096	glycolytic process	BP	7	58	1.5.E-05	1.9.E-03
	GO:0006633	fatty acid biosynthetic process	BP	7	65	1.8.E-05	2.3.E-03
	GO:0046872	metal ion binding	MF	30	1165	1.1.E-04	1.2.E-02
	GO:0016788	hydrolase activity, acting on ester bonds	MF	6	66	1.7.E-04	1.7.E-02
	GO:0020037	heme binding	MF	10	211	3.4.E-04	3.0.E-02
	GO:0042802	identical protein binding	MF	6	74	3.4.E-04	3.0.E-02
	GO:0051287	NAD binding	MF	7	105	4.0.E-04	3.5.E-02
	GO:0050661	NADP binding	MF	6	82	4.4.E-04	3.7.E-02
	GO:0005506	iron ion binding	MF	10	219	4.7.E-04	3.8.E-02

[a] category of the ontology. BP:Biological Process, MF:Molecular Function, CC:Cellular Component

[b] The number of differentially expressed genes with given GO term.

[c] Total number of genes with given GO term in the reference.

Table S1 GO analysis of the differentially expressed genes assigned to cluster 4

category	term	ontology[a]	numDEInCat[b]	numInCat[c]	pvalue	FDR
GO:0006537	glutamate biosynthetic process	BP	8	9	0	0
GO:0016040	glutamate synthase (NADH) activity	MF	8	8	0	0
GO:0051536	iron-sulfur cluster binding	MF	8	35	0	0
GO:0010181	FMN binding	MF	8	65	1.2.E-10	1.6.E-07
GO:0048589	developmental growth	BP	5	8	9.3.E-10	9.6.E-07
GO:0019676	ammonia assimilation cycle	BP	5	13	2.9.E-09	2.5.E-06
GO:0055114	oxidation-reduction process	BP	25	1640	4.4.E-09	3.3.E-06
GO:0050660	flavin adenine dinucleotide binding	MF	8	101	6.5.E-09	4.2.E-06
GO:0005506	iron ion binding	MF	10	219	1.2.E-08	6.9.E-06
GO:0009735	response to cytokinin	BP	6	76	4.6.E-07	2.4.E-04
GO:0009507	chloroplast	CC	13	802	7.6.E-05	2.8.E-02

[a] category of the ontology. BP:Biological Process, MF:Molecular Function, CC:Cellular Component

[b] The number of differentially expressed genes with given GO term.

[c] Total number of genes with given GO term in the reference.

Table S2 GO analysis of the differentially expressed genes assigned to cluster 6

category	term	ontology[a]	numDEInCat[b]	numInCat[c]	pvalue	FDR
GO:0030422	production of siRNA involved in RNA interference	BP	11	36	4.4.E-14	2.3.E-10
GO:0016442	RISC complex	CC	9	24	2.8.E-13	7.2.E-10
GO:0004525	ribonuclease III activity	MF	10	35	5.1.E-13	8.8.E-10
GO:0047134	protein-disulfide reductase activity	MF	6	9	3.7.E-11	4.8.E-08
GO:0001172	transcription, RNA-templated	BP	6	13	2.9.E-09	2.5.E-06
GO:0003968	RNA-directed RNA polymerase activity	MF	6	13	2.9.E-09	2.5.E-06
GO:0090502	RNA phosphodiester bond hydrolysis, endonucleolytic	BP	10	92	1.9.E-08	1.4.E-05
GO:0003993	acid phosphatase activity	MF	7	44	2.5.E-07	1.6.E-04
GO:0005623	cell	CC	7	46	3.9.E-07	2.2.E-04
GO:0019538	protein metabolic process	BP	6	42	1.4.E-06	7.5.E-04
GO:0045454	cell redox homeostasis	BP	9	114	2.4.E-06	1.1.E-03
GO:0031624	ubiquitin conjugating enzyme binding	MF	5	26	5.4.E-06	2.2.E-03
GO:0032436	positive regulation of proteasomal ubiquitin-dependent protein catabolic process	BP	5	25	5.6.E-06	2.2.E-03
GO:0000151	ubiquitin ligase complex	CC	5	39	4.2.E-05	1.4.E-02
GO:0055085	transmembrane transport	BP	13	369	1.5.E-04	4.1.E-02

[a] category of the ontology. BP:Biological Process, MF:Molecular Function, CC:Cellular Component

[b] The number of differentially expressed genes with given GO term.

[c] Total number of genes with given GO term in the reference.

Table S3 GO analysis of genes assigned to the magenta module

category	term	ontology[a]	numDEInCat[b]	numInCat[c]	pvalue	FDR
GO:0009800	cinnamic acid biosynthetic process	BP	8	8	0	0
GO:0045548	phenylalanine ammonia-lyase activity	MF	8	8	0	0
GO:0016168	chlorophyll binding	MF	14	28	3.2.E-20	5.5.E-17
GO:0009522	photosystem I	CC	13	29	5.1.E-18	6.4.E-15
GO:0018298	protein-chromophore linkage	BP	13	31	6.2.E-18	6.4.E-15
GO:0009768	photosynthesis, light harvesting in photosystem I	BP	10	15	2.3.E-16	2.0.E-13
GO:0006559	L-phenylalanine catabolic process	BP	8	9	2.0.E-15	1.4.E-12
GO:0003849	3-deoxy-7-phosphoheptulonate synthase activity	MF	9	12	2.2.E-15	1.4.E-12
GO:0031409	pigment binding	MF	9	14	1.3.E-14	7.4.E-12
GO:0003855	3-dehydroquinate dehydratase activity	MF	9	14	1.6.E-14	7.5.E-12
GO:0004764	shikimate 3-dehydrogenase (NADP+) activity	MF	9	14	1.6.E-14	7.5.E-12
GO:0009073	aromatic amino acid family biosynthetic process	BP	11	29	3.1.E-14	1.3.E-11
GO:0009718	anthocyanin-containing compound biosynthetic process	BP	8	11	7.2.E-14	2.9.E-11
GO:0055114	oxidation-reduction process	BP	56	1640	9.7.E-13	3.6.E-10
GO:0010287	plastoglobule	CC	11	47	6.7.E-12	2.3.E-09
GO:0019632	shikimate metabolic process	BP	7	14	1.5.E-10	5.0.E-08
GO:0016747	transferase activity, transferring acyl groups other than amino-acyl groups	MF	9	39	7.4.E-10	2.3.E-07
GO:0009416	response to light stimulus	BP	9	42	2.0.E-09	5.8.E-07
GO:0045552	dihydrokaempferol 4-reductase activity	MF	6	14	1.3.E-08	3.4.E-06
GO:0050662	coenzyme binding	MF	10	73	2.0.E-08	4.9.E-06
GO:0009535	chloroplast thylakoid membrane	CC	14	187	6.6.E-08	1.5.E-05
GO:0006633	fatty acid biosynthetic process	BP	9	65	7.3.E-08	1.6.E-05
GO:0016829	lyase activity	MF	8	81	8.3.E-06	1.4.E-03
GO:0015238	drug transmembrane transporter activity	MF	6	41	9.8.E-06	1.6.E-03
GO:0009813	flavonoid biosynthetic process	BP	8	95	2.1.E-05	3.4.E-03
GO:0020037	heme binding	MF	11	211	5.8.E-05	8.6.E-03
GO:0050661	NADP binding	MF	7	82	7.2.E-05	1.0.E-02
GO:0015297	antiporter activity	MF	6	59	8.9.E-05	1.2.E-02
GO:0046872	metal ion binding	MF	30	1165	9.2.E-05	1.2.E-02
GO:0006855	drug transmembrane transport	BP	6	70	1.8.E-04	2.2.E-02
GO:0003993	acid phosphatase activity	MF	5	44	1.9.E-04	2.4.E-02
GO:0080043	quercetin 3-O-glucosyltransferase activity	MF	6	80	4.5.E-04	4.9.E-02
GO:0009941	chloroplast envelope	CC	12	322	5.1.E-04	5.1.E-02
GO:0006096	glycolytic process	BP	5	58	5.1.E-04	5.1.E-02
GO:0044550	secondary metabolite biosynthetic process	BP	5	63	1.0.E-03	9.6.E-02

[a] category of the ontology. BP:Biological Process, MF:Molecular Function, CC:Cellular Component

[b] The number of differentially expressed genes with given GO term.

[c] Total number of genes with given GO term in the reference.

Table S4 GO analysis of genes assigned to modules negatively correlated with PA traits

category	term	ontology[a]	numDEInCat[b]	numInCat[c]	pvalue	FDR
module:plum2						
GO:0004674	protein serine/threonine kinase activity	MF	7	632	7.4.E-05	6.3.E-02
module:greenyellow						
GO:0030422	production of siRNA involved in RNA interfe	BP	11	36	3.7.E-15	1.9.E-11
GO:0016442	RISC complex	CC	9	24	5.2.E-13	1.4.E-09
GO:0004525	ribonuclease III activity	MF	9	35	1.8.E-11	3.1.E-08
GO:0047134	protein-disulfide reductase activity	MF	6	9	5.3.E-11	6.8.E-08
GO:0001172	transcription, RNA-templated	BP	5	13	1.6.E-08	1.4.E-05
GO:0003968	RNA-directed RNA polymerase activity	MF	5	13	1.6.E-08	1.4.E-05
GO:0005623	cell	CC	7	46	9.2.E-08	6.8.E-05
GO:0090502	RNA phosphodiester bond hydrolysis, endonu	BP	9	92	1.1.E-07	7.1.E-05
GO:0045454	cell redox homeostasis	BP	9	114	4.8.E-07	2.8.E-04
GO:0005737	cytoplasm	CC	23	1126	3.2.E-05	1.5.E-02
GO:0052696	flavonoid glucuronidation	BP	6	88	8.3.E-05	2.5.E-02
GO:0009813	flavonoid biosynthetic process	BP	6	95	1.3.E-04	3.5.E-02
GO:0080044	quercetin 7-O-glucosyltransferase activity	MF	5	79	4.9.E-04	9.8.E-02
module:darkred						
GO:0004568	chitinase activity	MF	7	17	0	0
GO:0006032	chitin catabolic process	BP	7	17	0	0
GO:0008061	chitin binding	MF	7	16	0	0
GO:0000272	polysaccharide catabolic process	BP	6	19	5.3.E-10	6.9.E-07
module:thistle1						
GO:0046910	pectinesterase inhibitor activity	MF	5	19	2.1.E-11	1.1.E-07
GO:0030599	pectinesterase activity	MF	5	33	4.5.E-10	1.2.E-06
GO:0043086	negative regulation of catalytic activity	BP	5	55	6.5.E-09	1.1.E-05
GO:0005618	cell wall	CC	5	201	4.3.E-06	3.2.E-03

[a] category of the ontology. BP:Biological Process, MF:Molecular Function, CC:Cellular Component

[b] The number of differentially expressed genes with given GO term.

[c] Total number of genes with given GO term in the reference.

Table S5 GO analysis of genes assigned to modules positively correlated with PA traits

category	term	ontology[a]	numDEInCat[b]	numInCat[c]	pvalue	FDR
module:grey60						
GO:0000786	nucleosome	CC	14	44	7.4.E-16	3.8.E-12
GO:0046982	protein heterodimerization activity	MF	13	76	1.9.E-11	4.8.E-08
GO:0015979	photosynthesis	BP	12	90	1.4.E-09	2.3.E-06
GO:0009535	chloroplast thylakoid membrane	CC	14	187	2.2.E-08	2.8.E-05
GO:0009522	photosystem I	CC	5	29	2.2.E-05	1.4.E-02
GO:0006334	nucleosome assembly	BP	5	32	5.7.E-05	2.7.E-02
GO:0009416	response to light stimulus	BP	5	42	1.7.E-04	4.6.E-02
GO:0016760	cellulose synthase (UDP-forming) activity	MF	5	48	2.7.E-04	6.9.E-02
module:blue						
GO:0008017	microtubule binding	MF	46	125	8.7.E-24	4.5.E-20
GO:0003777	microtubule motor activity	MF	39	89	6.5.E-23	1.7.E-19
GO:0007018	microtubule-based movement	BP	42	109	1.7.E-22	3.0.E-19
GO:0055114	oxidation-reduction process	BP	202	1640	2.1.E-17	2.7.E-14
GO:0005874	microtubule	CC	41	141	3.1.E-17	2.8.E-14
GO:0005871	kinesin complex	CC	30	76	3.2.E-17	2.8.E-14
GO:0006468	protein phosphorylation	BP	109	879	1.4.E-11	1.0.E-08
GO:0005524	ATP binding	MF	250	2702	1.2.E-10	8.0.E-08
GO:0032440	2-alkenal reductase [NAD(P)] activity	MF	36	183	1.4.E-10	8.2.E-08
GO:0016722	oxidoreductase activity, oxidizing metal ions	MF	15	33	2.0.E-10	1.0.E-07
GO:0005618	cell wall	CC	39	201	6.3.E-10	3.0.E-07
GO:0004672	protein kinase activity	MF	61	437	3.5.E-09	1.5.E-06
GO:0016887	ATPase activity	MF	34	169	1.8.E-08	7.1.E-06
GO:0016021	integral component of membrane	CC	488	5890	2.5.E-08	9.4.E-06
GO:0030570	pectate lyase activity	MF	9	16	3.3.E-07	1.1.E-04
GO:0051753	mannan synthase activity	MF	7	9	7.0.E-07	2.3.E-04
GO:0009505	plant-type cell wall	CC	27	146	1.3.E-06	4.0.E-04
GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds	MF	27	146	1.9.E-06	5.5.E-04
GO:0006855	drug transmembrane transport	BP	17	70	2.1.E-06	5.6.E-04
GO:0045490	pectin catabolic process	BP	13	43	3.6.E-06	9.4.E-04
GO:0031408	oxylipin biosynthetic process	BP	8	18	7.7.E-06	1.9.E-03
GO:0000796	condensin complex	CC	5	6	1.4.E-05	3.2.E-03
GO:0030247	polysaccharide binding	MF	17	75	1.5.E-05	3.4.E-03
GO:0016702	oxidoreductase activity, acting on single donors with inc	MF	8	21	1.6.E-05	3.4.E-03
GO:0007076	mitotic chromosome condensation	BP	6	9	1.6.E-05	3.4.E-03
GO:0042546	cell wall biogenesis	BP	10	28	2.5.E-05	5.0.E-03
GO:0016788	hydrolase activity, acting on ester bonds	MF	15	66	3.5.E-05	6.6.E-03
GO:0008152	metabolic process	BP	190	2195	3.6.E-05	6.7.E-03
GO:0006270	DNA replication initiation	BP	6	11	3.9.E-05	6.9.E-03
GO:0048046	apoplast	CC	24	140	4.3.E-05	7.1.E-03
GO:0000910	cytokinesis	BP	5	8	4.4.E-05	7.1.E-03
GO:0008447	L-ascorbate oxidase activity	MF	5	10	4.4.E-05	7.1.E-03
GO:0071555	cell wall organization	BP	21	121	4.9.E-05	7.6.E-03
GO:0009809	lignin biosynthetic process	BP	8	22	5.4.E-05	7.8.E-03
GO:0005975	carbohydrate metabolic process	BP	42	331	5.4.E-05	7.8.E-03
GO:0045550	geranylgeranyl reductase activity	MF	5	7	6.5.E-05	9.1.E-03
GO:0005576	extracellular region	CC	26	160	8.1.E-05	1.1.E-02
GO:0006542	glutamine biosynthetic process	BP	5	8	8.2.E-05	1.1.E-02
GO:0046658	anchored component of plasma membrane	CC	17	90	9.6.E-05	1.2.E-02
GO:0010311	lateral root formation	BP	6	16	1.0.E-04	1.3.E-02
GO:0004560	alpha-L-fucosidase activity	MF	5	8	1.4.E-04	1.7.E-02
GO:0009524	phragmoplast	CC	9	38	1.5.E-04	1.7.E-02
GO:0051276	chromosome organization	BP	7	19	1.6.E-04	1.8.E-02
GO:0004356	glutamate-ammonia ligase activity	MF	5	9	2.2.E-04	2.4.E-02
GO:0016762	xyloglucan:xyloglucosyl transferase activity	MF	8	24	2.3.E-04	2.5.E-02
GO:0015995	chlorophyll biosynthetic process	BP	8	27	2.4.E-04	2.6.E-02
GO:0016491	oxidoreductase activity	MF	44	364	2.5.E-04	2.6.E-02
GO:0000226	microtubule cytoskeleton organization	BP	5	13	2.6.E-04	2.7.E-02
GO:0010411	xyloglucan metabolic process	BP	8	25	2.7.E-04	2.7.E-02
GO:0016628	oxidoreductase activity, acting on the CH-CH group of d	MF	5	9	3.2.E-04	3.2.E-02
GO:0016760	cellulose synthase (UDP-forming) activity	MF	11	48	3.4.E-04	3.3.E-02
GO:0015238	drug transmembrane transporter activity	MF	10	41	3.6.E-04	3.3.E-02
GO:0045735	nutrient reservoir activity	MF	6	15	6.1.E-04	5.2.E-02
GO:0046274	lignin catabolic process	BP	6	16	6.8.E-04	5.4.E-02
GO:0052716	hydroquinone:oxygen oxidoreductase activity	MF	6	16	6.8.E-04	5.4.E-02
GO:0020037	heme binding	MF	27	211	1.2.E-03	9.0.E-02

GO:0009506	plasmodesma	CC	41	386	1.3.E-03	9.6.E-02
module:cyan						
GO:0030422	production of siRNA involved in RNA interference	BP	11	36	4.4.E-14	2.3.E-10
GO:0016442	RISC complex	CC	9	24	2.8.E-13	7.2.E-10
GO:0004525	ribonuclease III activity	MF	10	35	5.1.E-13	8.8.E-10
GO:0047134	protein-disulfide reductase activity	MF	6	9	3.7.E-11	4.8.E-08
GO:0001172	transcription, RNA-templated	BP	6	13	2.9.E-09	2.5.E-06
GO:0003968	RNA-directed RNA polymerase activity	MF	6	13	2.9.E-09	2.5.E-06
GO:0090502	RNA phosphodiester bond hydrolysis, endonucleolytic	BP	10	92	1.9.E-08	1.4.E-05
GO:0003993	acid phosphatase activity	MF	7	44	2.5.E-07	1.6.E-04
GO:0005623	cell	CC	7	46	3.9.E-07	2.2.E-04
GO:0019538	protein metabolic process	BP	6	42	1.4.E-06	7.5.E-04
GO:0045454	cell redox homeostasis	BP	9	114	2.4.E-06	1.1.E-03
GO:0031624	ubiquitin conjugating enzyme binding	MF	5	26	5.4.E-06	2.2.E-03
GO:0032436	positive regulation of proteasomal ubiquitin-dependent p	BP	5	25	5.6.E-06	2.2.E-03
GO:0000151	ubiquitin ligase complex	CC	5	39	4.2.E-05	1.4.E-02
GO:0055085	transmembrane transport	BP	13	369	1.5.E-04	4.1.E-02

GO analysis for magenta module was shown in Table S3.

[a] category of the ontology. BP:Biological Process, MF:Molecular Function, CC:Cellular Component

[b] The number of differentially expressed genes with given GO term.

[c] Total number of genes with given GO term in the reference.

Table S6 Functional classification and edge counts of genes for network visualization of the magenta module

node	Annotation	classification[a]	numEdge[b]
Contig11	phenylalanine ammonia-lyase	F	115
Contig4700	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	U	114
Contig2390	Cinnamate-4-hydroxylase [*Theobroma cacao*]	F	110
Contig16053	chalcone synthase	F	87
Contig8677	4-coumarate partial [*Populus nigra*]	F	85
Contig482	phospho-2-dehydro-3-deoxyheptonate aldolase chloroplastic-like	F	83
Contig4670	bifunctional 3-dehydroquinate dehydratase shikimate chloroplastic-like	F	82
Contig492	anthocyanidin synthase	F	81
Contig5640	3-N-debenzoyl-2-deoxytaxol N-benzoyltransferase	T	78
Contig34711	phosphoglycerate mutase	U	77
Contig11748	Glycosyl hydrolase family isoform 4 [*Theobroma cacao*]	U	74
Contig16195	phospho-2-dehydro-3-deoxyheptonate aldolase chloroplastic-like	F	73
Contig5021	4-coumarate-ligase 2-like	F	73
Contig628	4-coumarate-ligase 2-like	F	73
Contig512	anthocyanidin synthase	F	69
Contig9187	phosphoglycerate mutase-like [*Malus domestica*]	U	68
Contig9491	phenylalanine ammonia-lyase	F	63
Contig966	flavonoid 3,5-hydroxylase 2-like	F	56
Contig11937	TRANSPARENT TESTA 12	F	52
Contig24127	probable acyl-activating enzyme peroxisomal	U	47
Contig11936	TRANSPARENT TESTA 12	F	46
Contig2389	cinnamate 4-partial [*Populus trichocarpa*]	F	46
Contig2957	phospho-2-dehydro-3-deoxyheptonate aldolase chloroplastic-like	F	39
Contig4343	DETOXIFICATION 33	U	39
Contig5729	phenylalanine ammonia-lyase	F	39
Contig3	serine carboxypeptidase-like 18 isoform X1	T	36
Contig4319	limonoid UDP-glucosyltransferase	F	34
Contig10563	CBL-interacting kinase 14 [*Vitis vinifera*]	CK	33
Contig5765	trans-cinnamate 4-monooxygenase	F	33
Contig16617	---NA---	U	32
Contig43179	---NA---	U	31
Contig10633	3-dehydroquinate dehydratase shikimate dehydrogenase 2 [*Camellia sinensis*]	F	29
Contig10634	3-dehydroquinate dehydratase shikimate dehydrogenase 2 [*Camellia sinensis*]	F	29
Contig16293	PREDICTED: uncharacterized protein LOC102616627 isoform X1	CP	29
Contig4832	phospho-2-dehydro-3-deoxyheptonate aldolase chloroplastic-like	F	29
Contig46247	Alpha beta-Hydrolases superfamily isoform 2 [*Theobroma cacao*]	U	28
Contig5761	chalcone synthase	F	28
Contig10671	TRANSPARENT TESTA 12-like	F	27
Contig29114	PREDICTED: uncharacterized protein LOC105646674 [*Jatropha curcas*]	U	27
Contig15066	---NA---	U	26
Contig16840	---NA---	U	26
Contig30349	CBL-interacting serine threonine-kinase 21 isoform 4 [*Theobroma cacao*]	CK	26
Contig3538	serine carboxypeptidase-like 18	T	26
Contig54344	PREDICTED: uncharacterized protein LOC104448399 [*Eucalyptus grandis*]	U	24
Contig7111	4-coumarate-ligase 2-like	F	24
Contig14472	phenylalanine ammonia-lyase	F	23
Contig15352	MATE efflux family [*Theobroma cacao*]	F	23
Contig17020	---NA---	U	23
Contig9715	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase	OR	23
Contig35613	---NA---	U	22
Contig56713	Riboflavin synthase alpha isoform 1 [*Theobroma cacao*]	OR	22
Contig23632	secoisolariciresinol dehydrogenase-like	OR	21
Contig32030	---NA---	U	21
Contig55679	serine carboxypeptidase-like 18	T	21
Contig4831	3-deoxy-d-arabino-heptulosonate 7-phosphate synthase isoform 3 [*Theobroma cacao*]	F	20
Contig16955	hypothetical protein EUGRSUZ_B02529	CP	19
Contig16616	---NA---	U	18
Contig16841	---NA---	U	18
Contig301	bifunctional 3-dehydroquinate dehydratase shikimate chloroplastic-like isoform X1	F	18
Contig60933	---NA---	U	18
Contig64054	---NA---	U	18
Contig17191	---NA---	U	17
Contig63662	cytochrome P450 93A3-like	F	17
Contig10040	1-Cys peroxiredoxin	F	16
Contig15277	beta-D-glucan exohydrolase [*Arabidopsis thaliana*]	U	16
Contig5381	bifunctional 3-dehydroquinate dehydratase shikimate chloroplastic-like	F	16
Contig4830	phospho-2-dehydro-3-deoxyheptonate aldolase chloroplastic-like	F	15
Contig5760	chalcone synthase	F	14
Contig26072	sulfite oxidase isoform X1 [*Ricinus communis*]	OR	13
Contig52008	LOL1 [*Vitis vinifera*]	TF	13
Contig5671	phospho-2-dehydro-3-deoxyheptonate aldolase chloroplastic-like	F	13
Contig765	chalcone synthase	F	13

Contig7962	beta-glucosidase 3B-like	U	13
Contig10070	---NA---	U	12
Contig14773	4-coumarate-ligase 2-like	F	12
Contig15435	anthocyanidin 3-O-glucosyltransferase 2-like	F	12
Contig18636	PREDICTED: uncharacterized protein LOC105961982 [*Erythranthe guttata*]	U	12
Contig30350	CBL-interacting serine threonine-kinase 21 isoform X1	CK	12
Contig313	dihydrofolate reductase-like	OR	12
Contig54	enolase chloroplastic	CP	12
Contig44976	---NA---	U	11
Contig4669	bifunctional 3-dehydroquinate dehydratase shikimate chloroplastic-like	F	11
Contig11166	MYB transcription factor	TF	10
Contig21600	---NA---	U	10
Contig26724	MATE efflux family chloroplastic	F	10
Contig92435	---NA---	U	10
Contig18274	1-Cys peroxiredoxin	OR	9
Contig11289	---NA---	U	8
Contig11378	CBL-interacting serine threonine-kinase 21-like	CK	8
Contig30387	sulfite oxidase isoform X2	OR	8
Contig40760	phenylalanine ammonia-lyase	F	8
Contig40982	---NA---	U	8
Contig54652	---NA---	U	8
Contig64871	binding partner of ACD11 1-like isoform X1	U	8
Contig1851	bifunctional 3-dehydroquinate dehydratase shikimate chloroplastic-like isoform X2	F	7
Contig1852	bifunctional 3-dehydroquinate dehydratase shikimate chloroplastic-like isoform X2	F	7
Contig21975	kaempferol 3-O-beta-D-galactosyltransferase-like	F	7
Contig3236	PREDICTED: uncharacterized protein LOC104219419 [*Nicotiana sylvestris*]	U	7
Contig47825	---NA---	U	7
Contig24911	---NA---	U	6
Contig34524	---NA---	U	6
Contig10860	---NA---	U	5
Contig20409	bHLH transcription factor	TF	5
Contig29758	---NA---	U	5
Contig6422	Arabinanase levansucrase [*Theobroma cacao*]	U	5
Contig67422	sulfite oxidase-like	OR	5
Contig11787	3-ketoacyl-synthase 10	T	4
Contig13545	ankyrin repeat-containing At5g02620-like isoform X2 [*Citrus sinensis*]	U	4
Contig28376	Glycosyl hydrolase family 35 protein isoform 1	U	4
Contig41477	---NA---	U	4
Contig57439	---NA---	U	4
Contig6939	3-hydroxyisobutyryl-hydrolase 5	U	4
Contig7110	4-coumarate: partial [*Glycine max*]	F	4
Contig338	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase	OR	3
Contig36482	---NA---	U	3
Contig37228	---NA---	U	3
Contig5302	cinnamate 4-partial [*Pinus massoniana*]	F	3
Contig7441	chlorophyll a-b binding of LHCII type 1-like	CP	3
Contig78880	---NA---	U	3
Contig8185	---NA---	U	3
Contig30639	dihydroflavonol 4-reductase	F	1
Contig33506	PREDICTED: uncharacterized protein LOC104216650	U	1

Edges in the co-expression network of magenta module were filtered out by topological overlap threshold (> 0.27).

[a] Gene functional classification. CK:CBL-interacting protein kinase. F: Enzemes/transporters involved in flavonoid pathway. TF: Transcription factor. CP: Other chroloplastic genes. T: Other transferases. OR: Other oxidoreductases. U: Unknown/Others.

[b] The number of edges retained after the topological overlap cutoff.

Table S7 Top 100 genes for connectivity in the magenta module

Contig	annotation	accumulation		conc.		module		[a]connectivity	RPKM			
		[b]cor.	P	[c]cor.	P	[d]MM	P		[e]P-June	[e]P-July	[e]NP-June	[e]NP-July
Contig2390	Cinnamate-4-hydroxylase [*Theobroma cacao*]	0.635	2.9.E-04	0.769	1.7.E-06	0.971	1.0.E-17	137.99	105.35	22.17	217.30	133.79
Contig4700	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	0.698	3.7.E-05	0.731	1.0.E-05	0.975	1.3.E-18	137.39	22.90	5.67	67.42	39.15
Contig11	phenylalanine ammonia-lyase	0.691	4.6.E-05	0.801	3.1.E-07	0.970	1.9.E-17	136.64	149.82	32.02	286.47	198.08
Contig512	anthocyanidin synthase	0.690	4.9.E-05	0.795	4.3.E-07	0.977	4.9.E-19	135.21	404.75	37.88	1040.29	577.12
Contig482	phospho-2-dehydro-3-deoxyheptonate aldolase chloroplastic-like	0.685	5.8.E-05	0.771	1.6.E-06	0.973	4.3.E-18	133.77	63.08	16.17	155.08	96.18
Contig2957	phospho-2-dehydro-3-deoxyheptonate aldolase chloroplastic-like	0.646	2.0.E-04	0.769	1.7.E-06	0.972	7.0.E-18	133.50	83.69	17.79	201.04	117.76
Contig4670	bifunctional 3-dehydroquinate dehydratase shikimate chloroplastic-like	0.589	9.6.E-04	0.762	2.5.E-06	0.963	2.5.E-16	131.39	28.99	2.65	67.22	38.06
Contig4343	DETOXIFICATION 33	0.620	4.3.E-04	0.738	7.3.E-06	0.967	4.8.E-17	131.30	27.55	6.34	65.55	33.89
Contig1603	chalcone synthase	0.665	1.1.E-04	0.799	3.4.E-07	0.961	4.4.E-16	130.79	17.16	2.37	38.47	23.29
Contig3471	phosphoglycerate mutase	0.675	8.2.E-05	0.689	5.0.E-05	0.969	3.2.E-17	129.51	19.31	6.06	64.69	37.37
Contig492	anthocyanidin synthase	0.728	1.2.E-05	0.826	6.2.E-08	0.961	4.5.E-16	129.25	786.45	84.22	1763.95	1135.29
Contig5640	3 -N-debenzoyl-2 -deoxytaxol N-benzoyltransferase	0.765	2.1.E-06	0.756	3.3.E-06	0.963	2.2.E-16	127.87	16.80	1.28	48.88	33.30
Contig5381	bifunctional 3-dehydroquinate dehydratase shikimate chloroplastic-like	0.553	2.3.E-03	0.797	3.9.E-07	0.962	3.5.E-16	127.57	33.37	2.43	71.83	35.98
Contig11748	Glycosyl hydrolase family isoform 4 [*Theobroma cacao*]	0.627	3.6.E-04	0.743	6.0.E-06	0.958	1.2.E-15	127.55	27.64	8.76	61.16	37.89
Contig9491	phenylalanine ammonia-lyase	0.716	1.8.E-05	0.799	3.4.E-07	0.961	5.6.E-16	126.59	79.83	17.22	147.89	106.51
Contig02435	—NA—	0.615	4.9.E-04	0.706	2.7.E-05	0.967	5.1.E-17	126.13	1.42	0.13	4.91	2.33
Contig4669	bifunctional 3-dehydroquinate dehydratase shikimate chloroplastic-like	0.545	2.7.E-03	0.763	2.3.E-06	0.957	1.6.E-15	125.50	26.59	1.63	60.19	30.06
Contig313	dihydrofolate reductase-like	0.625	3.8.E-04	0.743	5.8.E-06	0.963	2.6.E-16	124.40	21.47	8.36	43.58	26.95
Contig10563	CBL-interacting kinase 14 [*Vitis vinifera*]	0.636	2.8.E-04	0.775	1.3.E-06	0.957	1.7.E-15	124.40	9.50	3.39	17.19	11.04
Contig14773	4-coumarate-- ligase 2-like	0.638	2.6.E-04	0.677	7.5.E-05	0.954	3.8.E-15	123.87	19.06	3.55	54.04	32.35
Contig7191	—NA—	0.574	1.4.E-03	0.749	4.5.E-06	0.945	4.1.E-14	123.55	55.72	22.36	106.59	55.50
Contig9187	phosphoglycerate mutase-like [*Malus domestica*]	0.713	2.1.E-05	0.707	2.6.E-05	0.956	2.4.E-15	123.48	19.45	5.09	60.97	36.77
Contig5021	4-coumarate-- ligase 2-like	0.669	9.8.E-05	0.666	1.1.E-04	0.947	2.6.E-14	123.16	14.78	3.00	42.41	28.65
Contig63662	cytochrome P450 93A3-like	0.571	1.5.E-03	0.772	1.5.E-06	0.944	5.2.E-14	121.69	71.09	21.21	137.73	73.46
Contig628	4-coumarate-- ligase 2-like	0.643	2.3.E-04	0.689	5.1.E-05	0.941	9.1.E-14	121.60	26.63	4.80	71.05	46.92
Contig5765	trans-cinnamate 4-monooxygenase	0.639	2.5.E-04	0.764	2.3.E-06	0.948	2.1.E-14	121.46	191.64	53.84	390.94	231.51
Contig6195	phospho-2-dehydro-3-deoxyheptonate aldolase chloroplastic-like	0.700	3.4.E-05	0.758	3.0.E-06	0.941	1.0.E-13	120.89	38.34	8.88	89.12	60.32
Contig24127	probable acyl-activating enzyme peroxisomal	0.651	1.8.E-04	0.735	8.4.E-06	0.946	2.9.E-14	120.74	7.12	2.38	14.18	9.39
Contig2208	dihydroflavonol 4-reductase	0.531	3.7.E-03	0.822	8.4.E-08	0.941	8.8.E-14	120.68	63.98	6.24	135.23	51.85
Contig8677	4-coumarate partial [*Populus nigra*]	0.677	7.6.E-05	0.726	1.3.E-05	0.937	2.1.E-13	120.16	24.79	4.65	59.51	42.55
Contig7228	—NA—	0.615	4.9.E-04	0.755	3.4.E-06	0.953	5.6.E-15	119.25	12.76	4.12	28.46	14.97
Contig3454	DNA photolyase	0.503	6.4.E-03	0.785	7.6.E-07	0.937	2.2.E-13	119.09	127.61	48.44	217.74	111.43
Contig11937	TRANSPARENT TESTA 12	0.611	5.5.E-04	0.698	3.6.E-05	0.942	8.2.E-14	118.91	16.36	4.22	40.09	22.74
Contig966	flavonoid 3 ,5 -hydroxylase 2-like	0.751	4.2.E-06	0.769	1.7.E-06	0.944	5.4.E-14	118.38	63.90	9.99	154.47	102.12
Contig20409	*bHLH transcription factor (DkMYC1)	0.571	1.5.E-03	0.713	2.1.E-05	0.937	2.1.E-13	118.01	15.86	7.13	27.95	17.19
Contig11936	TRANSPARENT TESTA 12	0.606	6.3.E-04	0.699	3.5.E-05	0.936	2.6.E-13	116.14	18.72	4.84	46.07	25.66
Contig10634	3-dehydroquinate dehydratase shikimate dehydrogenase 2 [*Camellia sinensis*]	0.559	2.0.E-03	0.764	2.2.E-06	0.937	2.0.E-13	114.96	29.84	2.61	64.40	36.93
Contig10633	3-dehydroquinate dehydratase shikimate dehydrogenase 2 [*Camellia sinensis*]	0.556	2.1.E-03	0.763	2.4.E-06	0.938	1.9.E-13	114.93	30.50	2.63	65.56	37.68
Contig54	enolase chloroplastic	0.581	1.2.E-03	0.634	2.9.E-04	0.940	1.2.E-13	113.06	21.47	15.71	47.14	29.21
Contig16617	—NA—	0.749	4.5.E-06	0.663	1.2.E-04	0.938	1.8.E-13	112.25	8.92	1.19	29.65	22.52
Contig5302	cinnamate 4- partial [*Pinus monticola*]	0.587	1.0.E-03	0.748	4.7.E-06	0.933	4.8.E-13	112.19	109.44	36.95	212.04	118.54
Contig8185	—NA—	0.538	3.1.E-03	0.827	5.8.E-08	0.926	1.7.E-12	111.36	41.02	5.30	81.88	39.30
Contig10004	dihydroflavonol 4-reductase	0.488	8.5.E-03	0.834	3.6.E-08	0.919	4.9.E-12	111.23	234.58	24.68	444.64	173.28
Contig11378	CBL-interacting serine threonine- kinase 21-like	0.593	8.8.E-04	0.782	8.7.E-07	0.934	4.2.E-13	110.70	13.96	4.81	25.73	15.98
Contig2445	dihydroflavonol 4-reductase	0.486	8.7.E-03	0.835	3.3.E-08	0.918	5.6.E-12	110.58	177.24	19.99	335.72	130.44
Contig04759	chlorophyll a-b binding of LHCII type 1-like	0.540	3.0.E-03	0.730	1.0.E-05	0.942	7.4.E-14	110.55	14.52	6.06	34.77	16.65
Contig2389	cinnamate 4- partial [*Populus trichocarpa*]	0.692	4.6.E-05	0.675	8.1.E-05	0.920	4.6.E-12	110.26	22.61	3.79	52.56	43.40
Contig3456	blue-light photoreceptor PHR2	0.468	1.2.E-02	0.788	6.4.E-07	0.920	4.2.E-12	109.89	122.01	44.25	188.76	97.93
Contig17999	—NA—	0.579	1.2.E-03	0.795	4.3.E-07	0.928	1.1.E-12	109.66	36.22	6.61	69.06	35.61
Contig4904	enolase chloroplastic	0.553	2.3.E-03	0.645	2.1.E-04	0.928	1.2.E-12	109.40	16.87	9.09	41.77	22.66
Contig8947	—NA—	0.557	2.1.E-03	0.801	3.0.E-07	0.922	3.2.E-12	108.95	41.76	4.69	93.45	40.93
Contig10759	chlorophyll a-b binding chloroplastic	0.568	1.6.E-03	0.778	1.1.E-06	0.928	1.3.E-12	108.94	13.45	2.96	32.97	13.82
Contig3538	serine carboxypeptidase-like 18	0.741	6.5.E-06	0.790	5.8.E-07	0.930	8.9.E-13	108.72	25.89	1.64	59.63	42.88
Contig13352	MATE efflux family [*Theobroma cacao*]	0.620	4.4.E-04	0.672	9.0.E-05	0.919	5.1.E-12	107.75	19.20	5.30	46.04	26.54
Contig5729	phenylalanine ammonia-lyase	0.705	2.8.E-05	0.790	5.9.E-07	0.912	1.4.E-11	107.36	88.06	20.34	162.21	121.93
Contig12030	—NA—	0.759	2.8.E-06	0.604	6.7.E-04	0.928	1.3.E-12	107.26	3.99	1.28	15.38	12.11
Contig3455	blue-light photoreceptor PHR2	0.479	9.8.E-03	0.785	7.4.E-07	0.913	1.3.E-11	107.23	116.10	42.85	193.72	93.87
Contig50	flavanone 3-hydroxylase	0.564	1.8.E-03	0.886	3.5.E-10	0.916	7.7.E-12	107.13	1123.75	190.24	1661.77	1027.42
Contig5671	phospho-2-dehydro-3-deoxyheptonate aldolase chloroplastic-like	0.781	9.6.E-07	0.719	1.6.E-05	0.926	1.8.E-12	106.24	65.52	15.78	179.44	120.67
Contig30879	—NA—	0.687	5.5.E-05	0.660	1.3.E-04	0.935	3.5.E-13	105.36	11.67	3.96	39.79	22.17
Contig10485	—NA—	0.586	1.0.E-03	0.733	9.2.E-06	0.921	4.1.E-12	105.10	6.16	0.89	17.45	7.03
Contig10825	—NA—	0.464	1.3.E-02	0.744	5.7.E-06	0.910	2.1.E-11	104.88	53.42	23.13	84.37	48.51
Contig4832	phospho-2-dehydro-3-deoxyheptonate aldolase chloroplastic-like	0.752	4.0.E-06	0.710	2.3.E-05	0.913	1.2.E-11	104.55	60.93	12.21	169.69	113.02
Contig2811	WAT1-related At5g47470	0.590	9.5.E-04	0.839	2.5.E-08	0.914	1.1.E-11	104.00	12.25	3.88	21.65	11.66
Contig4319	limonoid UDP-glucosyltransferase	0.798	3.7.E-07	0.751	4.2.E-06	0.907	2.7.E-11	103.27	134.58	16.98	306.54	249.87
Contig10040	1-Cys peroxiredoxin	0.643	2.2.E-04	0.705	2.9.E-05	0.926	1.6.E-12	102.58	29.90	0.57	121.43	60.93
Contig52008	*LOL1 [*Vitis vinifera*]	0.718	1.7.E-05	0.672	9.1.E-05	0.927	1.3.E-12	102.45	2.03	0.46	5.83	3.70
Contig9595	blue-light photoreceptor PHR2	0.463	1.3.E-02	0.754	3.6.E-06	0.907	2.9.E-11	101.92	124.90	61.84	203.28	108.60
Contig11166	*MYB transcription factor (DkMYB4)	0.668	1.0.E-04	0.688	5.2.E-05	0.928	1.2.E-12	101.84	7.11	1.79	15.46	11.86
Contig4831	3-deoxy-d-arabino-heptulosonate 7-phosphate synthase isoform 3 [*Theobroma cacao*]	0.752	4.0.E-06	0.712	2.1.E-05	0.909	2.2.E-11	101.82	54.77	11.19	151.49	100.52
Contig14177	—NA—	0.493	7.7.E-03	0.669	1.0.E-04	0.915	9.3.E-12	101.68	13.63	4.82	34.05	16.21
Contig16840	—NA—	0.759	2.8.E-06	0.669	9.9.E-05	0.907	3.1.E-11	101.36	7.06	1.63	17.86	14.22
Contig16841	—NA—	0.783	8.4.E-07	0.668	1.0.E-04	0.908	2.5.E-11	101.29	8.28	1.82	19.66	16.68
Contig338	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase	0.548	2.5.E-03	0.690	4.8.E-05	0.902	5.7.E-11	101.20	109.71	49.70	198.80	121.65
Contig12413	*bHLH transcription factor (DkMYC1)	0.502	6.5.E-03	0.713	2.1.E-05	0.901	6.0.E-11	100.89	9.19	3.48	16.13	8.41
Contig6422	Arabinanase levansucrase [*Theobroma cacao*]	0.790	5.8.E-07	0.661	1.3.E-04	0.926	1.6.E-12	100.09	3.17	0.58	11.97	7.97
Contig41252	—NA—	0.642	2.3.E-04	0.753	3.7.E-06	0.916	7.7.E-12	99.24	6.52	0.77	16.00	8.23
Contig15434	UDP-glucuronosyl/UDP-glucosyl transferase family [*Arabidopsis lyrata*]	0.604	6.7.E-04	0.790	5.8.E-07	0.916	8.1.E-12	97.95	7.58	0.74	14.32	8.12
Contig26724	MATE efflux family chloroplastic	0.583	1.1.E-03	0.701	3.3.E-05	0.907	3.0.E-11	97.87	3.71	1.06	10.05	4.86
Contig43179	—NA—	0.538	3.1.E-03	0.634	2.9.E-04	0.908	2.4.E-11	97.54	4.88	1.10	16.36	9.81
Contig30528	—NA—	0.677	7.7.E-05	0.639	2.5.E-04	0.917	6.8.E-12	97.44	1.40	0.19	4.34	2.63
Contig7441	chlorophyll a-b binding of LHCII type 1-like	0.601	7.3.E-04	0.779	1.1.E-06	0.916	8.6.E-12	96.98	62.05	26.13	131.60	74.75
Contig8324	blue-light photoreceptor PHR2	0.429	2.3.E-02	0.714	2.0.E-05	0.890	2.4.E-10	95.39	203.08	75.07	354.04	161.15
Contig40982	—NA—	0.594	8.7.E-04	0.666	1.1.E-04	0.909	2.1.E-11	95.32	1.93	0.29	4.76	3.03
Contig24120	CBL-interacting kinase 14 [*Vitis vinifera*]	0.522	4.4.E-03	0.780	1.0.E-06	0.905	3.9.E-11	95.13	7.34	2.01	12.14	7.41
Contig36141	—NA—	0.556	2.1.E-03	0.654	1.6.E-04	0.904	4.1.E-11	95.02	4.65	1.84	10.57	5.56
Contig21096	—NA—	0.676	7.9.E-05	0.856	6.5.E-09	0.907	2.9.E-11	94.77	6.15	0.69	9.85	6.92
Contig8060	dihydroflavonol 4-reductase	0.496	7.3.E-03	0.792	5.1.E-07	0.895	1.3.E-10	94.64	47.51	4.66	92.78	36.61
Contig10350	CBL-interacting serine threonine- kinase 21 isoform X1	0.555	2.2.E-03	0.639	2.5.E-04	0.898	9.5.E-11	94.55	3.73	1.45	8.09	5.53
Contig54344	PREDICTED: uncharacterized protein LOC104448399 [*Eucalyptus grandis*]	0.790	5.8.E-07	0.616	4.9.E-04	0.903	5.2.E-11	94.47	2.27	0.14	11.41	7.95
Contig30130	hypothetical protein	0.555	2.2.E-03	0.740	6.7.E-06	0.901	6.4.E-11	94.38	20.70	8.90	42.56	23.32
Contig44069	—NA—	0.578	1.3.E-03	0.652	1.7.E-04	0.908	2.5.E-11	94.01	1.87	0.25	4.79	2.75
Contig56713	Riboflavin synthase alpha isoform 1 [*Theobroma cacao*]	0.759	2.9.E-06	0.728	1.1.E-05	0.894	1.4.E-10	93.87	15.42	3.59	27.69	24.61
Contig3	serine carboxypeptidase-like 18 isoform X1	0.765	2.2.E-06	0.686	5.5.E-05	0.884	4.6.E-10	93.32	29.35	1.51	93.12	77.66
Contig5761	chalcone synthase	0.714	2.0.E-05	0.744	5.7.E-06	0.879	7.4.E-10	93.12	316.56	42.74	632.78	508.17
Contig15277	beta-D-glucan exohydrolase [*Arabidopsis thaliana*]	0.652	1.7.E-04	0.680	6.8.E-05	0.889	2.6.E-10	92.93	20.49	6.24	45.88	34.88
Contig10565	anthocyanidin reductase	0.475	1.1.E-02	0.830	4.6.E-08	0.885	4.1.E-10	92.79	34.90	6.07	49.39	26.03
Contig18636	PREDICTED: uncharacterized protein LOC105961982 [*Erythranthe guttata*]	0.821	8.9.E-08	0.606	6.3.E-04	0.909	2.2.E-11	92.05	2.56	0.31	14.79	10.37
Contig23479	—NA—	0.495	7.4.E-03	0.664	1.2.E-04	0.899	7.8.E-11	90.61	1.42	0.17	3.18	1.57
Contig18327	chlorophyll a-b binding chloroplastic-like	0.463	1.3.E-02	0.755	3.5.E-06	0.891	2.2.E-10	90.40	15.96	3.44	38.55	12.12

Transcription factors in this list were indicated with asterisk (*).

[b]cor.: correlation coefficient between PA traits (accumulation ratio and concentration) and given gene.

[d]MM: Module membership: correlation coefficient between module eigengene and given gene.

[a]connectivity: sum of TOM within the module, which was used for the indicator of centrality in this module.

[e]P: PCNA, [e]NP:non-PCNA