Article

Transcriptome-wide identification and quantification of Caffeoylquinic acid biosynthesis pathway and prediction of their putative BAHDs gene complex in *A. spathulifolius*

Jean Claude Sivagami [1,2], Sunmi Park [2] and SeonJoo Park *

[1]   [1]Department of Life Sciences, Yeungnam University, Gyeongsan, Gyeongbuk 38541, South Korea

[2]   [2]Institute of Natural Science, Yeungnam University, Gyeongsan, Gyeongbuk 38541, South Korea

*   Correspondence: sjpark01@ynu.ac.kr; Department of Life Sciences, Yeungnam University, Gyeongsan, Gyeongbuk 38541, South Korea

Abstract: The phenylpropanoid pathway is a major secondary metabolite pathway that helps plants overcome biotic and abiotic stress and produces various by-products that promote human health. Its by-product, chloroquinic acid (CQA), is a soluble phenolic compound present in many angiosperms. Hydroxycinnamate-CoA shikimate/quinate transferase(BAHDs superfamily enzyme) is a significant enzyme that plays a role in accumulating CQA biosynthesis. This study analyzed transcriptome-wide identification of the phenylpropanoid to chloroquinic acid biosynthesis candidate genes in *A. spathulifolius* flowers and leaves. Transcriptomic analyses of the flowers and leaves showed a differential expression of the PPP and CQA biosynthesis regulated unigenes. An analysis of PPP captive unigenes revealed the following: the major duplication of the key enzyme, *PAL,* 120 unigenes in leaves and 76 in flowers; the gene encoding *C3'H,* 169 unigenes in leaves and 140 unigenes in flowers; duplicated unigenes of *4CL,* 41 in leaves and 27 in flowers. In addition, *C4H* unigenes had 12 unigenes in the leaves of *A. spathulifolius* and four in the flowers. The characterization of the BAHDs superfamily members identified 82 in leaves and 72 in flowers. Among them, phylogenetic analysis showed that five unigenes encoded HQT and three encoded HCT in *A. spathulifolius*. The three HQT are common to both leaves and flowers, whereas the two HQT were specialized for leaves. The pattern of HQT synthesis was upregulated in flowers, whereas HCT was expressed strongly in the leaves of *A. spathulifolius*. Overall, *4CL*, *C4H*, and HQT are expressed strongly in flowers, and caffeic acid and *HCT* show more expression in leaves. Therefore, CQA biosynthesis occurs in the flowers of *A. spathulifolius* rather than leaves.

Keywords: Phenylpropanoid pathway, Caffeoylquinic acid, BAHDs, hydroxycinnamoyl-coenzyme A: quinate hydroxycinnamoyl transferase, hydroxycinnamoyl-coenzyme A: shikimate/quinate hy droxycinnamoyl transferase, *A. spathulifolius*, DEGs

1. Introduction

Plant secondary metabolites (PSM) are a group of organic compounds that assist in protecting plants against biotic and abiotic stress [1-4]. PSM metabolites are non-essential to normal life but impart competence to stressful environments [5, 6]. Secondary metabolites can be divided into three types: terpenoids, polyketides, and phenylpropanoid [7, 8]. The phenylpropanoid (PP) metabolism, the production of enormous compounds by the intermediate process of the shikimate pathway, is present in bacterial, fungi, and plants but absent in animals [9]. These phenolic compounds assist in the plant defense system against insects and fungi [10]. The main compounds of PP metabolites products of phenylalanine (Phe) are precursors that regulate many metabolites, such as flavonoids, tannins, lignins, and phenylpropanoid [11]. The shikimate pathway network connects the carbon metabolism and the AAA (Aromatic

Amino Acid) by converting phosphate phenol pyruvate and erthrose4-phosphate in glycolysis pentose phosphate pathway to chorismate [12]. Phe is the main compound that regulates the phenylpropanoid pathway [13]. PP metabolites are highly involved in many aspects of plant development and morphologically support and respond to both biotic and abiotic stress conditions [14]. In PPP, transamination of the phenylpyruvate and 4-hydroxyphenylpyruvate to form Phe via the trans-cinnamate 4-monooxygenase enzyme(*C4H*) occurs. Phe helps derive the byproducts of phenol compounds in stress environments and is involved in the production of iso-flavonoids, especially in diseased plants and flavonoids in UV irradiation during symbionts and salicylic acid in plant-pathogen interactions [15, 16]. Unlike caffeoyl-shikimate, CQA is rarely regulated in lignin biosynthesis but is involved in biotic and abiotic stresses, particularly UV radiation [16]. Hydroxycinnamic acids (HCs) are a major class of phenolic compounds in every plant. HCs include caffeic acid, which occurs mainly as an ester with quinic acid rather than shikimic acid, and is called 5-caffeoylquinic acid (CQA) [17]. HCs produce caffeic acid, coumaric acid, ferulic acid, and sinapic acid through PPP in plants. The effects of adding a combination of hydroxycinnamoyl-coenzyme A: shikimate hydroxycinnamoyl transferase (HCT*)* or hydroxycinnamoyl-coenzyme A: quinate hydroxycinnamoyl transferase (HQT) to *C3'H* help catalyze the H-unit of p-Coumaroyl CoA into Caffeoyl CoA. This process results in the production of chlorogenic acid or chloroquinic acid (CGA or CQAs) [18-21]. The HCT and HQT belong to BAHD acyltransferase superfamily. The HCT enzymes use coenzyme A-activated acyl donors and HQT uses quinate as a acceptor rather than shikimate acid [22] The HQT is involved in chloroquinic acid in some angiosperm [23]. *HQT* is estimated to be one of the key enzymes for the synthesis of CQA in plants [20]. These assist human diets. CQA has biological effects in blood circulation; its main function is to inhibit the oxidation of low-density lipoprotein in vitro [24, 25]. Various organic compounds of hydroxycinnamic acid are based on various types of BAHDs proteins abundant in apples, peaches, berry, carrot, and coriander. These compounds are chemically unstable, degradable and form other compounds [26, 27]. The derived byproduct of hydroxycinnamic acid protects against degenerative and age-related diseases in animals [28-31]. Most major soluble phenolic compounds can be found in Solanum species, such as eggplant, potato and tomato [32-34]. These compounds accumulate to substantial levels in coffee, apples, plums, and pears [35-38]. They are absorbed directly by the small intestine but most CQA is absorbed in the large intestine by the esterase of the gut microflora to release caffeic acid [39].Previous studies evaluated many Asteraceae species for PP derived by products in different tissue of plants. In particular, the sunflower family has been studied widely for CQA production in sprouts, leaves, and roots [40-43].   In this study, transcriptome and quantitative PCR were used to analyze the CQA biosynthesis response in the leaves and flowers of *A. spathulifolius*. The differential expression of PPP involved unigenes throughout the leaf and flower transcriptome was analyzed. In addition, this study analyzed the BAHDs enzyme members to the putative HXXXD domain of HQT (quinic acid) and HCT (shikimic acid). Finally, the putative PPP and BAHDs unigenes were quantified in different plants parts of *A. spathulifolius*. The identification of these PPP and BAHDs provided valuable insights

for elucidating the response metabolites of chlorogenic acid or chloroquinic acid biosynthesis in *A. spathulifolius* and an importans source of ornamental plants to useful drug discovery.

2. Results

2.1. *Assembly and gene annotation*

The assembled flower reads produced 146,337 unigenes, with an average contig 811.58 bp in length. The assembled bases of transcripts had an N50 value of 1,279 bp in length, and overall alignment rates of 91.71% to the flower paired-end reads. The GC content was 38.09% on average. In contrast, the leaf transcriptome was reported previously [44]. The assembly completeness was measured by BUSCO analysis to be 91.4 %(leaf) and 91.7 % (flower), which are the complete transcripts to the eudicots database via tblastn aligns (**Figure S1**), indicating the good quality of unique transcripts. The flower of *A. spathulifolius* retrieved using the Nr database was 65,129 and 48,896 against the KEGG database, and 70,019 unigenes were identified in the Pfam database (**Table 1**). The PP pathway-involved unigenes in flowers showed a transcript of 1,128 and in the leaf had 1,287 unigenes. The Gene ontology terms indicated 40.9% to the phenylpropanoid metabolic process, 31.8% to the response to wounding, 22.7% to the lignin biosynthetic process, and 18.2% to the cinnamic acid biosynthetic process to the biological process function (**Table 2**).

Table 1. The statistical comparisons of flower and leaf of *A. spathulifolius* to the annotation of unigenes.

S.No	sample	Total base	unigenes	Uniprot	KEGG	PPP	BAHD
1	Flower	98467589	146,337	65,129	70,019	1,128	82
2	leaf	71660029	98,860	48,896	39,890	1,287	72

2.2. *Functional characterization and DEGs of PP biosynthesis*

The phenylpropanoid biosynthesis-involved unique transcripts were annotated to the KEGG to explore the CQA synthesis pathway in *A. spathulifolius*. The perspective diagram of the CQA biosynthesis pathway in *A. spathulifolius* revealed three different possible routes through the leaf and flower transcriptome: i) coumaric acid to p-cinnamoyl quinic acid (*p-CQA*) to *C'3H*, ii) cinnamoyl-CoA with *C4H* followed by p-CQA, and iii) p-coumaric acid +*4CL*+Caffeoyl shikimic acid to Caffeoyl CoA with *HQT* to form CQA (Figure 1). The total identified PPP involved unigenes showed DEGs in the flower, and leaf reveals 464 different unigenes were upregulated in the flower. In the leaf transcriptome, there were 704 upregulated unigenes compared to the flower. Among the DEGs comparison, 482 unigenes were upregulated in both leaf and flower. One hundred and forty-two unigenes were downregulated in the leaf and flower of *A. spathulifolius* (Figure 2a). Regarding the DEGs unigenes, the *PAL* showed considerable duplication of the genes in both cases of leaves with 120 unigenes, and the flower had 76 copies of *PAL* genes. *PAL1* was highly regulated in the flower with a 793.68 FPKM value, whereas the leaf showed 765.03 FPKM. The last step to synthesis was the Cytochrome P450 (*C3'H*),

which catalyzes the 3'-hydroxylation of *p*-coumaric esters of shikimic/quinic acids to form CQA. *C3'H*

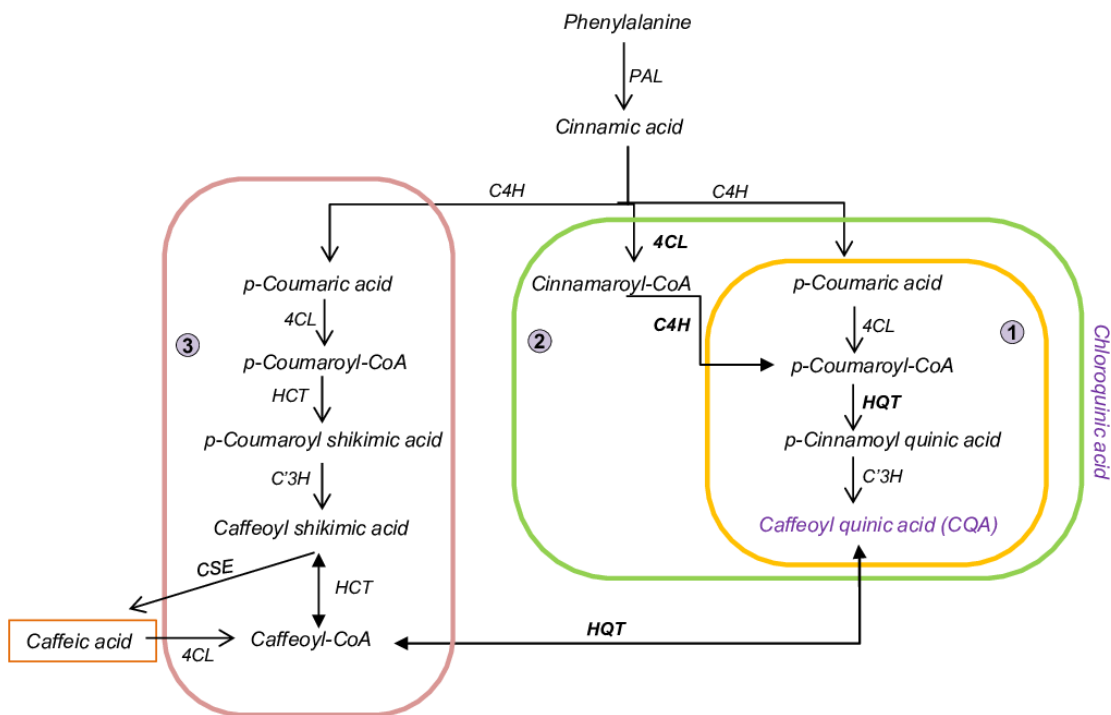**Phenylalanine Biosynthesis to Chloroquinic acid (CQA) in *A. spathulifolius***

Figure 1. Schematic diagram of Phenylalanine towards to Chloroquinic acid (CQA) biosynthesis in *A. spathulifolius*. 1, 2 and 3 indicates the different route pathway of chlorogenic acid production. **PAL**, *phenylalanine ammonia-lyase;* **C4H**, *Cinnamic acid 4-hydroxylase;* **4CL**, *4- coumarate--CoA ligase;* **C3'H**, *p-coumarate 3'Hydroxylases;* **CSE**, *Caffeoylshikimate esterase.* **HCT**, *Hydroxycinnamoyl CoA shikimate Hydroxycinnamoyl-transferase;* **HQT**, *Hydroxycinnamoyl CoA quinate Hydroxycinnamoyl-transferase.*

Table 2. Enriched GO term of Biological Process (BP) function in *A. spathulifolius* regarding PPP.

GO ID	BP TERM	%	p-value
GO:0009698	phenylpropanoid metabolic process	40.9	9.00E-21
GO:0009800	cinnamic acid biosynthetic process	18.2	1.10E-08
GO:2000762	regulation of phenylpropanoid metabolic process	18.2	3.80E-08
GO:0009611	response to wounding	31.8	4.90E-08
GO:0009699	phenylpropanoid biosynthetic process	18.2	6.90E-06
GO:0006559	L-phenylalanine catabolic process	13.6	3.10E-05
GO:0009809	lignin biosynthetic process	22.7	2.80E-04
GO:0008152	metabolic process	13.6	3.80E-04
GO:0009411	response to UV	13.6	2.40E-03
GO:0080167	response to karrikin	13.6	8.30E-03
GO:0009813	flavonoid biosynthetic process	13.6	1.10E-02

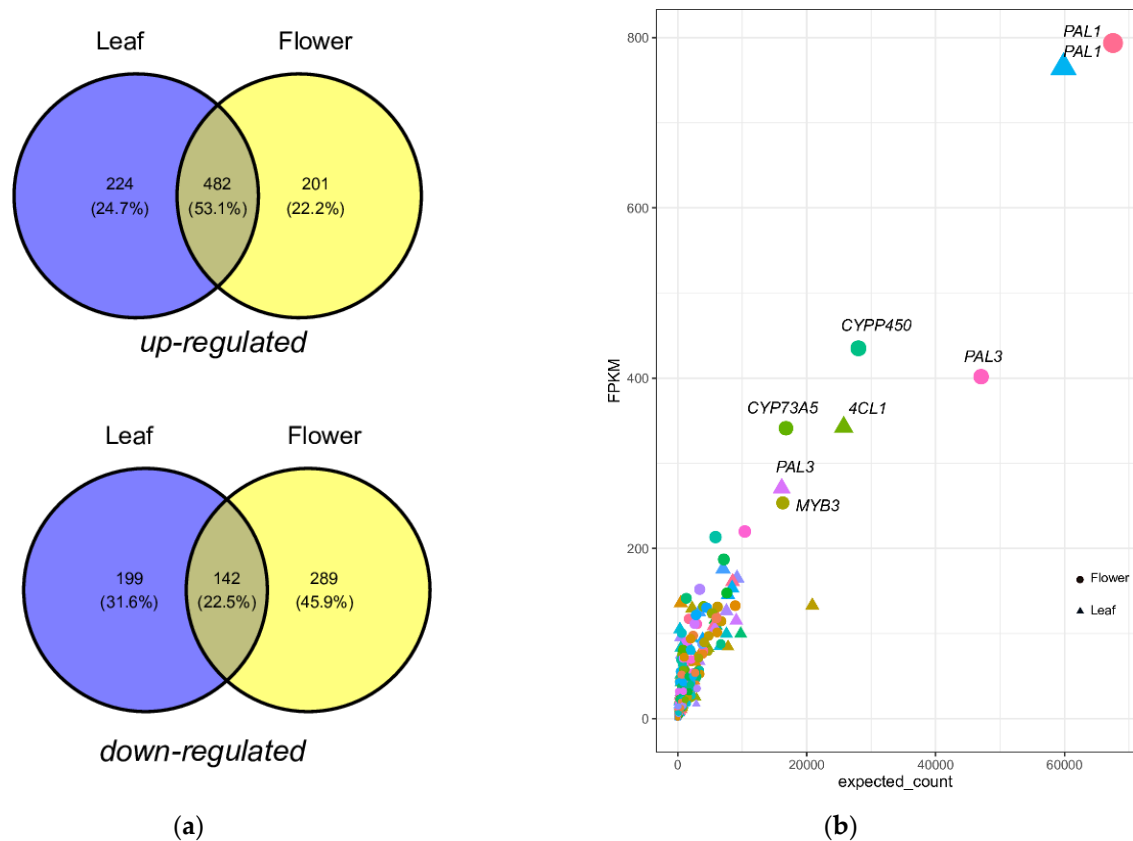(a)   (b)

Figure 2. (**a**) The geom point plot shows the up-regulated differential expressed genes of PPP in leaf and flower of *A. spathulifolius*.; (**b**) A Venn diagram shows up and down regulation of PPP in *A. spathulifolius* leaf and flower.

had169 unigenes in the leaf and 140 unigenes in the flower of *A. spathulifolius*. There were 27 unigenes in the flower and 41 in the leaf with regards to the *4CL* enzymes. The two *4CL* genes in the flower were highly regulated with 435~253.48 than in the leaf. The two leaf *4CL* with 332.62 and 165 FPKM has been observed. *C4H* protein was found to have 12 unigenes in the leaves and four unigenes in the flower. The four unigenes of *C4H* were highly upregulated in the leaves, while three unigenes of *C4H* were upregulated in the flowers with FPKM value 160~77 in range (Figure 2b) .

2.3. *Estimation of BAHDs superfamily*

The leaves and flowers of *A. spathulifolius* were compared. The BAHD superfamily genes were vastly duplicated in the flowers with 82 unigenes in that 30 unigenes with 'HXXXD' and 'DFGWG' were observed. In the leaf, 72 unigenes were related to the BAHD transferase family. Only 33 unigenes had two domains. The HXXXD domain is related to HQT ("HTLAD/HTLSD"), which promoted the production of CQA byproducts. In leaves, there are 704 upregulated PPP-involved unigenes; 82 genes were BAHD member unigenes among the upregulated unigenes. In contrast, 457 unigenes were found in the flower; 72 unigenes belonged to the BAHD clades, which were upregulated. The HXXXD domain recognizes the distribution of different BAHD complex genes; the most available domain was 'HTLSD' with five copies in the leaf and three unigenes in the flower (Figure 3a). The duplication of these unigenes showed different
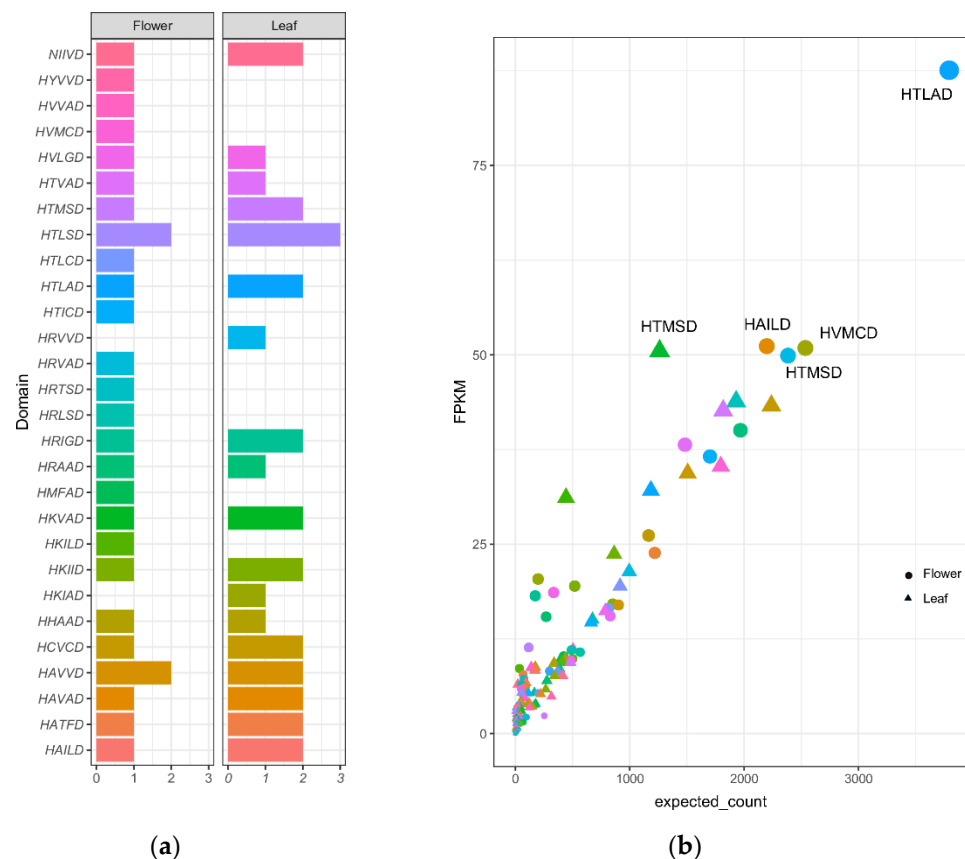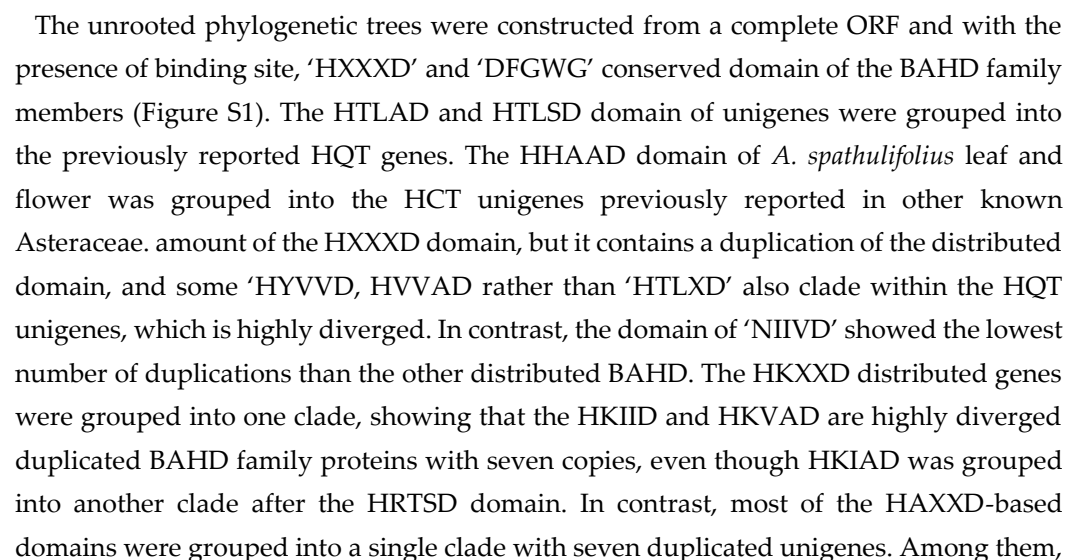
Figure 3. (**a**) Distribution of "*HXXXD*" domain among the leaf and flower of *A. spathulifolius*. The blank space indicates not presents.; (**b**) The DEGs expression of up and down regulated *BAHD transferase* unigenes among the different transcripts.

FPKM values with one another in both the leaf and flower. In addition, the distributed domain varied from the flower to leaf; the flower HXXXD domain maintained different amino acid codons than the leaf, even though the leaf has an equal amount of the HXXXD domain, but it contains a duplication of the distributed domain, and some 'HYVVD, HVVAD, HVMCD, HRVVD, and HKIAD' which specialized to flower, are not present. The domain of 'HAVVD' contains two copies in both the leaf and flower. The HRIGD, HKIID, HAVAD, HATFD, and HAILD domain presented one copy in flowers, whereas it had two copies each in the leaves. The HTLAD (HQT)-related unigenes tended to have three duplication genes in the leaf, whereas the flower showed only two copies (Figure 3a). In the flower, the 'HTLSD' (HQT) genes were upregulated according to DEGs analysis compared to the leaf HQT. The unknown function of the 'HTMSD' domain in both the leaf and flower was strongly expressed in the same value of FPKM (>50) in *A. spathulifolius*. The lowest expressed domain was the single duplication copy of the 'HRXXD' domain of the HCs unigenes (Figure 3b). Most of the BAHD transferase protein between the flower and leaf showed the highest FRKM value from the highest of 87.58 to the lowest of <2. The HQT with the 'HTLAD' domain contained unigenes in the flower expressed strongly with an FPKM value of 87.58. The unigenes leaf 'HAMSD' domain with 50.48 were detected, respectively. This indicates that CQA production in the flower of *A. spathulifolius* is higher than the leaf.

Figure 4. Phylogenetic tree shows the different type of up-regulated HXXXD among leaf and flower of *A. spathulifolius*.

2.4. *BAHDs protein Phylogenies*

The unrooted phylogenetic trees were constructed from a complete ORF and with the presence of binding site, 'HXXXD' and 'DFGWG' conserved domain of the BAHD family members (Figure S1). The HTLAD and HTLSD domain of unigenes were grouped into the previously reported HQT genes. The HHAAD domain of *A. spathulifolius* leaf and flower was grouped into the HCT unigenes previously reported in other known Asteraceae. amount of the HXXXD domain, but it contains a duplication of the distributed domain, and some 'HYVVD, HVVAD rather than 'HTLXD' also clade within the HQT unigenes, which is highly diverged. In contrast, the domain of 'NIIVD' showed the lowest number of duplications than the other distributed BAHD. The HKXXD distributed genes were grouped into one clade, showing that the HKIID and HKVAD are highly diverged duplicated BAHD family proteins with seven copies, even though HKIAD was grouped into another clade after the HRTSD domain. In contrast, most of the HAXXD-based domains were grouped into a single clade with seven duplicated unigenes. Among them,

the HRAAD domains showed the least diverged group of proteins in *A. spathulifolius*: two copies in the flower and one in the leaf. Based on the phylogenetic tree, the HXXXD domain highly diverged in leaf and flower of *A. spathulifolius*.

2.5. *Quantification of PPP unigenes*

Previous studies of *A. spathulifolius* leaf transcriptome-identified PP candidate genes were designed for reverse transcription (RT) to confirm the quality of isoform produced by assembled transcripts (Figure S2). The total assembled

(**a**)   (**b**)
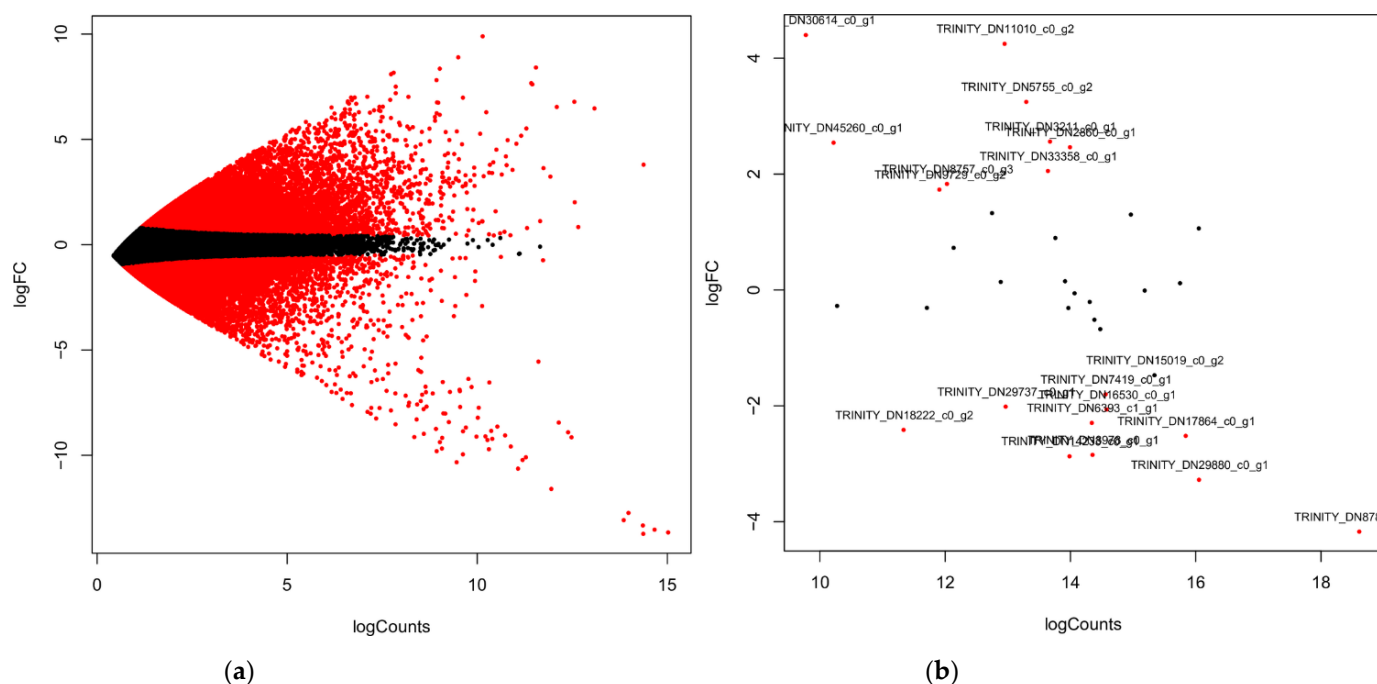
Figure 5. (**a**) MA plot of leaf and flower DEGs in *A. spathulifolius*. Black indicates the no expression, below and above red color indicates the down and up regulation.; (**b**) The plot shows "HXXXD" conserved domain (BAHDs superfamily members) down and up regulation unigenes in both leaf and flower of *A. sapthulifolius*.

unigenes in both leaf and flower produced 345,781 unigenes. The DEGs analysis of the leaf and flower revealed the up and downregulated unigenes shown in MA plot. The quantification of DEGs showed that 146,78 were upregulated and 9,789 were downregulated; the left unigenes showed no differential expression and were defined in 0 values in the smear plot (**Figure 5a**). The quantification flower hydroxycinnamoyl transferase unigenes showed the top nine upregulated genes, and 11 downregulated unigenes in the flower, which was also distributed in the leaf transcriptome (**Figure 5b**). The quantification flower hydroxycinnamoyl transferase unigenes showed the top nine upregulated genes, and 11 downregulated unigenes in the flower, which was also distributed in the leaf transcriptome (**Figure 5B**). The production of *CQA*-derived hydroxycinnamoyl transferase enzymes in both the leaf unigenes of PPP was identified in
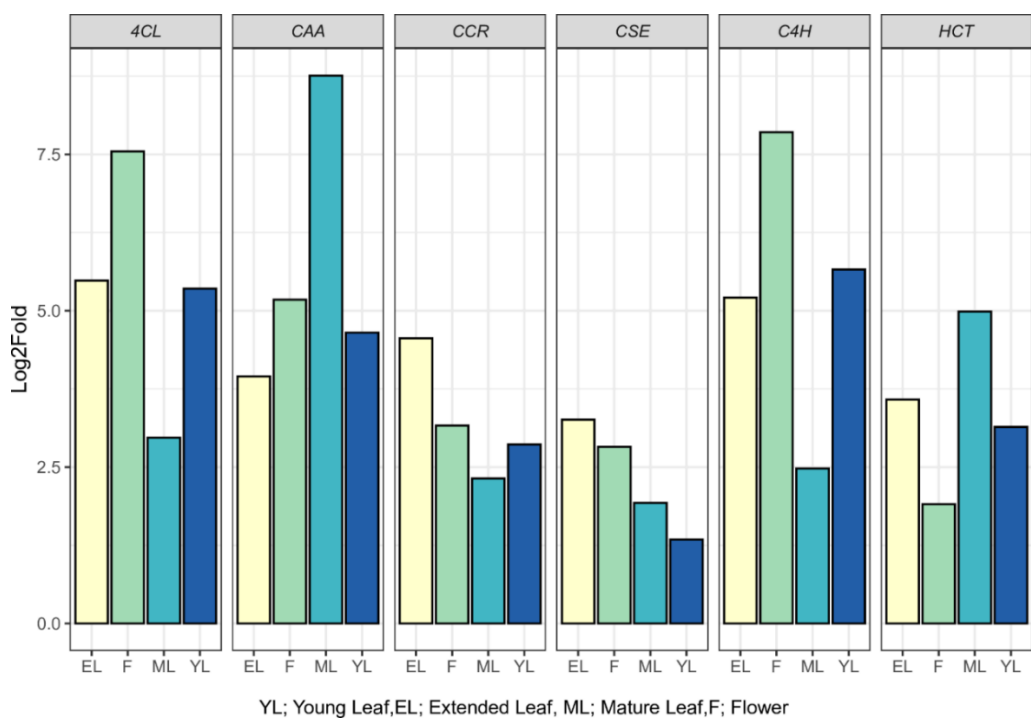
Figure 6. Quantification of PPP involved unigenes in different plant parts of *A. spathulifolius*. 4CL; *4-coumarate--CoA ligase*, CAA; *Coumaroyl-CoA*, CCR; *Cinnamoyl-CoA reductase*, CSE;   *Caffeoylshikimate esterase*.
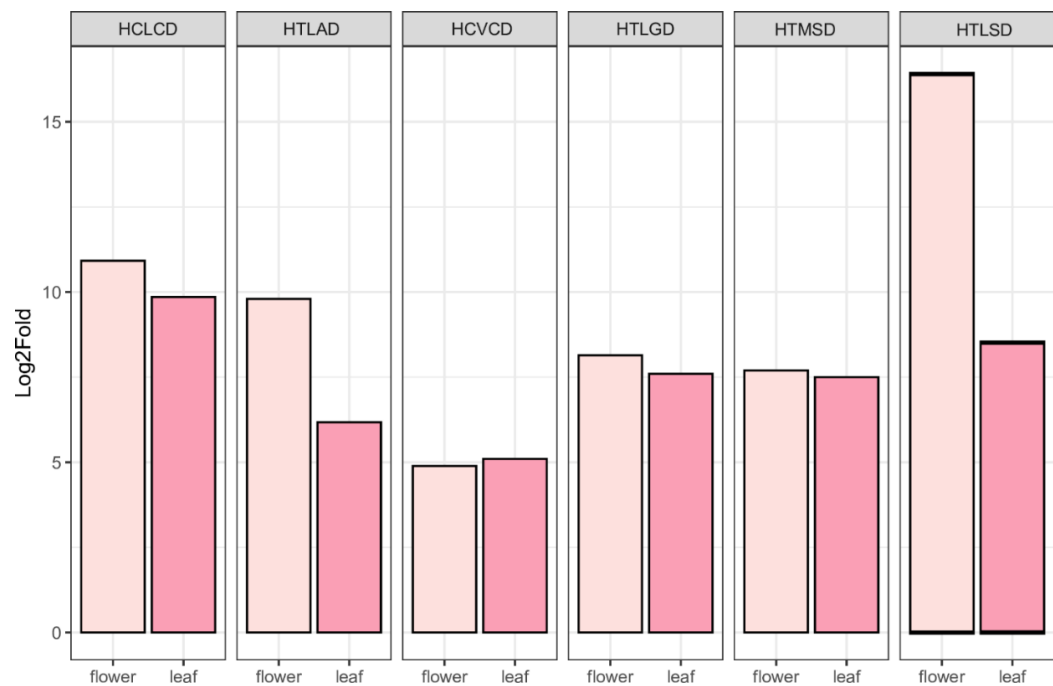

Figure 7. overlay of quantified expression of *HXXXD – Hydroxycinnamoyl transferase* unigenes in flower and leaf.

the different parts of the leaf, such as young leaf (YL), extended leaf (EL), and mature leaf (ML) stages (**Figure 6A**). qRT-PCR revealed the high production of *PAL*, showing large Log2fold changes (above >345 FPKM in mature and flower samples). The *4CL* protein shows the high fold changes in flower and the second-highest activity in the elongation leaf and young leaf of *A. spathulifolius*. CAA shows the largest fold change of

<9.5 in the mature leaf and >5 .5 to 4 fold change in the flower compared to the elongation leaf and young leaf of *A. spathulifolius*. The *CSE* unigenes were equal to the *4CL* genes, showing a high ΔCt value in the flower than the leaf of *A. spathulifolius*. Finally, the HCT showed the lowest ΔCt in the flower, which was highly activated in the leaf rather than the flower. The random BAHD protein family was also analyzed using real-time PCR. Quantification of the 'HCVCD' gene showed the lowest expression among the BADH unigenes chosen. The highest expression was observed in 'HTLSD' with ΔCt in flower and second most unigene 'HCLCD', followed by the 'HTLAD, HTMSD, and HTLGD' BAHD family in the leaf and flower of *A. spathulifolius* (**Figure 6**). Based on the random 'HXXXD' domain, the protein of *A. spathulifolius* showed an unknown function of 'HTMSD' in the pathway mechanism in *Aster*. Predominately, the HQT (HTLAD and HTLSD) showed the highest ΔCt value compared to the root. Hence, the synthesis of production and upregulation of *HQT* was higher in flowers of *A. spathulifolius* than in the leaf. In contrast, the leaf of *A. spathulifolius* showed the *HCT* (Cinnamic acid rather than quince acid) expressed in producing compunds like caffeic acid and other cinnamic acid.

3. Discussion

Phenylpropanoid biosynthesis produces various secondary metabolites, most of which have beneficial effects on human health [45]. These metabolisms are of concern regarding diabetes, obesity, cancer, and cardiovascular disease that are a significant burden on the world health care system. CQA is an abundant polyphenol compound in the human diet produced by almost all plants [19, 31, 39]. Polyphenols compounds are abundant in coffee, which is why coffee plants have attracted considerable attention for CQA production (12, 13 & 14). CQA affects the glucose and lipids metabolism, where intermediates regulate the breakdown of glucose and lipids [19, 25]. A range of mingled pathways was reported to produce *CQA* through the phenylpropanoid pathway. Based on the KEGG pathway, *A. spathulifolius* could produce CQA as a byproduct in a three-way route [46, 47]. DEGs analysis of the transcriptome-produced PPP involved 1128 unigenes in the flower transcriptome and 1287 in the leaf transcriptome (**Table 1**).   Among the PPP-involved genes, the *PAL* 1 enzyme showed the up regulation in both leaf and flower, which is the key product of PPP in the shikimate pathway [48]. The phenylalanine ammonia-lyase (*PAL*)-dependent phenylpropanoid route produces the byproduct of CQA isomers that assist in protein adducts [39, 49, 50]. *PAL* catalyzes the conversion of Phe to trans-cinnamic acid, which is the gateway to the PPP that leads to various derivative-like flavonoids, isoflavonoids, coumarin, anthocyanins, and lignins. This ultimately results in the synthesis of CQA and caffeic acid in plants, which respond differentially to biotic or abiotic stresses [11, 51]. Duplicated *PAL* unigenes in leaves and flowers have 120 and 76 copies, respectively, which was higher than Cynara species (Asteraceae) (102 *PAL* genes) [52, 53]. Overall, with the improvement of *PAL* activity and stability, they can be used as an efficient biocatalyst for the enzymatic production of Phe. *PAL*, *C4H*, *4CL*, *HCT*, and *HQT* are important enzymes in this pathway, facilitating the biosynthesis of flavonoids and other important secondary metabolites from

phenylalanine. In addition, there are 27 and 41 *4CL* unigene copies in the flower and leaf of *A. spathulifolius.* The *4CL* protein was found at its highest levels after the *PAL* enzymes, which play important roles in PP biosynthesis and belong to the cytochrome P-450 family. A higher level of chlorogenic acid production was observed based on the *4CL* activity even towards coumaric acid, ferulic acid, and cinnamic acid [33, 54, 55]. *4CL* might also participate in the biosynthesis of flavonoids and other soluble phenolic compounds [56]. This duplication of *PAL* and *4CL* highlights the importance of PP biosynthesis in *A. spathulifolius*. The *C3'H* enzyme belongs to the p450 monooxygenases family of the CYP98 family. This enzyme does not use p-coumaric acid as a substrate; it uses the shikimate (*HCT*) and quinic (*HQT*) ester of p-coumaric acid instead. This process catalyzes the 3'-hydroxylation of p-coumaric esters of shikimic/quinic acids to form CQA [18, 46, 57]. The duplication of *C3'H* 169 copies in the leaf and 140 unigenes in the flower indicates the involvement of Caffeoyl CoA in the synthesis of CQAs and other flavonoids. The up-or-down regulation of phenylpropanoid biosynthesis through the leaf and flower transcriptome shows 704 unigenes up regulated in the leaf and 413 unigenes in the flower (Fig 2b). Duplication of the PPP-involved unigenes suggests that the complexity of the molecular mechanism of PP biosynthesis in *A. spathulifolius* to produce various byproducts. A comparison of the leaf and flower DEGs revealed 482 unigenes equally upregulated than downregulated (142) in *A. spathulifolius*. This duplication of the core components unigenes suggests that they may have been recruited for major plant phenylpropanoid metabolites for their specialized tissues in *A. spathulifolius*. Hence PP biosynthesis depends strongly in an enzyme-dependent manner and accumulates in the flower and leaf of *A. spathulifolius.* The DEGs of HCs (HXXXD) in the leaf and flower show the binding sites of shikimate and quinic acid, which have competition sites to synthesize CQA or caffeic acid in PPP. The superfamily of BADH is a large class of acyl CoA-dependent transferase, which is distinct between the conserved domains. The HXXXDG and DFGWG consensus sequences are highly conserved among plants. There are two conserved domains in BAHDs: A C-terminal 'DFGWG', which may or may not be present in all BAHD, but it provides structural stability of enzymes; HXXXD, which has a solid role in catalyzing the acyltransferase group [58, 59]. *A. spathulifolius* BAHDs retain a functional copy of the HXXXD conserved site was predicted to have leaf (82) and flower (71). Two HCT proteins contain the same motif HHAAD in the middle part. In contrast, three HQT share the motif HTLS/AD. The two motifs of HXVVD also show a close relation to HQT (HTLXD) proteins, showing that although they do not share the same amino acid conserved motif, they share other conserved codons to the HQT protein with 78.7% of pairwise identity.

The HQT enzymes conserved the domain with HTLSD and HTLAD compared to Cynara and Helianthus of the Asteraceae family. Furthermore, two new domains, 'HXVVD', are the same as HQT with the phylogeny tree and alignment of codons. A *H. annuus* study revealed a large amount of CQAs in the sprout [43], where *C. cardunculus* whole-genome mapping showed that among 32 BAHDs proteins, only three of them showed similarity to the HCT and HQT enzymes [60]. In vivo studies of CQA synthesis in *C. intybus* revealed two HCT and three HQT genes involved in *HCT* or *HQT* with *C3'H*

to synthesize Caffeoyl CoA[61]. The HTLSD and HTLAD (HQT) sites were predicted in the leaf and flower *of A. spathulifolius*, which is the same as in *C. cardunclus* L, var. scolymus. In that study, 69 BAHDs proteins with the catalytic site of 'HXXXD' were identified genome-wide with three HQT and two *HCT* proteins. The new domain 'H(Y/R) VVD' may also be one of the *HQT* enzymes, but more analysis will be needed for confirmation. HCT and HQT are directly involved in CQA biosynthesis in tobacco [20]. Performing the quantitative PCR on HQT and HCT unigenes showed that HQT was expressed dramatically compared to the HCT unigenes. These duplication and new clade of BAHDs super family may indicates the PSM is essential to the endurance and successive reproductive fitness of a green plant species in its natural habitat. These results highlight the evolutionary relationships and conserved domain concurrence among the different HQT and HCT isoforms in the Asteraceae family plants, such as chicory and globe artichoke. Based on the transcriptome-wide identification and characterization of CQA in *A. spathulifolius*, the candidate genes involved in the core caffeoylquinic acid synthesis pathway (*PAL*, *C4H*, *4CL*, *HCT*, *C3'H*, *HQT*) were highly distributed and duplicated throughout the genome. Therefore, the diverged domain of the BAHDs gene complex could affect CQA synthesis in *A. spathulifolius* leaves and flowers. In addition, the HQT is strongly expressed in flowers, and *HCT* is expressed in the leaves. These metabolites are a diverse set of compounds with numerous functions in plant-environment interactions. The results suggest that PSM increases the diversity of higher plants, including *A. spathulifolius*. Overall, these results revealed substantial CQA biosynthesis in *A. spathulifolius* that could benefit human health.

4. Materials and Methods

4.1. *RNA Isolation, cDNA Library Construction, and Illumina Sequencing*

The total RNA of the whole flower of *Aster spathulifolius* was extracted using the protocol reported by Bretial *et al*. [62]. cDNA Library construction and RNA-Sequencing were performed using the Genomics Macrogen Laboratory (South Korea). The TruSeq method was used to make short fragments of mRNA. The short fragments were then as used as templates for the cDNA library. All short fragments were linked to the sequencing adapter, and the fragments were sequenced for Paired-End (PE) reads using illumina sequencing. For leaf transcriptome, the material used in previous studies (SRR10724565) [44] and flower RNA-sequences submitted under NCBI-SRA database with accession no: SRR14001926 were applied.

4.2. *Denovo assembly and functional annotation*

The raw reads were checked for Fastq quality control [63]. Below-quality value ≤ 30% (Q20) reads were removed using the trimmomatic tool [64]. The clean reads were assembled to retrieve unique transcripts using the trinity program with *kmer* size 25 with the following pipeline: inchworm, chrysalis, and butterfly [65]. Finally, the unique reads transcripts were checked for coding function annotations using the 'Trinotate and TrinotateWeb' pipeline, as mentioned in a previous study [66]. The gene annotation for all assembled unigenes was aligned to the Swiss-Prot protein database (https://www.uniprot.org/), and the Kyoto Encyclopedia of Genes and Genomes (KEGG,

http://www.genome.jp/kegg/kegg2.html) database. HMMER[67] was used to analyze the complete protein family (Pfam) [68]. The Gene Ontology was analyzed by Trinotate pipeline (https://github.com/Trinotate/Trinotate.github.io/wiki) and DAVID tools (https://david.ncifcrf.gov/tools.jsp) to retrieve the Biological process (BP) term. The quality of the transcripts coding completeness was analyzed using the BUSCO through tblastn alignments against the eudicots database lineage [69].

4.3. *Identification of PPP unigenes*

The peptide sequence of the Arabidopsis thaliana database of PPP protein was downloaded from the TAIR website [70, 71]. The identity of the PPP unigenes was retrieved [63] using local blastX at a value of 1×10^5 using the NCBI-Blast tool [72] and against the KEGG database [73] and the BioCyc for pathway prediction (https://biocyc.org/ARA/organism-summary?object=ARA). In addition, the PPP involved unigenes of *A. spathulifolius* were set as a database (as a reference database) to blast against the total transcripts to predict the unrecognized sequences for reverification.

4.4. *Differentially expressed genes (DEGs) analysis*

Transcripts of *Aster spathulifolius*, differential gene expressions (DEGs) of the abundance unigenes were determined by RSEM (RNA-Seq by Expectation-Maximization) [74, 75], which first generates and processes the total transcripts and then aligns them to raw reads of *A. spathulifolius* flower and leaf. RSEM is used to calculate the fragments per kilobase per million (FPKM) and the transcripts per million (TPM) values of the total unigenes to analyze the maximum number of genes expressed or abundant in *A. spathulifolius*. The genes were defined as DEGs with the options of FDR-corrected $p < 0.01$ and log2FC > 1. Both FPKM and TPM values were calculated to estimate the expression levels of the unigenes involved in the phenylpropanoid pathway (PPP) and caffeoylquinic acid biosynthesis. Complete DEGs analysis was performed under the Trinity pipeline [65] of RSEM with the option of Bowtie2 [75, 76] and EdgeR [77]. The thresholds for considering the significance of DEGs were FDR ≤ 0.01 and |log2 (fold change)| ≥ 1, which is the visualization of result under ggplot2 (https://ggplot2.tidyverse.org/reference/geom_point.html).

4.5. *Structure of BAHD family member unigenes*

The protein sequences of BAHD from *Cynara*, *Cichorium* from Asteraceae (same family), and Solanum [20, 61, 78, 79] plants were chosen to identify the HCT and HQT domain and predict the conserved domain of (HXXXD) of HCT and HQT under many BAHD family member proteins. The BAHD structural variance was differentiated by aligning the protein sequences by MAFFT[80] with the default parameters and the PHYML (LG) maximum likelihood tree with 100 bootstrap replications [81].

4.6. *qRT-PCR expression studies:*

Three stages of the leaf (young, extended, and mature leaf), a whole flower, and root were used to quantify the CQA biosynthesis-involved unigenes in *A. spathulifolius*. The PP pathway to CQA involving the unigenes from transcriptome leaf data was identified through KEGG annotation and BLAST-p analysis. The unigenes were then searched for the full open reading frame (ORF) to design the forward and reverse primers for further analysis. An amplicon size and primer structure of seven candidate unigenes and random

BAHDs unigenes were selected to quantify the HQT genes (**Table S1**). cDNA was synthesized by Reverse Transcription System (A3500, Promega, South Korea), the final 10 μl product was incubated at 70° C for 5 min to synthesize the c-DNA using Oligo(dT)$_{15}$ Primer. Finally, a quantification assay was carried out using GoTaq® qPCR Master Mix (A6001, Promega) followed by standard cycling conditions as a guideline (Applied Biosystems 7500 step one plus).

6. Patents

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1.   1.   Arnold, P. A.; Kruuk, L. E. B.; Nicotra, A. B., How to analyse plant phenotypic plasticity in response to a changing climate. New Phytologist 2019, 222, (3), 1235-1241.
2.   2.   Austen, N.; Walker, H. J.; Lake, J. A.; Phoenix, G. K.; Cameron, D. D., The Regulation of Plant Secondary Metabolism in Response to Abiotic Stress: Interactions Between Heat Shock and Elevated CO2. Frontiers in Plant Science 2019, 10.
3.   3.   Berini, J. L.; Brockman, S. A.; Hegeman, A. D.; Reich, P. B.; Muthukrishnan, R.; Montgomery, R. A.; Forester, J. D., Combinations of Abiotic Factors Differentially Alter Production of Plant Secondary Metabolites in Five Woody Plant Species in the Boreal-Temperate Transition Zone. Frontiers in Plant Science 2018, 9.
4.   4.   Yang, L.; Wen, K. S.; Ruan, X.; Zhao, Y. X.; Wei, F.; Wang, Q., Response of Plant Secondary Metabolites to Environmental Factors. Molecules 2018, 23, (4).
5.   5.   Carrington, Y.; Guo, J.; Le, C. H.; Fillo, A.; Kwon, J.; Tran, L. T.; Ehlting, J., Evolution of a secondary metabolic pathway from primary metabolism: shikimate and quinate biosynthesis in plants. Plant J 2018, 95, (5), 823-833.
6.   6.   Isah, T., Stress and defense responses in plant secondary metabolites production. Biol Res 2019, 52.
7.   7.   Biala, W.; Jasinski, M., The Phenylpropanoid Case - It Is Transport That Matters. Front Plant Sci 2018, 9, 1610.
8.   8.   Tholl, D., Biosynthesis and biological functions of terpenoids in plants. Adv Biochem Eng Biotechnol 2015, 148, 63-106.
9.   9.   Vogt, T., Phenylpropanoid biosynthesis. Mol Plant 2010, 3, (1), 2-20.
10.   10.   Bhattacharya, A.; Sood, P.; Citovsky, V., The roles of plant phenolics in defence and communication during Agrobacterium and Rhizobium infection. Mol Plant Pathol 2010, 11, (5), 705-19.
11.   11.   Yoo, H.; Widhalm, J. R.; Qian, Y.; Maeda, H.; Cooper, B. R.; Jannasch, A. S.; Gonda, I.; Lewinsohn, E.; Rhodes, D.; Dudareva, N., An alternative pathway contributes to phenylalanine biosynthesis in plants via a cytosolic tyrosine:phenylpyruvate aminotransferase. Nat Commun 2013, 4, 2833.
12.   12.   Maeda, H.; Dudareva, N., The shikimate pathway and aromatic amino Acid biosynthesis in plants. Annu Rev Plant Biol 2012, 63, 73-105.
13.   13.   Herrmann, K. M.; Weaver, L. M., The Shikimate Pathway. Annu Rev Plant Physiol Plant Mol Biol 1999, 50, 473-503.
14.   14.   Tzin, V.; Galili, G., The Biosynthetic Pathways for Shikimate and Aromatic Amino Acids in Arabidopsis thaliana. Arabidopsis Book 2010, 8, e0132.
15.   15.   Dixon, R. A.; Achnine, L.; Kota, P.; Liu, C. J.; Reddy, M. S.; Wang, L., The phenylpropanoid pathway and plant defence- a genomics perspective. Mol Plant Pathol 2002, 3, (5), 371-90.

16.   16.   Cle, C.; Hill, L. M.; Niggeweg, R.; Martin, C. R.; Guisez, Y.; Prinsen, E.; Jansen, M. A., Modulation of chlorogenic acid biosynthesis in Solanum lycopersicum; consequences for phenolic accumulation and UV-tolerance. Phytochemistry 2008, 69, (11), 2149-56.

17.   17.   El-Seedi, H. R.; El-Said, A. M.; Khalifa, S. A.; Goransson, U.; Bohlin, L.; Borg-Karlson, A. K.; Verpoorte, R., Biosynthesis, natural sources, dietary intake, pharmacokinetic properties, and biological activities of hydroxycinnamic acids. Journal of agricultural and food chemistry 2012, 60, (44), 10877-95.

18.   18.   Hoffmann, L.; Maury, S.; Martz, F.; Geoffroy, P.; Legrand, M., Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. J Biol Chem 2003, 278, (1), 95-103.

19.   19.   Meng, S.; Cao, J.; Feng, Q.; Peng, J.; Hu, Y., Roles of chlorogenic Acid on regulating glucose and lipids metabolism: a review. Evid Based Complement Alternat Med 2013, 2013, 801457.

20.   20.   Niggeweg, R.; Michael, A. J.; Martin, C., Engineering plants with increased levels of the antioxidant chlorogenic acid. Nat Biotechnol 2004, 22, (6), 746-54.

21.   21.   Mølgaard, P.; Ravn, H., Evolutionary aspects of caffeoyl ester distribution In Dicotyledons. Phytochemistry 1988, 27, (8), 2411-2421.

22.   22.   D'Auria, J. C., Acyltransferases in plants: a good time to be BAHD. Curr Opin Plant Biol 2006, 9, (3), 331-340.

23.   23.   Clifford, M. N., Chlorogenic acids and other cinnamates - nature, occurrence, dietary burden, absorption and metabolism. J Sci Food Agr 2000, 80, (7), 1033-1043.

24.   24.   Cheng, J. C.; Dai, F.; Zhou, B.; Yang, L.; Liu, Z. L., Antioxidant activity of hydroxycinnamic acid derivatives in human low density lipoprotein: Mechanism and structure-activity relationship. Food Chem 2007, 104, (1), 132-139.

25.   25.   Rahman, M. A.; Abdullah, N.; Aminudin, N., Antioxidative Effects and Inhibition of Human Low Density Lipoprotein Oxidation In Vitro of Polyphenolic Compounds in Flammulina velutipes (Golden Needle Mushroom). Oxid Med Cell Longev 2015, 2015, 403023.

26.   26.   Liang, N.; Kitts, D. D., Role of Chlorogenic Acids in Controlling Oxidative and Inflammatory Stress Conditions. Nutrients 2015, 8, (1).

27.   27.   Schroder, J., Probing plant polyketide biosynthesis. Nat Struct Biol 1999, 6, (8), 714-6.

28.   28.   Santana-Galvez, J.; Cisneros-Zevallos, L.; Jacobo-Velazquez, D. A., Chlorogenic Acid: Recent Advances on Its Dual Role as a Food Additive and a Nutraceutical against Metabolic Syndrome. Molecules 2017, 22, (3).

29.   29.   Abrankó, L.; Clifford, M. N., An Unambiguous Nomenclature for the Acyl-quinic Acids Commonly Known as Chlorogenic Acids. Journal of agricultural and food chemistry 2017, 65, (18), 3602-3608.

30.   30.   Miyamae, Y.; Kurisu, M.; Han, J.; Isoda, H.; Shigemori, H., Structure-activity relationship of caffeoylquinic acids on the accelerating activity on ATP production. Chem Pharm Bull (Tokyo) 2011, 59, (4), 502-7.

31.   31.   Shimoda, H.; Seki, E.; Aitani, M., Inhibitory effect of green coffee bean extract on fat accumulation and body weight gain in mice. BMC Complement Altern Med 2006, 6, 9.

32.   32.   Sun, J.; Song, Y. L.; Zhang, J.; Huang, Z.; Huo, H. X.; Zheng, J.; Zhang, Q.; Zhao, Y. F.; Li, J.; Tu, P. F., Characterization and quantitative analysis of phenylpropanoid amides in eggplant (Solanum melongena L.) by high performance liquid chromatography coupled with diode array detection and hybrid ion trap time-of-flight mass spectrometry. Journal of agricultural and food chemistry 2015, 63, (13), 3426-36.

33.   33.   Rigano, M. M.; Raiola, A.; Docimo, T.; Ruggieri, V.; Calafiore, R.; Vitaglione, P.; Ferracane, R.; Frusciante, L.; Barone, A., Metabolic and Molecular Changes of the Phenylpropanoid Pathway in Tomato (Solanum lycopersicum) Lines Carrying Different Solanum pennellii Wild Chromosomal Regions. Front Plant Sci 2016, 7, 1484.

34.   34.   Docimo, T.; Francese, G.; Ruggiero, A.; Batelli, G.; De Palma, M.; Bassolino, L.; Toppino, L.; Rotino, G. L.; Mennella, G.; Tucci, M., Phenylpropanoids Accumulation in Eggplant Fruit: Characterization of Biosynthetic Genes and Regulation by a MYB Transcription Factor. Front Plant Sci 2015, 6, 1233.

35.   35.   Awad, M. A.; de Jager, A.; van Westing, L. M., Flavonoid and chlorogenic acid levels in apple fruit: characterisation of variation. Sci Hortic-Amsterdam 2000, 83, (3-4), 249-263.

36.   36.   Xi, Y.; Cheng, D.; Zeng, X.; Cao, J.; Jiang, W., Evidences for Chlorogenic Acid--A Major Endogenous Polyphenol Involved in Regulation of Ripening and Senescence of Apple Fruit. PLoS One 2016, 11, (1), e0146940.

37.   37.   Cui, T.; Nakamura, K.; Ma, L.; Li, J. Z.; Kayahara, H., Analyses of arbutin and chlorogenic acid, the major phenolic constituents in Oriental pear. Journal of agricultural and food chemistry 2005, 53, (10), 3882-7.

38.   38.   Jeon, J. S.; Kim, H. T.; Jeong, I. H.; Hong, S. R.; Oh, M. S.; Yoon, M. H.; Shim, J. H.; Jeong, J. H.; Abd El-Aty, A. M., Contents of chlorogenic acids and caffeine in various coffee-related products. J Adv Res 2019, 17, 85-94.

39.   39.   Olthof, M. R.; Hollman, P. C.; Katan, M. B., Chlorogenic acid and caffeic acid are absorbed in humans. J Nutr 2001, 131, (1), 66-71.

40.   40.   Comino, C.; Hehn, A.; Moglia, A.; Menin, B.; Bourgaud, F.; Lanteri, S.; Portis, E., The isolation and mapping of a novel hydroxycinnamoyltransferase in the globe artichoke chlorogenic acid pathway. BMC Plant Biol 2009, 9, 30.

41.   41.   Schutz, K.; Kammerer, D.; Carle, R.; Schieber, A., Identification and quantification of caffeoylquinic acids and flavonoids from artichoke (Cynara scolymus L.) heads, juice, and pomace by HPLC-DAD-ESI/MS(n). Journal of agricultural and food chemistry 2004, 52, (13), 4090-6.

42. 42.   Gonzalez-Perez, S.; Merck, K. B.; Vereijken, J. M.; van Koningsveld, G. A.; Gruppen, H.; Voragen, A. G., Isolation and characterization of undenatured chlorogenic acid free sunflower (Helianthus annuus) proteins. Journal of agricultural and food chemistry 2002, 50, (6), 1713-9.

43. 43.   Cheevarungnapakul, K.; Khaksar, G.; Panpetch, P.; Boonjing, P.; Sirikantaramas, S., Identification and Functional Characterization of Genes Involved in the Biosynthesis of Caffeoylquinic Acids in Sunflower (Helianthus annuus L.). Front Plant Sci 2019, 10, 968.

44. 44.   Jean Claude, S.; Park, S., Aster spathulifolius Maxim. a leaf transcriptome provides an overall functional characterization, discovery of SSR marker and phylogeny analysis. PLoS One 2020, 15, (12), e0244132.

45. 45.   Xu, H.; Park, N. I.; Li, X.; Kim, Y. K.; Lee, S. Y.; Park, S. U., Molecular cloning and characterization of phenylalanine ammonia-lyase, cinnamate 4-hydroxylase and genes involved in flavone biosynthesis in Scutellaria baicalensis. Bioresour Technol 2010, 101, (24), 9715-22.

46. 46.   Mahesh, V.; Million-Rousseau, R.; Ullmann, P.; Chabrillange, N.; Bustamante, J.; Mondolot, L.; Morant, M.; Noirot, M.; Hamon, S.; de Kochko, A.; Werck-Reichhart, D.; Campa, C., Functional characterization of two p-coumaroyl ester 3'-hydroxylase genes from coffee tree: evidence of a candidate for chlorogenic acid biosynthesis. Plant Mol Biol 2007, 64, (1-2), 145-59.

47. 47.   Clifford, M. N.; Jaganath, I. B.; Ludwig, I. A.; Crozier, A., Chlorogenic acids and the acyl-quinic acids: discovery, biosynthesis, bioavailability and bioactivity. Nat Prod Rep 2017, 34, (12), 1391-1421.

48. 48.   Kong, J. Q., Phenylalanine ammonia-lyase, a key component used for phenylpropanoids production by metabolic engineering. Rsc Adv 2015, 5, (77), 62587-62603.

49. 49.   Tsai, C. J.; Harding, S. A.; Tschaplinski, T. J.; Lindroth, R. L.; Yuan, Y., Genome-wide analysis of the structural genes regulating defense phenylpropanoid metabolism in Populus. New Phytol 2006, 172, (1), 47-62.

50. 50.   Kalinowska, M.; Wroblewska, A. M.; Gryko, K., Chlorogenic acids in natural products and their antioxidant and antimicrobial properties. Przem Chem 2017, 96, (11), 2259-2264.

51. 51.   Weitzel, C.; Petersen, M., Enzymes of phenylpropanoid metabolism in the important medicinal plant Melissa officinalis L. Planta 2010, 232, (3), 731-42.

52. 52.   De Paolis, A.; Pignone, D.; Morgese, A.; Sonnante, G., Characterization and differential expression analysis of artichoke phenylalanine ammonia-lyase-coding sequences. Physiol Plant 2008, 132, (1), 33-43.

53. 53.   Zhang, X.; Liu, C. J., Multifaceted regulations of gateway enzyme phenylalanine ammonia-lyase in the biosynthesis of phenylpropanoids. Mol Plant 2015, 8, (1), 17-27.

54. 54.   Hamberger, B.; Hahlbrock, K., The 4-coumarate:CoA ligase gene family in Arabidopsis thaliana comprises one rare, sinapate-activating and three commonly occurring isoenzymes. Proc Natl Acad Sci U S A 2004, 101, (7), 2209-14.

55. 55.   Li, Y.; Kim, J. I.; Pysh, L.; Chapple, C., Four Isoforms of Arabidopsis 4-Coumarate: CoA Ligase Have Overlapping yet Distinct Roles in Phenylpropanoid Metabolism. Plant Physiol 2015, 169, (4), 2409-2421.

56. 56.   Shi, R.; Sun, Y. H.; Li, Q.; Heber, S.; Sederoff, R.; Chiang, V. L., Towards a systems approach for lignin biosynthesis in Populus trichocarpa: transcript abundance and specificity of the monolignol biosynthetic genes. Plant Cell Physiol 2010, 51, (1), 144-63.

57. 57.   Schoch, G.; Goepfert, S.; Morant, M.; Hehn, A.; Meyer, D.; Ullmann, P.; Werck-Reichhart, D., CYP98A3 from Arabidopsis thaliana is a 3 '-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway. Journal of Biological Chemistry 2001, 276, (39), 36566-36574.

58. 58.   Unno, H.; Ichimaida, F.; Suzuki, H.; Takahashi, S.; Tanaka, Y.; Saito, A.; Nishino, T.; Kusunoki, M.; Nakayama, T., Structural and mutational studies of anthocyanin malonyltransferases establish the features of BAHD enzyme catalysis. J Biol Chem 2007, 282, (21), 15812-22.

59. 59.   Haslam, T. M.; Manas-Fernandez, A.; Zhao, L. F.; Kunst, L., Arabidopsis ECERIFERUM2 Is a Component of the Fatty Acid Elongation Machinery Required for Fatty Acid Extension to Exceptional Lengths. Plant Physiol 2012, 160, (3), 1164-1174.

60. 60.   Menin, B.; Comino, C.; Moglia, A.; Dolzhenko, Y.; Portis, E.; Lanteri, S., Identification and mapping of genes related to caffeoylquinic acid synthesis in Cynara cardunculus L. Plant Sci 2010, 179, (4), 338-347.

61. 61.   Legrand, G.; Delporte, M.; Khelifi, C.; Harant, A.; Vuylsteker, C.; Morchen, M.; Hance, P.; Hilbert, J. L.; Gagneul, D., Identification and Characterization of Five BAHD Acyltransferases Involved in Hydroxycinnamoyl Ester Metabolism in Chicory. Front Plant Sci 2016, 7, 741.

62. 62.   Breitler, J. C.; Campa, C.; Georget, F.; Bertrand, B.; Etienne, H., A single-step method for RNA isolation from tropical crops in the field. Sci Rep 2016, 6, 38368.

63. 63.   Roser, L. G.; Aguero, F.; Sanchez, D. O., FastqCleaner: an interactive Bioconductor application for quality-control, filtering and trimming of FASTQ files. BMC Bioinformatics 2019, 20, (1), 361.

64. 64.   Bolger, A. M.; Lohse, M.; Usadel, B., Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014, 30, (15), 2114-20.

65. 65.   Grabherr, M. G.; Haas, B. J.; Yassour, M.; Levin, J. Z.; Thompson, D. A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q. D.; Chen, Z. H.; Mauceli, E.; Hacohen, N.; Gnirke, A.; Rhind, N.; di Palma, F.; Birren, B. W.; Nusbaum, C.; Lindblad-Toh, K.; Friedman, N.; Regev, A., Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 2011, 29, (7), 644-U130.

66.   66.   Bryant, D. M.; Johnson, K.; DiTommaso, T.; Tickle, T.; Couger, M. B.; Payzin-Dogru, D.; Lee, T. J.; Leigh, N. D.; Kuo, T. H.; Davis, F. G.; Bateman, J.; Bryant, S.; Guzikowski, A. R.; Tsai, S. L.; Coyne, S.; Ye, W. W.; Freeman, R. M., Jr.; Peshkin, L.; Tabin, C. J.; Regev, A.; Haas, B. J.; Whited, J. L., A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors. Cell Rep 2017, 18, (3), 762-776.

67.   67.   Johnson, L. S.; Eddy, S. R.; Portugaly, E., Hidden Markov model speed heuristic and iterative HMM search procedure. BMC Bioinformatics 2010, 11, 431.

68.   68.   El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S. R.; Luciani, A.; Potter, S. C.; Qureshi, M.; Richardson, L. J.; Salazar, G. A.; Smart, A.; Sonnhammer, E. L. L.; Hirsh, L.; Paladin, L.; Piovesan, D.; Tosatto, S. C. E.; Finn, R. D., The Pfam protein families database in 2019. Nucleic Acids Res 2019, 47, (D1), D427-D432.

69.   69.   Simao, F. A.; Waterhouse, R. M.; Ioannidis, P.; Kriventseva, E. V.; Zdobnov, E. M., BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 2015, 31, (19), 3210-2.

70.   70.   Huala, E.; Dickerman, A. W.; Garcia-Hernandez, M.; Weems, D.; Reiser, L.; LaFond, F.; Hanley, D.; Kiphart, D.; Zhuang, M.; Huang, W.; Mueller, L. A.; Bhattacharyya, D.; Bhaya, D.; Sobral, B. W.; Beavis, W.; Meinke, D. W.; Town, C. D.; Somerville, C.; Rhee, S. Y., The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. Nucleic Acids Res 2001, 29, (1), 102-5.

71.   71.   Swarbreck, D.; Wilks, C.; Lamesch, P.; Berardini, T. Z.; Garcia-Hernandez, M.; Foerster, H.; Li, D.; Meyer, T.; Muller, R.; Ploetz, L.; Radenbaugh, A.; Singh, S.; Swing, V.; Tissier, C.; Zhang, P.; Huala, E., The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucleic Acids Res 2008, 36, (Database issue), D1009-14.

72.   72.   Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T. L., BLAST+: architecture and applications. BMC Bioinformatics 2009, 10, 421.

73.   73.   Kanehisa, M.; Sato, Y., KEGG Mapper for inferring cellular functions from protein sequences. Protein Sci 2020, 29, (1), 28-35.

74.   74.   Wu, D. C.; Yao, J.; Ho, K. S.; Lambowitz, A. M.; Wilke, C. O., Limitations of alignment-free tools in total RNA-seq quantification. BMC Genomics 2018, 19, (1), 510.

75.   75.   Bray, N. L.; Pimentel, H.; Melsted, P.; Pachter, L., Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 2016, 34, (5), 525-7.

76.   76.   Langmead, B.; Salzberg, S. L., Fast gapped-read alignment with Bowtie 2. Nat Methods 2012, 9, (4), 357-9.

77.   77.   Robinson, M. D.; McCarthy, D. J.; Smyth, G. K., edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010, 26, (1), 139-40.

78.   78.   Comino, C.; Lanteri, S.; Portis, E.; Acquadro, A.; Romani, A.; Hehn, A.; Larbat, R.; Bourgaud, F., Isolation and functional characterization of a cDNA coding a hydroxycinnamoyltransferase involved in phenylpropanoid biosynthesis in Cynara cardunculus L. BMC Plant Biol 2007, 7, 14.

79.   79.   Scaglione, D.; Reyes-Chin-Wo, S.; Acquadro, A.; Froenicke, L.; Portis, E.; Beitel, C.; Tirone, M.; Mauro, R.; Lo Monaco, A.; Mauromicale, G.; Faccioli, P.; Cattivelli, L.; Rieseberg, L.; Michelmore, R.; Lanteri, S., The genome sequence of the outbreeding globe artichoke constructed de novo incorporating a phase-aware low-pass sequencing strategy of F1 progeny. Sci Rep 2016, 6, 19427.

80.   80.   Katoh, K.; Standley, D. M., MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 2013, 30, (4), 772-80.

81.   81.   Guindon, S.; Dufayard, J. F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O., New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 2010, 59, (3), 307-21.