

Genome-wide analysis of BAHD gene family and in vivo characterization of HQT-like genes involved in chlorogenic acid synthesis in tobacco (*Nicotiana tabacum* L.)

Lin Xia

x504107743@163.com

China Tobacco Gene Research Center, Zhengzhou Tobacco Research Institute of CNTC

Weiwei Song

407902360@qq.com

China Tobacco Hunan Industrial Co., Ltd

Zhengtai Liu

2024204039@stu.njau.edu.cn

Jiangsu Province and Chinese Academy of Sciences

Yangwei Mu

diolatp@gmail.com

China Tobacco Gene Research Center, Zhengzhou Tobacco Research Institute of CNTC

Jie Li

LJ734820531@163.com

Jiangsu Province and Chinese Academy of Sciences

Mengru Liu

liumengru54@163.com

China Tobacco Gene Research Center, Zhengzhou Tobacco Research Institute of CNTC

Junping Gao

junpinggao@163.com

China Tobacco Hunan Industrial Co., Ltd

Zefeng Li

379359757@qq.com

China Tobacco Gene Research Center, Zhengzhou Tobacco Research Institute of CNTC

Pingping Liu

Liu_pingping2012@163.com

China Tobacco Gene Research Center, Zhengzhou Tobacco Research Institute of CNTC

Zhong Wang

zhongwang1123@126.com

China Tobacco Gene Research Center, Zhengzhou Tobacco Research Institute of CNTC

Jun Yang

yangjun@ztri.com.cn

China Tobacco Gene Research Center, Zhengzhou Tobacco Research Institute of CNTC

Shixiao Xu

xushixiao@henau.edu.cn

Henan Agricultural University

Sheng Xu

xusheng@cnbg.net

Jiangsu Province and Chinese Academy of Sciences

Mingzhu Wu

wumz@ztri.com.cn

China Tobacco Gene Research Center, Zhengzhou Tobacco Research Institute of CNTC

Research Article

Keywords: Tobacco, BAHD gene family, HQT, CGA, abiotic and biotic stresses

DOI: <https://doi.org/>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Genome-wide analysis of *BAHD* gene family and *in vivo* characterization of *HQT*-like genes involved in chlorogenic acid synthesis in tobacco (*Nicotiana tabacum* L.)

Lin Xia^{1, 3}, Weiwu Song², Zhengtai Liu⁴, Yangwei Mu¹, Jie Li⁴, Mengru Liu¹, Junping Gao², Zefeng Li¹, Pingping Liu¹, Zhong Wang¹, Jun Yang¹, Shixiao Xu^{3, *}, Sheng Xu^{4 *}, Mingzhu Wu^{1, *}

1 China Tobacco Gene Research Center, Zhengzhou Tobacco Research Institute of CNTC, Zhengzhou 450001, China

2 China Tobacco Hunan Industrial Co., Ltd., Changsha 410007, China

3 College of Tobacco Science, Henan Agricultural University, Zhengzhou 450002, China

4 Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing, 210014, China

Corresponding author: Mingzhu Wu, wumz@ztri.com.cn

Abstract

In plants, the diverse enzymes known as BAHD acyltransferases use acyl coenzyme A (acyl-CoA) as their substrate and fulfill crucial biological functions. In this study, 94 *NtBAHD* genes were identified

24 in the tobacco (*Nicotiana tabacum*) genome. These genes displayed
25 considerable diversity in their physicochemical properties,
26 phylogenetic relationships, gene structures, and conserved motifs.
27 Chlorogenic acid (CGA), which is the predominant polyphenol found
28 in tobacco leaves, is mainly synthesised via the hydroxycinnamoyl-
29 CoA:quinate hydroxycinnamoyl transferase (HQT) pathway. Four
30 *HQT* genes (*NtHQT1-NtHQT4*) were identified in the *BAHD* gene
31 family of *N. tabacum*. Of these, *NtHQT1* and *NtHQT2* were highly
32 expressed in leaves, while *NtHQT3* and *NtHQT4* were root-specific.
33 Following heterologous expression *in vitro*, three of the four HQT
34 proteins (*NtHQT2*, *NtHQT3*, and *NtHQT4*) were obtained in soluble
35 form; however, only *NtHQT2* and *NtHQT3* exhibited enzymatic
36 activity. Downregulation of *NtHQT1*, *NtHQT2*, *NtHQT3*,
37 and *NtHQT4* significantly reduced both total polyphenol and
38 chlorogenic acid (CGA) content, whereas the levels of rutin,
39 scopoletin, and lignin remained unchanged. Furthermore, the
40 expression of *NtHQT* genes was differentially induced under abiotic
41 and biotic stresses, including drought, cold, and *Phytophthora*
42 *nicotianae* infection. In contrast, their responses to various
43 phytohormones exhibited distinct expression profiles. These results
44 suggest that whereas *in vitro* assays identified *NtHQT2* and *NtHQT3*
45 as the catalytically active isoforms, all four *NtHQT* genes contribute

to chlorogenic acid (CGA) biosynthesis and environmental adaptation.

Trial registration This study does not involve clinical trials or human participants and, as such, does not require clinical trial registration.

Keywords: Tobacco; *BAHD* gene family; HQT; CGA; abiotic and biotic stresses

Introduction

The growth, development, and defense mechanisms of plants are profoundly affected by secondary metabolites [1]. The acylation of substrates containing nitrogen and oxygen to produce amides and esters is a prevalent and biologically significant process for the modification of secondary metabolites [2]. The BAHD acyltransferase superfamily, characterized by the use of coenzyme A (CoA) thioesters as acyl donors and amines or alcohols as acceptors, is essential for the biosynthesis of diverse plant secondary metabolites, including esters, anthocyanins, terpenoids, and lignin monomers [2, 3]. The acronym BAHD is derived from the combination of initials representing the initial four enzymes that have been biochemically characterized: benzylalcohol O-acetyltransferase (BEAT) synthesizes the floral ester benzyl acetate [4]; anthocyanin O-hydroxycinnamoyl transferase (AHCT) plays a crucial

68 role in catalyzing anthocyanin synthesis [5]; N-hydroxycinnamoyl
69 anthranilate benzoyl transferase (HCBT) is pivotal in initiating
70 dianthramide biosynthesis [6]; Deacetylvindoline 4-O-
71 acetyltransferase (DAT) is involved in the last stage of producing the
72 alkaloid vindoline [7]. The BAHD family encompasses a diverse array
73 of enzymes with varied functions, exhibiting considerable
74 divergence in their primary sequences, with only a 10–30% amino
75 acid similarity. However, it is noteworthy that the majority of
76 enzymes within the BAHD family possess two conserved motifs:
77 HXXXD and DFGWG [2, 3]. Most characterized BAHD enzymes
78 exhibit both motifs; however, deletions or modifications of either
79 motif results in a substantial reduction in enzymatic activity [8].

80 The *BAHD* genes are essential for producing suberin, wax, cutin,
81 and polyphenols, which are crucial for plant defense mechanisms [9,
82 10]. Chlorogenic acid (CGA) is widely present in the human diet and
83 is one of the most abundant polyphenols, which are primarily
84 sourced from coffee [11], tea [12], fruits [13], vegetables [14-16],
85 and medicinal plants [17, 18]. CGA is recognized for its strong
86 antioxidant properties, specifically, its capacity to defend against
87 and potentially mitigate DNA damage [19, 20]. Therefore, CGA
88 exhibits several therapeutic effects, such as its ability to act as an
89 antioxidant, antihypertensive, anti-inflammatory, hepatoprotective,

cardioprotective, antibacterial, antiviral activities, and anti-obesity [21].

The biosynthesis of CGA is a multi-step process within the phenylpropanoid pathway, relying on the coordinated action of several key enzymes [16]. HQT, a BAHD-family enzyme, is responsible for the biosynthesis of CGA from its precursors, quinic acid and caffeoyl-CoA. In solanaceous plants, this conversion is widely recognized as the principal pathway for the synthesis of CGA [14, 16, 18]. Alternative pathways include *p*-coumarate 3'-hydroxylase (C3H) and hydroxyl cinnamoyl CoA:shikimate/quinic acid hydroxycinnamoyl transferase (HCT). However, despite the expression of C3H and HCT in *Arabidopsis*, CGA is absent, thereby challenging this hypothesis [16, 18]. A potential alternative pathway for CGA biosynthesis exists, in which hydroxyl cinnamoyl D-glucose: quinate hydroxycinnamoyl transferase (HCGQT) utilizes caffeoyl-D-glucose as a substrate [16].

The BAHD family has been identified and studied across various species through genetic and biochemical research, including *Arabidopsis* and *populus* [22], pear [23], cacao [24], banana [25], onion [26], and *Taxus mairei* [27]. However, information on tobacco, an economically important crop and model plant, is limited [28]. Therefore, this study aims to characterize the *BAHD* gene family in

112 tobacco and to clarify the role of the *HQT* genes in chlorogenic acid
113 biosynthesis and stress responses.

114 **Materials and methods**

115 **Plant materials**

116 Common tobacco (*Nicotiana tabacum*) cultivar K326 seeds were
117 acquired from Guizhou Academy of Tobacco Science (Guiyang, China)
118 and treated with 10% sodium hypochlorite for surface sterilization.
119 After rinsing the seeds several times with distilled water, they were
120 planted in 1/2 MS medium for germination. Transplanting and
121 cultivation conditions for tobacco seedlings followed previously
122 described methods [29]. Once the seedlings developed 5–6 true
123 leaves, they were transferred to open fields to facilitate further
124 growth. Various plant tissues at the flowering stage were collected
125 for RNA extraction, including lateral roots, fibrous roots, stems, old
126 leaves, middle leaves, new leaves, leaf veins, axillary buds, flower
127 buds, calyces, crowns, pistils, and stamens.

128 Tobacco seedlings (3–5 leaves) cultivated under controlled growth
129 conditions, were subjected to drought (10 d), cold (4 °C, 24 h), and
130 dark (7 d) stresses. *Phytophthora nicotianae* Breda de Haan
131 zoospores were prepared and inoculated as previously described
132 [30]. The tobacco seedling roots were placed in 10 µM abscisic acid

133 (ABA), 10 μ M gibberellin A3 (GA3), 50 μ M methyl jasmonate (MeJA),
134 or 100 μ M salicylic acid (SA) solution containing 1/2 Hoagland
135 solution for a period of 2 h. Seedlings treated with only the 1/2
136 Hoagland solution served as controls.

137 **Identification and classification of NtBAHDs**

138 Candidate genes of the BAHD family were identified from the
139 tobacco database (<http://lifenglab.hzau.edu.cn/Nicomics/index.php>)
140 using a Hidden Markov Model search in TBtools, with the
141 characteristic PF02458 domain as the query and an E-value
142 threshold of 10^{-5} . Domain presence was confirmed using Pfam
143 (<http://pfam-legacy.xfam.org/>) and NCBI CD-search
144 (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>).

145 Furthermore, the characteristic conserved motifs (HXXXD and
146 DFGWG) were verified via the online tool MEME ([http://meme-](http://meme-suite.org/tools/meme)
147 [suite.org/tools/meme](http://meme-suite.org/tools/meme)). Sequences containing the HXXXD and
148 DFGWG motifs were included in the study due to their crucial role in
149 BAHD enzyme functionality [23]. The length of amino acids,
150 molecular weight (MW), and isoelectric point (pI) were analyzed
151 using ExPASy (<https://www.expasy.org/>).

152 **Chromosome mapping**

153 Physical maps for each tobacco chromosome were acquired from

Nicomics (<http://lifenglab.hzau.edu.cn/Nicomics/>). These maps provided information on the length of the chromosomes, number of *NtBAHDs*, and their start/end sites. Subsequently, simple maps were created to visually depict the distribution of *NtBAHDs* across tobacco chromosomes.

Phylogeny, gene structures, and conserved proteins of BAHD

A Maximum Likelihood phylogenetic tree was constructed in MEGA12 based on the protein sequences, supported by 1,000 bootstrap replicates, and visualized using iTOL v5 [31]. To determine the locations of the exons and introns, sequence alignment between each BAHD genomic DNA sequence and its corresponding cDNA sequence was performed. Ten conserved motifs (6–30 amino acids wide) were detected in the BAHD proteins using MEME Suite [32].

Gene duplication analyses of NtBAHDs

The MCScanX algorithm in TBtools was applied to identify gene duplication events within the NtBAHD family [33]. Collinear blocks were identified both within the allotetraploid *N. tabacum* genome and between *N. tabacum* and *A. thaliana* to assess intra-species and inter-genome homology. The syntenic relationships were visualized using the Advanced Circos module in TBtools, facilitating the interpretation of gene duplication patterns and evolutionary

divergence.

Cloning and quantitative real-time PCR (qRT-PCR) analysis of *NtHQT* genes

The coding sequences of NtHQT1-4 were cloned by PCR amplification from cDNA isolated from tobacco leaves and roots with gene-specific primers (Supplementary Table S1). Total RNA isolation and first-strand cDNA synthesis were performed following established protocols [29]. TB Green kit (TaKaRa Bio Inc.) was used for qRT-PCR analysis with a Bio-Rad CFX96 instrument (Bio-Rad, Laboratories, Hercules, CA, USA). The reference gene used in the study was *NtL25* [34]. The primers used for qRT-PCR are listed in Supplementary Table S1.

RNA-Seq analysis

The RNA-seq data were obtained from the NCBI database (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA208209>). The processing of mRNA-Seq data was conducted in accordance with the methods described [35]. Gene expression levels and differential gene expressions (DEGs) were calculated and identified using Cufflinks (v2.2.1) and Cuffdiff 2, respectively. TB-Tools were utilized to create a heat map illustrating gene expression levels [36].

Subcellular localization analysis of NtHQTs

The four *NtHQT* genes were cloned into the pCAMBIA1300-GFP plasmid using the Cloning Kit (Wuhan Genenode Biotechnology Co., Ltd.) to create N-terminal green fluorescent protein (GFP)-tagged NtHQT proteins. After sequence confirmation, we transformed *Agrobacterium tumefaciens* LBA4404 with the 35S::NtHQT1-GFP, 35S::NtHQT2-GFP, 35S::NtHQT3-GFP, 35S::NtHQT4-GFP, and empty 35S::GFP vectors. Tobacco (*Nicotiana benthamiana*) plants aged 4–5 weeks were chosen for subcellular localization experiments. *Agrobacterium* infection of *N. benthamiana* was conducted following previous methods [37]. Observations and photography were conducted with a Carl Zeiss AG LSCM 800 confocal laser-scanning microscope (Germany).

Structural modeling, molecular docking and enzyme activity analysis of NtHQTs

AlphaFold3 was used to generate 3D structural models of NtHQTs using established methods [38]. A homology model was constructed based on the crystal structure of *Helianthus annuus* HQT3. Subsequent model optimization and assessment were carried out using the Modeler program and the Profiles-3D module integrated into Discovery Studio 4.5 (Accelrys Inc., San Diego, CA, USA). This software suite was further used for 3-D visualization and molecular

217 docking simulations involving caffeoyl-CoA and quinic acid as
218 ligands.

219 **Prokaryotic expression and purification of NtHQTs**

220 Using a One Step Cloning Kit (Beijing Creative Spark Technology Co.,
221 Ltd.), the coding sequences of NtHQTs were directionally cloned into
222 a 6×His-tagged pET-28a(+) vector to generate recombinant
223 plasmids. After sequence verification, the resulting constructs (pET-
224 28a-NtHQT1 to -NtHQT4) were transformed into *Escherichia coli*
225 BL21(DE3) cells. Transformants were cultured in LB-kanamycin
226 medium at 37 °C until the OD₆₀₀ reached 0.6–0.8. Protein expression
227 was then induced with 0.1 mM IPTG at 15 °C for 12 h. Cells were
228 harvested by centrifugation (4,000 × g, 15 min, 4 °C) and stored at -
229 20 °C. The His-tagged proteins were purified following a published
230 protocol and analyzed by SDS-PAGE [39].

231 **Enzymatic activity assays for recombinant NtHQTs**

232 The enzymatic activity of NtHQT was assayed as described with
233 modifications [14]. The 60 µL reaction mixture, containing 100 mM
234 sodium phosphate buffer (pH 7.5), 1 mM EDTA, purified enzyme (1–
235 50 µg), and substrates (50 µM–10 mM), was initiated by enzyme
236 addition and incubated at 37 °C for 3 h. Reactions were quenched
237 with an equal volume of HPLC solvent [0.1% formic acid, 5%

238 acetonitrile], filtered through a 0.22 µm membrane, and analyzed by
239 HPLC.

240 The catalytic properties of purified NtHQT toward quinate and
241 caffeoyl CoA were characterized by assessing its optimal
242 temperature (25–60 °C) and pH (5.0–9.0, using NaAC-HAC,
243 phosphate, and Tris-HCl buffers). Kinetic parameters were
244 determined by varying substrate concentrations: for the K_m of
245 quinate, reactions contained 5 µg enzyme, 0.5 mM caffeoyl CoA, and
246 0.05–10 mM quinate; for caffeoyl CoA, they contained 5 µg enzyme,
247 4 mM quinate, and 0.05–2 mM caffeoyl CoA. The K_m and V_{max}
248 values were calculated from Michaelis–Menten plots fitted using
249 Origin 8.5 (OriginLab, Northampton, MA, USA).

250 **Obtaining NtHQTs virus-induced gene silencing (VIGS)** 251 **transgenic tobacco**

252 Transgenic tobacco plants exhibiting reduced expressions of
253 *NtHQTs* were generated through VIGS following previously reported
254 methods [29]. For plasmid construction, primers were designed with
255 gene-specific sequences and a 20-bp pTRV2 homology arm
256 (Supplementary Table S1). The resulting plasmids (pTRV2-NtHQT1,
257 pTRV2-NtHQT2, pTRV2-NtHQT3, pTRV2-NtHQT4) alongside control
258 vectors (pTRV1, pTRV2) were transformed into *A. tumefaciens*

259 LBA4404 using a freeze-thaw method. The subsequent knockdown
260 efficiency of NtHQTs was then quantified via qRT-PCR.

261 **Measurement of polyphenols**

262 Total phenolic content was quantified with the Folin-Ciocalteu
263 reagent and expressed as gallic acid equivalents [40]. CGA, rutin,
264 and scopoletin were analyzed using an Agilent 1260/6490 HPLC
265 system coupled to a triple quadrupole mass spectrometer, following
266 established procedures [35].

267 **Determination of lignin content**

268 The lignin assay kit (Sangon Biotech Co., Ltd.) was utilized for the
269 assessment of lignin content following the manufacturer's
270 instructions. TRACETM 1310-ISQ gas chromatography mass
271 spectrometry (GC-MS; Thermo Fisher Scientific Inc., Waltham, MA,
272 USA) was used to analyze the content of Syringyl (S), guaiacyl (G),
273 and *p*-hydroxyphenyl (H) lignin monomers, following established
274 methods [29, 41].

275 **Results**

276 **Genome-wide identification and chromosomal distribution of**
277 **BAHDs in tobacco**

278 Comprehensive analysis of the tobacco genome identified 212
279 putative BAHD acyltransferases (Supplementary data 1). Among
280 these, 94 contained the two conserved motifs (HXXXD and DFGWG;
281 Supplementary Figure S1) [42]. The characteristics of these genes
282 and their encoded proteins are summarized in Table 1, with CDS
283 lengths, protein lengths, and molecular weights ranging from 810-
284 2,961 bp, 269-986 amino acids, and 30.18-108.54 kDa, respectively.
285 Among these, 49 proteins were predicted to have a $pI < 6.5$,
286 suggesting that certain BAHD enzymes exhibit alkaline
287 characteristics, whereas others display acidic properties.

288 The NtBAHDs was renamed from NtBAHD1 to NtBAHD94 based
289 on its chromosomal location, starting from chromosome 1. In cases
290 where multiple genes are on the same chromosome, the first
291 positioned gene is renamed as “first.” Among the 94 *NtBAHD* genes,
292 93 were distributed across the chromosomes as follows: 11 were
293 found on chromosome 24, nine on chromosome 23, eight each on
294 chromosomes 4, and 20, seven on chromosome 16, five each on
295 chromosomes 5, 19, and 21, four each on chromosomes 6, and 17,
296 three each on chromosomes 1, 3, 9, 11, 14, and 15, and two each on
297 chromosomes 10, 13, 18, and 22. Only one *NtBAHD* gene was
298 identified on chromosome 12 (Figure 1). Additionally, one gene was
299 located on a chromosome scaffold (Table 1). These findings indicated

that *NtBAHD* genes are unevenly distributed in the tobacco genome.

Phylogenetic analyses of BAHDs

The phylogenetic relationships among the 94 tobacco and 63 *Arabidopsis* BAHD proteins were inferred using the Maximum Likelihood (ML) method in MEGA12. The resulting tree classified all members into five major clades (I-V; Figure 2). Clades II, IV, and V were further subdivided into two subclades each. Among them, Clade I consisted of five *AtBAHD* members, and Clade III contained two. Clade II comprised 45 *NtBAHDs*, distributed as three in subclade II-a and 42 in II-b. Clade IV included 27 *NtBAHDs*, with 11 in subclade IV-a and 16 in IV-b, while Clade V contained 22 *NtBAHDs*, of which seven were grouped into subclade V-a and 15 into V-b.

Gene structure and conserved motif analyses of *NtBAHDs*

The gain and loss of introns in *NtBAHD* genes exhibited diverse patterns among the three clades in the ML tree, with intron numbers ranging from zero to four in the same cluster (Figure 3A, B). Clade II-a genes uniformly contained a single intron. In contrast, the majority of Clade II-b genes were intronless, with the exceptions of seven genes possessing one intron and a single gene possessing four. Clade IV genes exhibited variation in both intron number (ranging from one to three) and spatial distribution. Similarly, 11 genes in

Clade V lacked introns, with nine containing one intron and two containing two.

By using MEME software, we identified ten conserved motifs in NtBAHD proteins (Figure 3C). All members of the NtBAHD family possess both motif 4 and motif 8, with the exception of NtBAHD92, which contains distinct sequences: GCCLSHKVG DGYS LWNFLNAWAELARGGS and LTISSLCRFPFYDADFGWGKP, respectively. These two motifs are associated with the HXXXD and DFGWG motifs, respectively. Notably, although NtBAHD92 lacks the canonical motif 4, it retains the essential HXXXD domain. The distribution of other motifs also varied across phylogenetic subclades. Motifs 9 and 10 were absent in all members of subclade II-a. In subclade II-b, most members contained the full complement of ten motifs, while a few possessed only five to nine. Furthermore, motif 10 was not detected in any members of subclades IV or V. The conserved motifs within the same subclades exhibited similar types and distributions, thereby providing further support for the classification of evolutionary trees.

Synteny analysis of NtBAHD family

Multiple duplication mechanisms, including whole-genome (WGD) and segmental, proximal (PD), tandem (TD), dispersed (DSD), and

transposed (TRD) duplications, collectively drove the evolutionary expansion of the *NtBAHD* gene family [23]. MCSanX analysis revealed distinct contributions of these mechanisms to gene family expansion: WGD or segmental duplications accounted for 68.09% (64/94) of duplication events, followed by PDs (12.77%, 12/94), TDs (12.77%, 12/94), and DSDs (6.38%, 6/94) (Supplementary Figure S2 and Table S2). These results demonstrate that WGD or segmental duplications serve as the major driver of *NtBAHD* gene family expansion, while PD, TD, DSD, and TRD collectively contribute to its diversification.

Additionally, synteny blocks were detected in the genome. As shown in Figure 4A, the 94 *BAHDs* of *N. tabacum* were distributed on 21 chromosomes, and 34 syntenic gene pairs in *N. tabacum* were revealed. The interspecific collinearity analysis between *N. tabacum* and *A. thaliana* identified 15 homologous gene pairs among the *NtBAHD* genes, suggesting conserved syntenic relationships between the two species (Figure 4B).

BAHDs related to CGA

Our analysis focused on Clade IV-a, a subgroup of hydroxycinnamoyltransferases implicated in the biosynthesis of hydroxycinnamoyl spermine and lignin monomeric intermediates.

whereas, NtBAHD33 and NtBAHD36 were excluded due to their distinct role as spermidine hydroxycinnamoyl transferases involved in the biosynthesis of various other secondary metabolites, including terpenoids, and alkaloids [2]. Among the remaining Clade I-b members (NtBAHD19, 23, 31, 34, 37, 56, 71, 87, and 88), only NtBAHD31, 34, 37, and 71 possessed the conserved HTLSD (HXXXD) and DEFEG motifs characteristic of HQT (Supplementary Figure S3) [10].

The genes *NtBAHD31*, *71*, *34*, and *37* (NCBI references: XM_009800234.1, XM_016628801.1, AJ582651, XM_009626002.3) were designated as NtHQT1, NtHQT2, NtHQT3, and NtHQT4, respectively. Sequence analysis revealed a high degree of pairwise similarity among these four NtHQTs. Specifically, NtHQT1 and NtHQT2 shared 95.04% identity, indicating strong conservation, while NtHQT3 and NtHQT4 showed a comparatively lower identity of 82.61%.

Expression analysis and subcellular localization of NtHQTs

The expression profiles of the *NtHQTs* at various stages of tobacco development were examined through RNA-seq (Figure 5A, Supplementary data 1). High expression of *NtHQT1* was detected in young leaves, whereas NtHQT2 was highly expressed in mature

flowers. The expression levels of *NtHQT3* and *NtHQT4* were significantly elevated in the roots. Each *NtHQT* gene expression was determined by qRT-PCR at the flowering stage. *NtHQT1* and *NtHQT2* exhibit elevated expression levels in leaves, especially in young and middle leaves (Figure 5A, B). However, *NtHQT3* and *NtHQT4* were mainly expressed in the lateral and fibrous roots (Figure 5C, D).

Subcellular localization of NtHQT1–NtHQT4 was assessed by infiltrating *N. benthamiana* leaves with agrobacteria harboring 35S:NtHQT1–4:GFP or a 35S:GFP control vector. The four NtHQT GFP-tagged proteins were observed to localize in both the nucleus and the cytoplasm, with nuclear localization confirmed by co-localization with DAPI staining. Additionally, these proteins accumulated at the cell periphery, exhibiting a distribution pattern similar to that of 35S:GFP (Figure 6).

Biochemical characterization of recombinant NtHQT enzymes

Given that extensive efforts to enhance the soluble expression of NtHQT1—including the optimization of induction conditions, the testing of alternative expression vectors, and fusion with a SUMO tag—yielded only insoluble protein (Supplementary Figure S4). NtHQT2, NtHQT3, and NtHQT4 were successfully expressed in a

405 prokaryotic system at concentrations ranging from 0.24 to 0.64
406 mg/mL. The expressed fusion proteins, with theoretical molecular
407 weights of 49.68, 49.29, and 42.36 kDa respectively, were confirmed
408 by SDS-PAGE and western blot analysis (Supplementary Figure S5).

409 Despite extensive optimization of key reaction parameters,
410 including protein concentration, buffer composition, temperature,
411 and pH, HQT4 exhibited no detectable catalytic activity. In contrast,
412 NtHQT2 and NtHQT3 displayed stronger activity. We therefore
413 conducted in vitro experiments to assess the pH dependence and
414 temperature sensitivity of their reduction activity (Figures 7A-D).
415 Specifically, NtHQT2 showed peak activity at pH 9.0 and 40 °C,
416 while NtHQT3 was most active at pH 8.5 and maintained high
417 activity over a broader temperature range (37-45 °C). The kinetic
418 parameters of the recombinant enzymes NtHQT2 and NtHQT3 were
419 determined with quinic acid and caffeoyl-CoA [14]. The two enzymes
420 displayed distinct substrate affinities: NtHQT2 exhibited a lower K_m
421 for quinic acid (0.35 mM) than for caffeoyl-CoA (0.91 mM), while
422 NtHQT3 showed the opposite preference, with a lower K_m for
423 caffeoyl-CoA (0.26 mM) than for quinic acid (0.85 mM, Figures 7E,
424 F). Consistent with this substrate preference, NtHQT3 demonstrated
425 significantly higher catalytic efficiency (k_{cat}/K_m), which was 5.21-

fold and 30.57-fold greater than that of NtHQT2 for quinic acid and
caffeoyl-CoA, respectively (Figures 7G, H).

Predicted protein structures of NtHQTs

Homology modeling of NtHQT1, NtHQT2, NtHQT3, and NtHQT4 was
performed using AlphaFold3, followed by substrate docking with
quinic acid and caffeoyl-CoA. The four models and the corresponding
docking structures exhibited a high degree of similarity. The
reaction cavity of NtHQT1, NtHQT2, NtHQT3, and NtHQT4 contains
two distinct intervals, which specifically bind to quinic acid and
caffeoyl-CoA, respectively (Figure 8). The protein sequence
surrounding the cavity exhibits a relatively high degree of
conservation. Take HQT1 as an example, upon the entry of quinic
acid (orange) into the upper reaction site, TYR-404 and ASN-370
facilitate its binding through hydrogen bonding interactions.
Subsequently, the caffeoyl group of caffeoyl-CoA (red) enters and
occupies the lower reaction site, where it is stabilized by π -bonding
with THR38, along with support from adjacent sites. This interaction
ultimately leads to the formation of chlorogenic acid as a product of
the reaction. The protein sequence retained by NtHQT1, NtHQT2,
and NtHQT3 (yellow region) further restricts the reaction cavity,
positioning the substrate quinine acid in a more reactive location,

thereby enhancing the efficiency of the reaction (Supplementary Figure S6A-C). In model NtHQT4, the absence of the yellow protein sequence creates a cavity in the upper part of the reaction chamber (Supplementary Figure S6D). Quinic acid can detach from the reaction site through this missing position, while caffeoyl-CoA deviates from its original binding site due to the lack of corresponding residue (THR38). This results in differences in activity compared to other HQT variants.

NtHQT modulate the biosynthesis of CGA

Transgenic tobacco plants were generated in which NtHQT1, NtHQT2, NtHQT3, and NtHQT4 were silenced using VIGS. However, these plants exhibited no obvious phenotypic alterations (Figure 9A). Furthermore, qRT-PCR analysis confirmed the silencing efficacy, demonstrating significant downregulation of NtHQT1, NtHQT2, NtHQT3, and NtHQT4 in leaves (Supplementary Figure S7). The total polyphenol content in wild-type (WT) tobacco leaves was measured at 209.86 µg/g. In comparison, the polyphenol content in VIGS-NtHQT1, VIGS-NtHQT2, VIGS-NtHQT3, and VIGS-NtHQT4 leaves exhibited a decrease ranging from 66.41% to 72.25% (Figure 9B). Compared to the control group, the CGA content in the leaves of VIGS-NtHQT1, VIGS-NtHQT2, VIGS-NtHQT3, and VIGS-NtHQT4

showed significant decreases of 62.02%, 66.25%, 64.64%, and 71.33%, respectively (Figure 9C). The rutin and scopoletin content in the leaves of VIGS-NtHQT1, VIGS-NtHQT2, VIGS-NtHQT3, and VIGS-NtHQT4 plants did not exhibit a significant difference (Figure 9D, E).

Lignin content and G, H, and S lignin monomers in the NtHQT1, NtHQT2, NtHQT3, and NtHQT4-silenced plants were also analyzed. No H monomers were detected in any of the samples (Table 2). The lignin content, the levels of G and S units, did not exhibit significant changes when compared to the WT tobacco as well as the S/G ratio. These findings indicate that NtHQT1, NtHQT2, NtHQT3, and NtHQT4 are essential for CGA biosynthesis.

Diverse roles of NtHQTs in abiotic and biotic stress responses

To investigate the potential functions of *NtHQTs* in plant regulatory networks, 1500-bp upstream regions of these two genes were defined as putative promoters. Cis-regulatory elements within these promoters were analyzed using the PlantCare database [43]. The identified elements were systematically categorized into three major groups: light signaling, phytohormone response, and biotic/abiotic stress responses (Figure 10A). Fourteen light-responsive elements were detected, among which the G-Box and GT1-motif were present

in all four *NtHQT* promoters, suggesting a crucial role in light-mediated transcription. Phytohormone response elements comprised six types, including ABA-responsive elements (ABRE and O₂-site), JA/SA-related elements (TCA-motif and TGA-motif), and GA-responsive elements (GARE-motif, and TATC-box). Specific cis-elements mediate responses to different environmental stresses: LTR (low-temperature responsive element) to cold stress, ARE (antioxidant response element) to oxidative stress, GGTCA-motif to pathogen infection, and TC-rich repeats to pathogen/wounding stress. Additionally, several MYB-binding elements such as MBS and MRE were identified, known to participate in stress responses, light signaling, and phenylpropanoid-derived metabolism..

To functionally link the cis-elements with potential gene regulation, we further analyzed the transcriptional profiles of *NtHQT* genes in response to diverse stress and hormone stimuli. Under various stress conditions, *NtHQT1* and *NtHQT2* displayed similar expression patterns, with both genes up-regulated under cold and drought stress while down-regulated in response to darkness and *P. nicotianae* infection (Figure 10B). Similarly, *NtHQT3* and *NtHQT4* showed comparable expression profiles, as both were induced by cold stress and *P. nicotianae* infection but repressed under dark conditions. Furthermore, *NtHQT4* expression was also up-regulated

under drought stress. These results highlight the distinct roles of *NtHQTs* in tobacco's response to biotic and abiotic stress. As shown in Figure 10C, the expression of *NtHQT1* was significantly up-regulated following treatment with abscisic acid (ABA), gibberellin (GA), and salicylic acid (SA). In contrast, *NtHQT2* expression was upregulated only after SA treatment, whereas methyl jasmonate (MeJA) treatment led to a significant downregulation of *NtHQT2*. The expression of *NtHQT3* was induced following treatments with ABA, GA, SA, and MeJA. Meanwhile, *NtHQT4* expression was upregulated in response to GA and MeJA treatments.

Discussion

Tobacco is widely recognized as a vital economic crop and is frequently used as a model system for investigating the metabolism of secondary metabolites, including alkaloids, terpenoids, phenylpropanoids [44]. The acylation of secondary metabolites represents a widespread and biologically important modification in plants [2]. BAHD enzymes are key contributors to the acetylation process of various plant secondary metabolites [2, 3, 10]. In this study, 94 *BAHDs* were identified in tobacco. The identified *NtBAHDs* exhibited significant diversity in gene structure, chromosomal distribution, physicochemical properties, and putative functions, consistent with previous findings regarding the *BAHD* family [2, 3,

10].

The tobacco and *Arabidopsis* NtBAHD family was categorized into five clades: I, II, III, IV, and V. This classification aligns with the results observed for BAHDs in *Arabidopsis* and pears [23, 45]. Clade I includes the *Arabidopsis* gene *AT4G24510.1* (*CER2*), which regulates cuticular wax biosynthesis [46]. Clade II-b is characterized by 42 NtBAHDs that catalyze the acetylation of small- to medium-chain and aromatic alcohols [45]. The functional roles of *Arabidopsis* genes in Clade IV-a involve the biosynthesis of hydroxycinnamoyl spermine and lignin monomeric intermediates, including shikimate hydroxycinnamoyl transferase activity [14, 41, 47]. Clade IV-b has been functionally linked to the modification of aromatic and aliphatic alcohols [45, 48]. In Clade V, the two *Arabidopsis* genes *AT3G29590* (*At5MAT*) and *AT1G03940* (*At3AT1*) are established contributors to anthocyanin biosynthesis [2, 49]. Furthermore, an analysis of the conserved motifs in NtBAHDs showed that their type and distribution were similar within the same subfamily. This observation further supported the authenticity of the classification results.

Different gene duplication patterns make distinct contributions to the expansion of gene families [50]. In this study, WGD or segmental duplication (68.09%) significantly contributed to the expansion of

BAHDs in tobacco, while other modes of replication played a relatively minor role. WGD, TD, and DSD play a pivotal role in eukaryotic genome evolution by facilitating the genesis of novel functions within genetic systems [51, 52]. The majority of gene expansion in the *Arabidopsis* pedigree is believed to be due to WGDs, which accounts for approximately 90% of the observed expansion [53]. Notably, WGD and DSD were key factors in expanding *BAHDs* in apple, and different varieties of pears [23].

As the key enzyme catalyzing the final step of chlorogenic acid (CGA) biosynthesis [14, 16], HQT is widely conserved and essential for CGA production in diverse species, including artichoke, tobacco, potato, and tomato. A finding consistently observed across these species is that the suppression of HQT expression results in markedly reduced CGA levels [14, 15, 16, 18]. *NtHQT1* and *NtHQT2* were highly expressed in tobacco leaves, while *NtHQT3* and *NtHQT4* showed elevated expression in roots. Similarly, potato *HQT* and *Taraxacum antungense* *TaHQT2* were predominantly expressed in leaves, whereas *TaHQT1* demonstrates peak expression levels in roots [16, 54]. In globe artichoke, *HQT1*, *HQT2*, and *HQT3* all displayed low expression in leaves, with *HQT3* showed high expression levels in stems [10]. In contrast, the *Lonicera japonica* *HQT* gene was most highly expressed in flower buds, followed by

leaves. Notably, the tissue-specific distribution of HQT expression correlates with chlorogenic acid (CGA) accumulation patterns across these species [47].

Unlike the primarily cytosolic localization of most BAHD members [2, 5], the four NtHQT-GFP fusion proteins displayed localization to both the cytoplasm and the nucleus. A similar localization pattern was also reported for the HQT2-GFP fusion protein in globe artichoke and the GFP-TaHQT2 fusion in *Taraxacum antungense* [10, 54]. Typically, molecules weighing less than 40 kDa can easily enter and exit the nucleus through nuclear pores [55]. However, the size of the four NtHQT-GFP fusion proteins is too large for passive nuclear import, implying the requirement for active nuclear localization signal (NLS)-dependent transport through the nuclear pore complex [56]. Computational analysis using WoLF PSORT predicted that NtHQT1, NtHQT2, NtHQT3, and NtHQT4 all contain putative nuclear localization signals (data not shown). This suggests that the four NtHQT-GFP fusion proteins likely enter the nucleus via an active, signal-dependent mechanism. To date, however, there is no clear biochemical evidence for chlorogenic acid (CGA) accumulation within the nucleus.

Among the four investigated HQT proteins, only NtHQT2, NtHQT3, and NtHQT4 were successfully obtained as soluble forms; of these,

only NtHQT2 and NtHQT3 displayed detectable enzymatic activity. The optimal pH for NtHQT2 and NtHQT3 was determined to be 9.0 and 8.5, respectively, aligning with the pH optimum of 8.8 reported for the recombinantly expressed rice serotonin N-acetyltransferase [57]. Notably, NtHQT3 exhibited a significantly higher catalytic efficiency than NtHQT2, yet its efficiency was comparable to that of the well-characterized tomato SlHQT [14]. Structural analysis via homology modeling suggested that the loss of activity in NtHQT4 may be due to a cavity above the active site, formed by the absence of a specific protein sequence, which could permit premature substrate release. Collectively, these findings suggest that NtHQT3 plays a more prominent role in chlorogenic acid biosynthesis in tobacco.

CGA, rutin, and scopoletin represent the major polyphenols in tobacco leaves, all biosynthesized through the phenylpropanoid pathway [58]. Leaves of VIGS-silenced NtHQTs plants exhibited significantly reduced levels of both CGA and total polyphenols, consistent with the established role of HQT in phenolic metabolism [16, 18]. In contrast, no significant changes were observed in the contents of rutin or scopoletin. Conversely, in tomato, knockout of HQT resulted in decreased CGA accumulation, while leading to elevated levels of hydroxycinnamoyl-glucose and flavonoids [59].

Furthermore, the leaves of NtHQT-silenced lines showed no significant alterations in total lignin content or in the levels of G, H, and S monomers. This finding is consistent with previous reports that HQT suppression does not impact lignin biosynthesis in tobacco [18]. Collectively, these results indicate that the metabolic flux diverted from CGA biosynthesis was channeled toward the production of phenylpropanoid compounds other than rutin, scopoletin or lignin.

CGA, the predominant polyphenol in solanaceous plants, plays a protective role against biotic and abiotic stresses [14, 59, 60]. This function is supported by studies across species: in tomato, *slhqt* mutants with reduced CGA accumulation exhibit enhanced sensitivity to UV-B irradiation [59], while cold and drought stress induce the expression of certain HQTs in sweet potato [60]. Consistent with these reports, our findings demonstrate that NtHQTs are significantly upregulated in response to cold, drought, and *P. nicotianae* infection. Furthermore, their expression is closely associated with key stress-related phytohormones such as ABA, GA, SA, and MeJA [22, 61, 62]. In line with previous findings that ABA and GA alleviate biotic and abiotic stresses by modulating plant metabolism [63], our results implicate NtHQTs in both chlorogenic acid (CGA) biosynthesis and stress response pathways.

Conclusions

This study identified 94 *BAHD* genes in the *N. tabacum* genome, from which four putative *NtHQT* genes were pinpointed through phylogenetic and conserved motif analyses. Consistent with the established HQT-dependent pathway for CGA biosynthesis in solanaceous plants, expression profiling showed that NtHQT1 and NtHQT2 were highly expressed in leaves, while NtHQT3 and NtHQT4 were predominantly expressed in roots. *In vitro* enzyme assays confirmed that both NtHQT2 and NtHQT3 use quinic acid and caffeoyl-CoA as substrates, with NtHQT3 exhibiting the highest catalytic efficiency. Functional studies demonstrated that downregulating the four *NtHQT* genes significantly reduced total polyphenol and CGA content, whereas the levels of rutin, scopoletin, lignin, and lignin monomers were largely unaffected. Furthermore, the expression of different *NtHQT* genes was variably induced by cold, drought, and *P. nicotianae* infection, and displayed distinct responses to various phytohormones. In summary, our findings provide a comprehensive characterization of the NtBAHD family and underscore the central role of four NtHQT enzymes in CGA biosynthesis and stress adaptation. Future research should focus on identifying the phenylpropanoid derivatives that accumulate upon the redirection of metabolic flux following NtHQT suppression.

Supplementary Information

The online version contains supplementary material available at

Acknowledgements

We gratefully acknowledge all our colleagues for their technical assistance and valuable discussions.

Authors' contributions

L. X. data curation, investigation, writing-original draft. W. S. investigation, writing-original draft. Z. L. investigation. Y. M. investigation. J. L. investigation. M. L. investigation. J. G. investigation. Z. L. investigation. P. L. investigation. Z. W. data curation, investigation. J. Y. conceptualisation, supervision, investigation. S. X. Review & editing. M. W. conceptualization and supervision. The final version of the manuscript has been approved by all authors.

Funding

This work was supported by Major scientific and technological projects of CNTC (110202101043(JY-20)), the Natural Science Foundation of Henan Province (Grant No. 232300420032), the Science and Technology Project of Henan Province (Grant No.222102110164, No. 242102111147). The funding bodies played no role in the design of the study and collection, analysis, interpretation of data, and in writing the manuscript.

Data availability

688 The data generated and/or analyzed during this study have been
689 deposited in the GenomeSequence Archive (GSA) database at the
690 National Genomics Data Center, ChinaNational Center for
691 Bioinformation. The dataset is accessible at:
692 <https://ngdc.cncb.ac.cn/omix/submitList>. The DNA and RNA
693 sequences are accessible under OMIX ID: OMIX012777, and the
694 DNA and RNA sequencing data under OMIX ID: OMIX012776.

695 **Declarations**

696 **Ethics approval and consent to participate**

697 Not applicable.

698 **Consent for publication**

699 Data is provided within the manuscript or supplementary information files.

700 **Competing interests**

701 The data are provided in the manuscript or as part of the supplementary
702 information.

References

1. Huang XQ, Dudareva N. Plant specialized metabolism. *Curr Biol.* 2023; 33 (11): R473–8. <https://doi.org/10.1016/j.cub.2023.01.057>.
2. D'Auria JC. Acyltransferases in plants: a good time to be BAHD. *Curr Opin Plant Biol.* 2006; 9 (3): 331–40. <https://doi.org/10.1016/j.pbi.2006.03.016>.
3. Xu D, Wang Z, Zhuang W, Wang T, Xie Y. Family characteristics, phylogenetic reconstruction, and potential applications of the plant BAHD acyltransferase family. *Front Plant Sci.* 2023; 14: 1218914. <https://doi.org/10.3389/fpls.2023.1218914>.
4. Dudareva N, D'Auria JC, Nam KH, Raguso RA, Pichersky E. Acetyl-CoA:benzylalcohol acetyltransferase--an enzyme involved in floral scent production in *Clarkia breweri*. *Plant J.* 1998; 14 (3): 297–304. <https://doi.org/10.1046/j.1365-313x.1998.00121.x>.
5. Fujiwara H, Tanaka Y, Yonekura-Sakakibara K, Fukuchi-Mizutani M, Nakao M, Fukui Y, Yamaguchi M, Ashikari T, Kusumi T. cDNA cloning, gene expression and subcellular localization of anthocyanin 5-aromatic acyltransferase from *Gentiana triflora*. *Plant J.* 1998; 16 (4): 421–31. <https://doi.org/10.1046/j.1365-313x.1998.00312.x>.
6. Yang Q, Reinhard K, Schiltz E, Matern U. Characterization and heterologous expression of hydroxycinnamoyl/benzoyl-CoA:anthranilate ase from elicited cell cultures of carnation, *Dianthus caryophyllus* L. *Plant Mol Biol.* 1997; 35 (6): 777–89. <https://doi.org/10.1023/a:1005878622437>.

- 725 7. St-Pierre B, Laflamme P, Alarco AM, De Luca V. The terminal O-
726 acetyltransferase involved in vindoline biosynthesis defines a new class of
727 proteins responsible for coenzyme A-dependent acyl transfer. *Plant J.* 1998;
728 14 (6): 703–13. <https://doi: 10.1046/j.1365-313x.1998.00174.x>.
- 729 8. Bayer A, Ma X, Stöckigt J. Acetyltransfer in natural product biosynthesis-
730 functional cloning and molecular analysis of vinorine synthase. *Bioorg Med*
731 *Chem.* 2004; 12 (10): 2787–95. <https://doi.org/10.1016/j.bmc.2004.02.029>.
- 732 9. Moglia A, Lanteri S, Comino C, Hill L, Knevitt D, Cagliero C, Rubiolo P,
733 Bornemann S, Martin C. Dual catalytic activity of hydroxycinnamoyl-
734 coenzyme A quinate transferase from tomato allows it to moonlight in the
735 synthesis of both mono- and dicaffeoylquinic acids. *Plant Physiol.* 2014;
736 166 (4): 1777–87. <https://doi.org/10.1104/pp.114.251371>.
- 737 10. Moglia A, Acquadro A, Eljounaidi K, Milani AM, Cagliero C, Rubiolo P,
738 Genre A, Cankar K, Beekwilder J, Comino C. Genome-wide identification
739 of BAHD acyltransferases and In vivo characterization of HQT-like
740 enzymes involved in caffeoylquinic acid synthesis in globe artichoke. *Front*
741 *Plant Sci.* 2016; 7: 1424. <https://doi.org/10.3389/fpls.2016.01424>.
- 742 11. Lepelley M, Cheminade G, Tremillon N, Simkin A, Caillet V, McCarthy J.
743 Chlorogenic acid synthesis in coffee: an analysis of CGA content and real-
744 time RT-PCR expression of *HCT*, *HQT*, *C3H1*, and *CCoAOMT1* genes
745 during grain development in *C. canephora*. *Plant Sci.* 2007; 172 (5): 978–
746 96. <https://doi.org/10.1016/j.plantsci.2007.02.004>.

- 747 12. Hayakawa S, Ohishi T, Miyoshi N, Oishi Y, Nakamura Y, Isemura M. Anti-
748 cancer effects of green tea epigallocatechin-3-gallate and coffee
749 chlorogenic acid. *Molecules*. 2020; 25 (19): 4553.
750 <https://doi.org/10.3390/molecules25194553>.
- 751 13. Meinhart AD, Damin FM, Caldeirão L, de Jesus Filho M, da Silva LC, da
752 Silva Constant L, Filho JT, Wagner R, Godoy HT. Chlorogenic and caffeic
753 acids in 64 fruits consumed in Brazil. *Food Chem*. 2019; 286: 51-63.
754 <https://doi.org/10.1016/j.foodchem.2019.02.004>.
- 755 14. Niggeweg R, Michael AJ, Martin C. Engineering plants with increased
756 levels of the antioxidant chlorogenic acid. *Nat Biotechnol*. 2004; 22 (6):
757 746-54. <https://doi.org/10.1038/nbt966>.
- 758 15. Comino C, Hehn A, Moglia A, Menin B, Bourgaud F, Lanteri S, Portis E.
759 The isolation and mapping of a novel hydroxycinnamoyltransferase in the
760 globe artichoke chlorogenic acid pathway. *BMC Plant Biol*. 2009; 9: 30.
761 <https://doi.org/10.1186/1471-2229-9-30>.
- 762 16. Payyavula RS, Shakya R, Sengoda VG, Munyaneza JE, Swamy P, Navarre
763 DA. Synthesis and regulation of chlorogenic acid in potato: Rerouting
764 phenylpropanoid flux in HQT-silenced lines. *Plant Biotechnol J*. 2015; 13
765 (4): 551-64. <https://doi.org/10.1111/pbi.12280>.
- 766 17. Chaowuttikul C, Palanuvej C, Ruangrunsi N. Pharmacognostic
767 specification, chlorogenic acid content, and in vitro antioxidant activities
768 of *Lonicera japonica* flowering bud. *Pharmacognosy Res*. 2017; 9 (2): 128-

- 769 32. https://doi.org/10.4103/pr.pr_117_16.
- 770 18. Cardenas CL, Costa MA, Laskar DD, Moinuddin SGA, Lee C, Davin LB,
 771 Lewis NG. RNAi modulation of chlorogenic acid and lignin deposition in
 772 *Nicotiana tabacum* and insufficient compensatory metabolic cross-talk. *J*
 773 *Nat Prod.* 2021; 84 (3): 694–706.
 774 <https://doi.org/10.1021/acs.jnatprod.1c00054>.
- 775 19. Kweon MH, Hwang HJ, Sung HC. Identification and antioxidant activity of
 776 novel chlorogenic acid derivatives from bamboo (*Phyllostachys edulis*). *J*
 777 *Agric Food Chem.* 2001; 49 (10): 4646–55.
 778 <https://doi.org/10.1021/jf010514x>.
- 779 20. Miao M, Xiang L. Pharmacological action and potential targets of
 780 chlorogenic acid. *Adv Pharmacol.* 2020; 87: 71–88.
 781 <https://doi.org/10.1016/bs.apha.2019.12.002>.
- 782 21. Naveed M, Hejazi V, Abbas M, Kamboh AA, Khan GJ, Shumzaid M, Ahmad
 783 F, Babazadeh D, FangFang X, Modarresi-Ghazani F, WenHua L, XiaoHui
 784 Z. Chlorogenic acid (CGA): A pharmacological review and call for further
 785 research. *Biomed Pharmacother.* 2018; 97: 67–74.
 786 <https://doi.org/10.1016/j.biopha.2017.10.064>.
- 787 22. Yu X, Zhang W, Zhang Y, Zhang X, Lang D, Zhang X. The roles of methyl
 788 jasmonate to stress in plants. *Funct Plant Biol.* 2019; 46 (3): 197–212.
 789 <https://doi.org/10.1071/FP18106>.
- 790 23. Liu C, Qiao X, Li Q, Zeng W, Wei S, Wang X, Chen Y, Wu X, Wu J, Yin H,

- 791 Zhang S. Genome-wide comparative analysis of the BAHD superfamily in
792 seven Rosaceae species and expression analysis in pear (*Pyrus*
793 *bretschneideri*). BMC Plant Biol. 2020; 20 (1): 14.
794 <https://doi.org/10.1186/s12870-019-2230-z>.
- 795 24. Abdullah, Faraji S, Heidari P, Poczai P. The *BAHD* gene family in Cacao
796 (*Theobroma cacao*, Malvaceae): genome-wide identification and
797 expression analysis. Front Ecol Evol. 2021; 9: 707708.
798 <https://doi.org/10.3389/fevo.2021.707708>.
- 799 25. Xu Y, Tie W, Yan Y, Xu B, Liu J, Li M, Yang J, Zeng J, Hu W, Jin Z.
800 Identification and expression of the BAHD family during development,
801 ripening, and stress response in banana. Mol Biol Rep. 2021; 48 (2): 1127–
802 38. <https://doi.org/10.1007/s11033-020-06132-9>.
- 803 26. Liu L, Xu H, Zhang W, Xing J, Zhu M, Zhang Y, Wang Y. Genome-wide
804 analysis of the BAHD family in welsh onion and CER2-LIKEs involved in
805 wax metabolism. Genes (Basel). 2023; 14(6): 1286.
806 <https://doi.org/10.3390/genes14061286>.
- 807 27. Xu D, Wang Z, Zhuang W, Zhang F, Xie Y, Wang T. Genome-wide
808 identification and expression pattern analysis of BAHD acyltransferase
809 family in *Taxus mairei*. Int J Mol Sci. 2024; 25 (7): 3777.
810 <https://doi.org/10.3390/ijms25073777>.
- 811 28. Sun H, Sun X, Wang H, Ma X. Advances in salt tolerance molecular
812 mechanism in tobacco plants. Hereditas. 2020; 157 (1): 5.

<https://doi.org/10.1186/s41065-020-00118-0>.

29. Wu M, Li Y, Liu Z, Xia L, Xiang Y, Zhao L, Yang X, Li Z, Xie X, Wang L, Wang R, Xu S, Yang J. Genome-wide identification of the CAD gene family and functional analysis of putative bona fide CAD genes in tobacco (*Nicotiana tabacum* L.). Front Plant Sci. 2024; 15: 1400213.

<https://doi.org/10.3389/fpls.2024.1400213>.

30. Zhang C, Feng C, Zheng Y, Wang J, Wang F. Root exudates metabolic profiling suggests distinct defense mechanisms between resistant and susceptible tobacco cultivars against black shank disease. Front Plant Sci. 2020; 11: 559775. <https://doi.org/10.3389/fpls.2020.559775>.

31. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021; 49(W1): W293–6. <https://doi.org/10.1093/nar/gkab301>.

32. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. Nucleic Acids Res. 2015; 43(W1): W39–49. <https://doi.org/10.1093/nar/gkv416>.

33. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, Kissinger JC, Paterson AH. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012; 40 (7): e49. <https://doi.org/10.1093/nar/gkr1293>.

34. Schmidt GW, Delaney SK. Stable internal reference genes for normalization of real-time RT-PCR in tobacco (*Nicotiana tabacum*) during development and abiotic stress. Mol. Genet. Genomics. 2010; 283 (3): 233–

- 835 41. <https://doi.org/10.1007/s00438-010-0511-1>.
- 836 35. Wang Z, Ma L, Liu P, Luo Z, Li Z, Wu M, Xu X, Pu W, Huang P, Yang, J.
837 Transcription factor NtWRKY33a modulates the biosynthesis of
838 polyphenols by targeting *NtMYB4* and *NtHCT* genes in tobacco. Plant Sci.
839 2023; 326: 111522. <https://doi.org/10.1016/j.plantsci.2022.111522>.
- 840 36. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools:
841 an integrative toolkit developed for interactive analyses of big biological
842 data. Mol Plant. 2020; 13 (8): 1194–202.
843 <https://doi.org/10.1016/j.molp.2020.06.009>.
- 844 37. Li X, Tao N, Xu B, Xu J, Yang Z, Jiang C, Zhou Y, Deng M, Lv J, Zhao K.
845 Establishment and application of a root wounding-immersion method for
846 efficient virus-induced gene silencing in plants. Front. Plant Sci. 2024; 15:
847 1336726. <https://doi.org/10.3389/fpls.2024.1336726>.
- 848 38. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R,
849 Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. SWISS-
850 MODEL: homology modelling of protein structures and complexes. Nucleic
851 Acids Res. 2018; 46: W296–303. <https://doi: 10.1093/nar/gky427>.
- 852 39. Sun B, Wang P, Wang R, Li Y, Xu, S. Molecular Cloning and
853 Characterization of a meta/para-O-Methyltransferase from *Lycoris aurea*.
854 Int J Mol Sci. 2018; 19: 1911. <https://doi: 10.3390/ijms19071911>.
- 855 40. Ainsworth EA, Gillespie KM. Estimation of total phenolic content and other
856 oxidation substrates in plant tissues using Folin-Ciocalteu reagent. Nat

857 Protoc. 2007; 2 (4): 875–7. <https://doi.org/10.1038/nprot.2007.102>.

858 41. Besseau S, Hoffmann L, Geoffroy P, Lapierre C, Pollet B, Legrand M.

859 Flavonoid accumulation in Arabidopsis repressed in lignin synthesis

860 affects auxin transport and plant growth. Plant Cell. 2007; 19 (1): 148–62.

861 <https://doi.org/10.1105/tpc.106.044495>.

862 42. Molina I, Kosma D. Role of HXXXD-motif/BAHD acyltransferases in the

863 biosynthesis of extracellular lipids. Plant Cell Rep. 2015; 34 (4): 587–601.

864 <https://doi.org/10.1007/s00299-014-1721-5>.

865 43. Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouzé

866 P, Rombauts S. PlantCARE, a database of plant cis-acting regulatory

867 elements and a portal to tools for in silico analysis of promoter sequences.

868 Nucleic Acids Res. 2002; 30 (1): 325–7.

869 <https://doi.org/10.1093/nar/30.1.325>.

870 44. Matros A, Amme S, Kettig B, Buck-Sorlin GH, Sonnewald U, Mock HP.

871 Growth at elevated CO₂ concentrations leads to modified profiles of

872 secondary metabolites in tobacco cv. SamsunNN and to increased

873 resistance against infection with potato virus Y. Plant Cell Environ. 2006;

874 29 (1): 126–37. <https://doi.org/10.1111/j.1365-3040.2005.01406.x>.

875 45. Yu XH, Gou JY, Liu CJ. BAHD superfamily of acyl-CoA dependent

876 acyltransferases in Populus and Arabidopsis: bioinformatics and gene

877 expression. Plant Mol Biol. 2009; 70 (4): 421–22.

878 <https://doi.org/10.1007/s11103-009-9482-1>.

- 879 46. Legrand G, Delporte M, Khelifi C, Harant A, Vuylsteker C, Mörchen M,
880 Hance P, Hilbert JL, Gagneul D. Identification and characterization of five
881 BAHD acyltransferases involved in hydroxycinnamoyl ester metabolism in
882 chicory. *Front Plant Sci.* 2016; 7: 741.
883 <https://doi.org/10.3389/fpls.2016.00741>.
- 884 47. Peng X, Li W, Wang W, Bai G. Cloning and characterization of a cDNA
885 coding a hydroxycinnamoyl-CoA quinate hydroxycinnamoyl transferase
886 involved in chlorogenic acid biosynthesis in *Lonicera japonica*. *Planta Med.*
887 2010; 76 (16): 1921–26. <https://doi.org/10.1055/s-0030-1250020>.
- 888 48. Kosma DK, Molina I, Ohlrogge JB, Pollard M. Identification of an
889 Arabidopsis fatty alcohol:caffeoyl-Coenzyme A acyltransferase required
890 for the synthesis of alkyl hydroxycinnamates in root waxes. *Plant Physiol.*
891 2012; 160 (1): 237–48. <https://doi.org/10.1104/pp.112.201822>.
- 892 49. Luo J, Nishiyama Y, Fuell C, Taguchi G, Elliott K, Hill L, Tanaka Y,
893 Kitayama M, Yamazaki M, Bailey P, Parr A, Michael AJ, Saito K, Martin C.
894 Convergent evolution in the BAHD family of acyl transferases:
895 identification and characterization of anthocyanin acyl transferases from
896 Arabidopsis thaliana. *Plant J.* 2007; 50 (4): 678–95.
897 <https://doi.org/10.1111/j.1365-313X.2007.03079.x>.
- 898 50. Freeling M. Bias in plant gene content following different sorts of
899 duplication: tandem, whole-genome, segmental, or by transposition. *Annu*
900 *Rev Plant Biol.* 2009; 60: 433–53.

- 901 <https://doi.org/10.1146/annurev.arplant.043008.092122>.
- 902 51. Friedman R, Hughes AL. Pattern and timing of gene duplication in animal
 903 genomes. *Genome Res.* 2001; 11 (11): 1842-7.
 904 <https://doi.org/10.1101/gr.200601>.
- 905 52. Moore RC, Purugganan MD. The early stages of duplicate gene evolution.
 906 *Proc Natl Acad Sci U S A.* 2003; 100 (26): 15682-87.
 907 <https://doi.org/10.1073/pnas.2535513100>.
- 908 53. Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de
 909 Peer Y. Modeling gene and genome duplications in eukaryotes. *Proc Natl*
 910 *Acad Sci USA.* 2005; 102 (15): 5454-5459.
 911 <https://doi.org/10.1073/pnas.0501102102>.
- 912 54. Liu Q, Yao L, Xu Y, Cheng H, Wang W, Liu Z, Liu J, Cui X, Zhou Y, Ning W.
 913 In vitro evaluation of hydroxycinnamoyl CoA:quinate hydroxycinnamoyl
 914 transferase expression and regulation in *Taraxacum antungense* in
 915 relation to 5-caffeoylquinic acid production. *Phytochemistry.* 2019; 162:
 916 148-56. <https://doi.org/10.1016/j.phytochem.2019.02.014>.
- 917 55. Keminer O, Peters R. Permeability of single nuclear pores. *Biophys J.* 1999;
 918 77 (1): 17-28. [https://doi.org/10.1016/S0006-3495\(99\)76883-9](https://doi.org/10.1016/S0006-3495(99)76883-9).
- 919 56. Lin DH, Hoelz A. The structure of the nuclear pore complex (an update).
 920 *Annu Rev Biochem.* 2019; 88: 725-83. [https://doi.org/10.1146/annurev-](https://doi.org/10.1146/annurev-biochem-062917-011901)
 921 [biochem-062917-011901](https://doi.org/10.1146/annurev-biochem-062917-011901).
- 922 57. Kang K, Lee K, Park S, Byeon Y, Back K. Molecular cloning of rice serotonin

923 N-acetyltransferase, the penultimate gene in plant melatonin biosynthesis.
 924 J Pineal Res. 2013; 55(1):7–13. [https://doi: 10.1111/jpi.12011](https://doi.org/10.1111/jpi.12011).

925 58. Zou X, Bk A, Rauf A, Saeed M, Al-Awthan YS, A Al-Duais M, Bahattab O,
 926 Hamayoon Khan M, Suleria HAR. Screening of polyphenols in tobacco
 927 (*Nicotiana tabacum*) and determination of their antioxidant activity in
 928 different tobacco varieties. ACS Omega. 2021; 6 (39): 25361–71.
 929 <https://doi.org/10.1021/acsomega.1c03275>.

930 59. D'Orso F, Hill L, Appelhagen I, Lawrenson T, Possenti M, Li J, Harwood W,
 931 Morelli G, Martin C. Exploring the metabolic and physiological roles of
 932 HQT in *S. lycopersicum* by gene editing. Front Plant Sci. 2023; 14:
 933 1124959. <https://doi.org/10.3389/fpls.2023.1124959>.

934 60. Medison MB, Pan R, Peng Y, Medison RG, Shalmani A, Yang X, Zhang W.
 935 Identification of HQT gene family and their potential function in CGA
 936 synthesis and abiotic stresses tolerance in vegetable sweet potato. Physiol
 937 Mol Biol Plants. 2023; 29 (3): 361–76. [https://doi.org/10.1007/s12298-023-](https://doi.org/10.1007/s12298-023-01299-4)
 938 [01299-4](https://doi.org/10.1007/s12298-023-01299-4).

939 61. Gul N, Masoodi KZ, Ramazan S, Mir JI, Aslam S. Study on the impact of
 940 exogenously applied methyl jasmonate concentrations on *Solanum*
 941 *lycopersicum* under low temperature stress. BMC Plant Biol. 2023;
 942 23(1):437. [https://doi: 10.1186/s12870-023-04449-8](https://doi.org/10.1186/s12870-023-04449-8).

943 62. Wu Z, Han S, Zhou H, Tuang ZK, Wang Y, Jin Y, Shi H, Yang W. Cold stress
 944 activates disease resistance in *Arabidopsis thaliana* through a salicylic

acid dependent pathway. Plant Cell Environ. 2019; 42 (9): 2645–63.

<https://doi.org/10.1111/pce.13579>.

63. Murcia G, Fontana A, Pontin M, Baraldi R, Bertazza G, Piccoli PN. ABA and

GA3 regulate the synthesis of primary and secondary metabolites related

to alleviation from biotic and abiotic stresses in grapevine. Phytochemistry.

2017; 135: 34–52. <https://doi.org/10.1016/j.phytochem.2016.12.007>.

Fig. 1 The location of each *NtBAHD* on the chromosomes of
Nicotiana tabacum. Scale represents Mb chromosomal distance.

967

968 **Fig. 2** Phylogenic analysis of tobacco and *Arabidopsis* BAHD
969 isoforms. The branches show the percentage of bootstrap values.

970

971 **Fig. 3** Phylogeny, gene structure, and conserved motifs of *BAHD*
972 genes in tobacco. A Neighbor-joining analysis of NtBAHD proteins,
973 lines of varying colors denote distinct groups; B Gene structure of
974 NtBAHDs, the yellow rectangle represents non-coding regions, the
975 green rectangle signifies exons, and the black line denotes introns;
976 C Distribution of conserved motifs into NtBAHD proteins, different
977 colored rectangles represent different motifs.

978

979 **Fig. 4** Collinearity analysis of BAHD family members in *Nicotiana*
980 *tabacum*. A Intra-genomic synteny analysis of *NtBAHD* genes, red
981 lines connect syntenic gene pairs; B Comparative synteny between
982 *N. tabacum* and *A. thaliana*, gray lines indicate all syntenic regions,
983 while red lines highlight syntenic *BAHD* gene pairs.

984

985 **Fig. 5** Analysis of NtHQTs expression in various tobacco tissues. A
986 The color scale represents the relative gene expression (Z-scores);
987 B-E Relative expression levels of *NtHQTs* in lateral roots, fibrous
988 roots, stems, old leaves, middle leaves, new leaves, axillary buds,

flower buds, calyces, crowns, pistils, and stamens. Bars with different letters represent statistically differences at $P < 0.05$ based on Duncan's multiple test.

Fig. 6 Subcellular localization of NtHQTs. pCAMBIA1300-GFP was used as the empty vector control. 4',6-Diamidino-2-Phenylindole (DAPI) was used to label the nucleus. Scale bar = 20 μ m.

Fig. 7 Enzymatic characterization of NtHQTs. A-D Optimal pH and temperature for NtHQT activity; E-H Kinetic analysis of NtHQTs using quinic acid and caffeoyl-CoA as substrates.

Fig. 8 Structural modeling and molecular docking of NtHQT proteins. Computational docking analysis of A NtHQT1, B NtHQT2, C NtHQT3, and D NtHQT4 with the ligands caffeoyl-CoA (red) and quinic acid (orange).

Fig. 9 Effects of VIGS-mediated silencing of *NtHQT* genes on tobacco growth (A) and polyphenolic substances contents (B-E). B The contents of total polyphenols, C chlorogenic acid, D rutin, and E scopoletin were measured. Bars labeled with different letters indicate statistically significant differences ($P < 0.05$) as determined by Duncan's multiple range test.

1013 **Fig. 10** Analysis of *cis*-acting elements and expression profiles of
1014 *NtHQT* genes. A *cis*-acting elements identified in the NtHQT
1015 promoters; B Expression patterns of *NtHQT* genes in response to
1016 phytohormone treatments; C Expression patterns of *NtHQT* genes
1017 under various abiotic stress conditions.

Figures

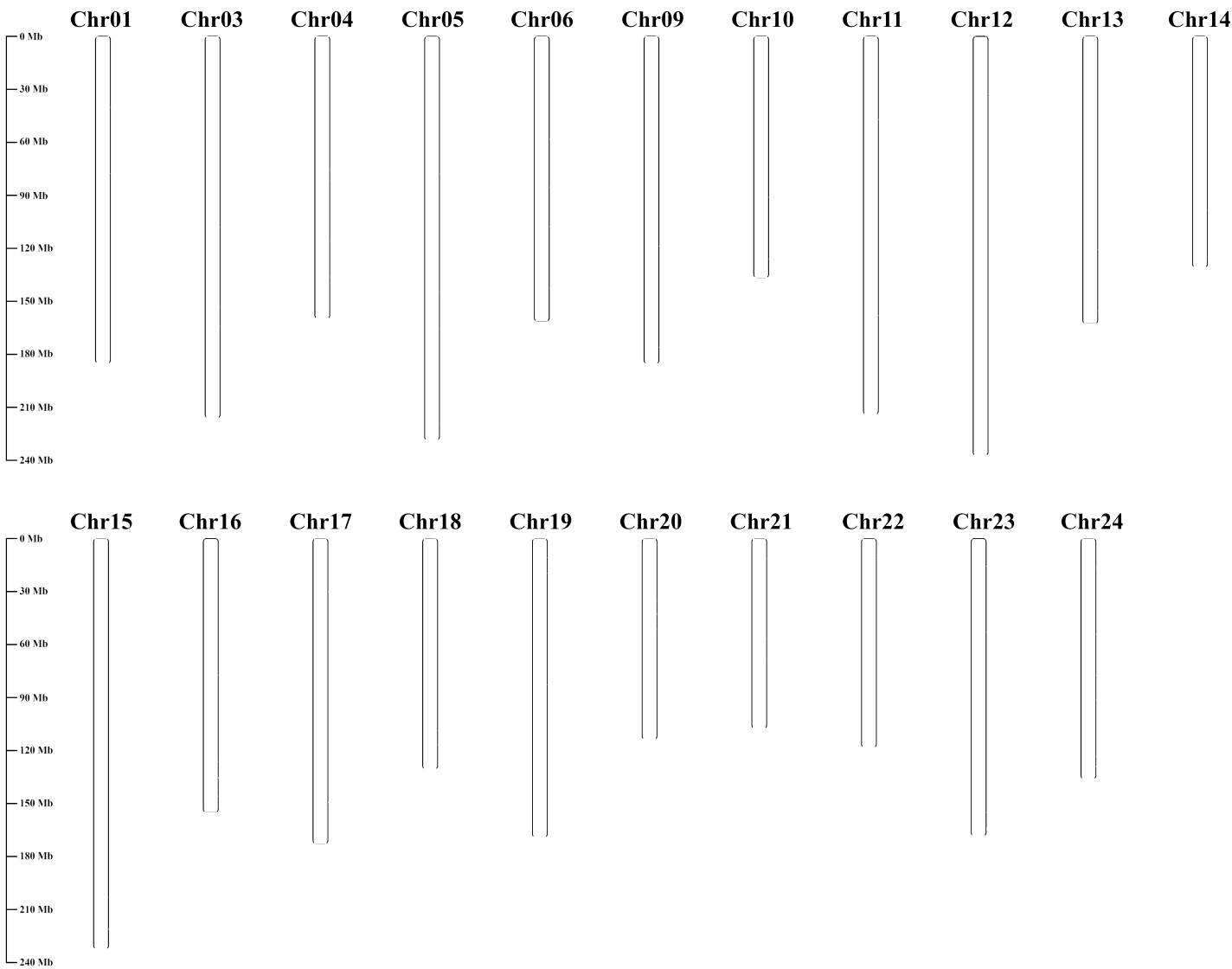


Figure 1

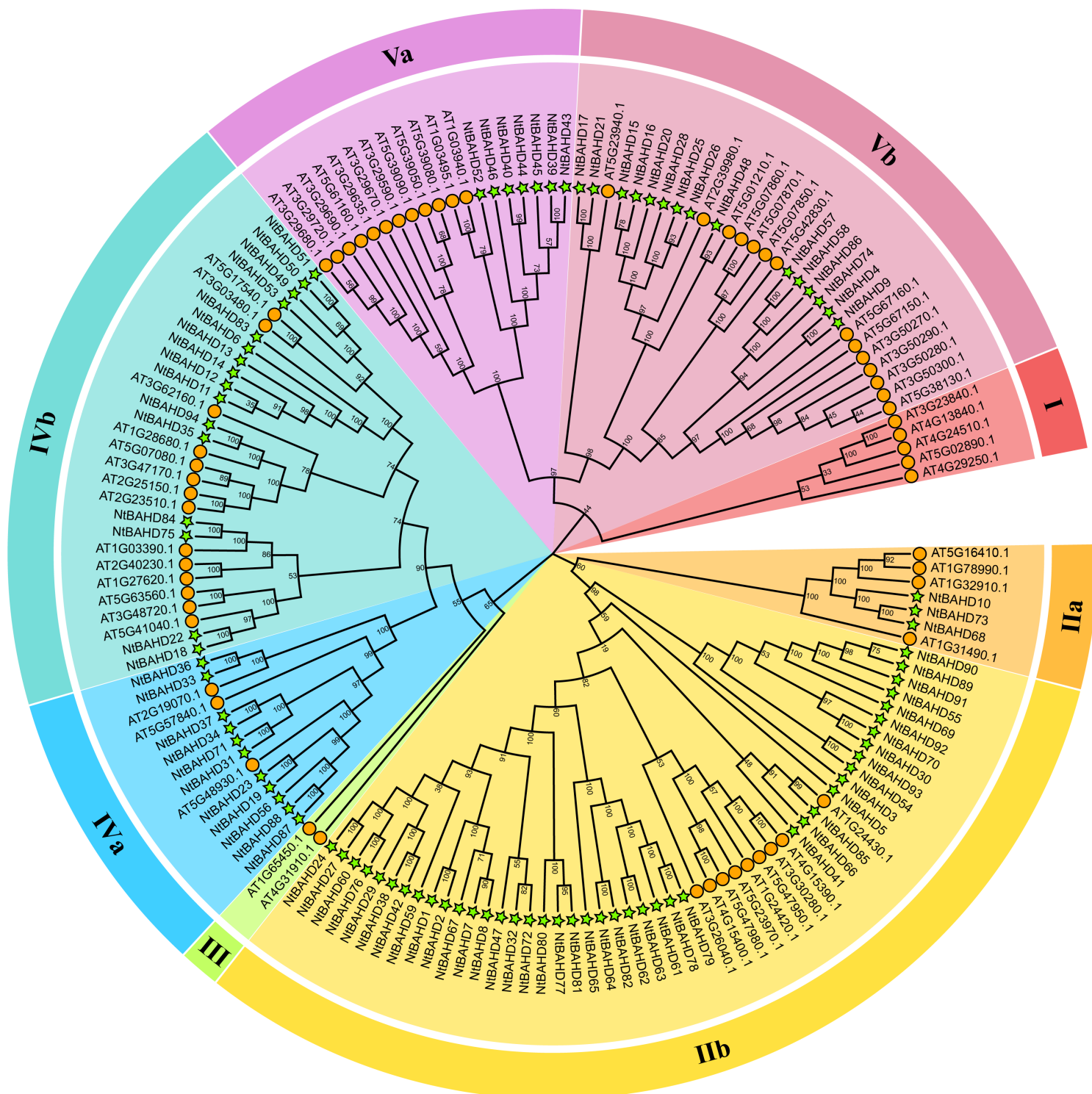


Figure 2



Figure 3



Figure 4

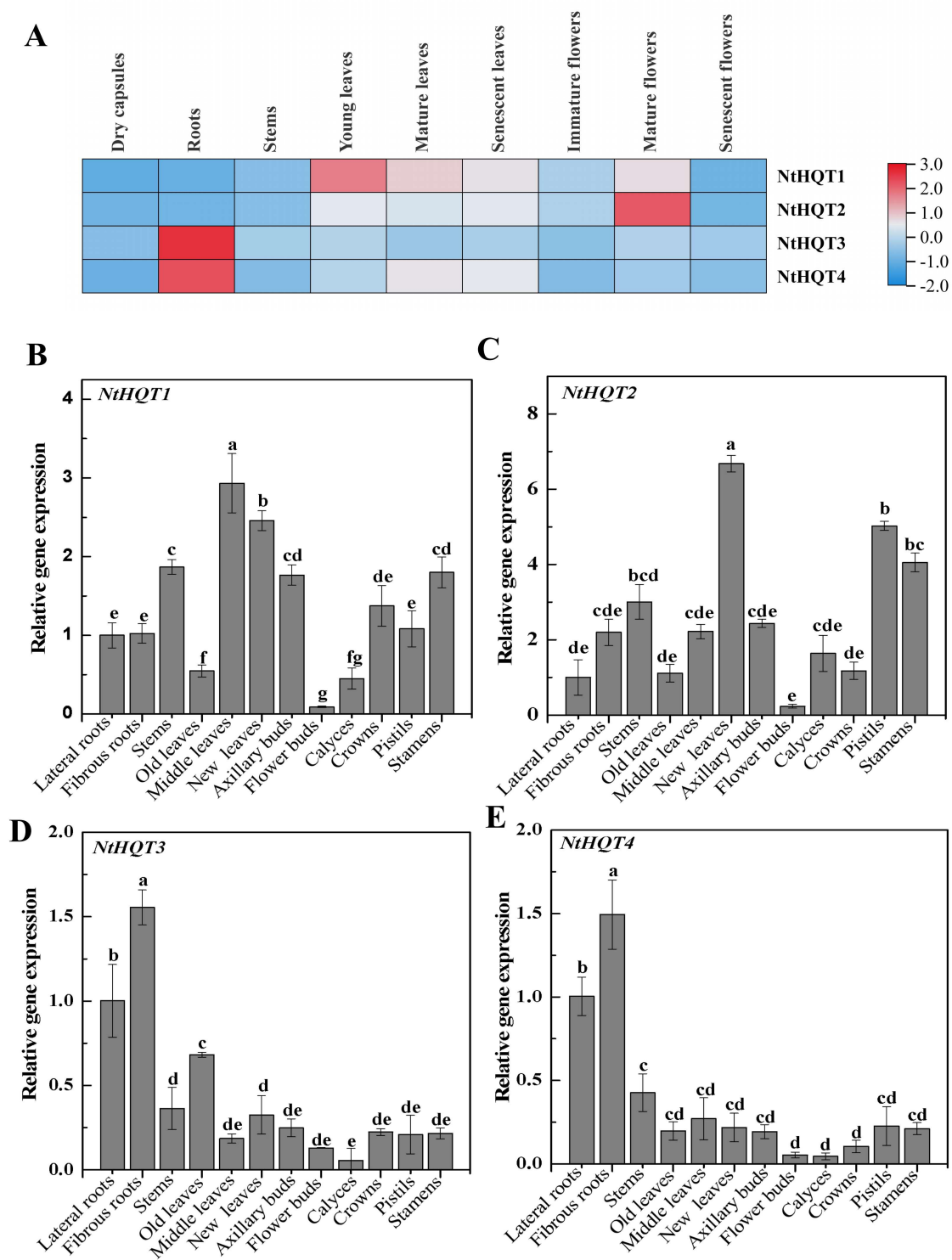


Figure 5

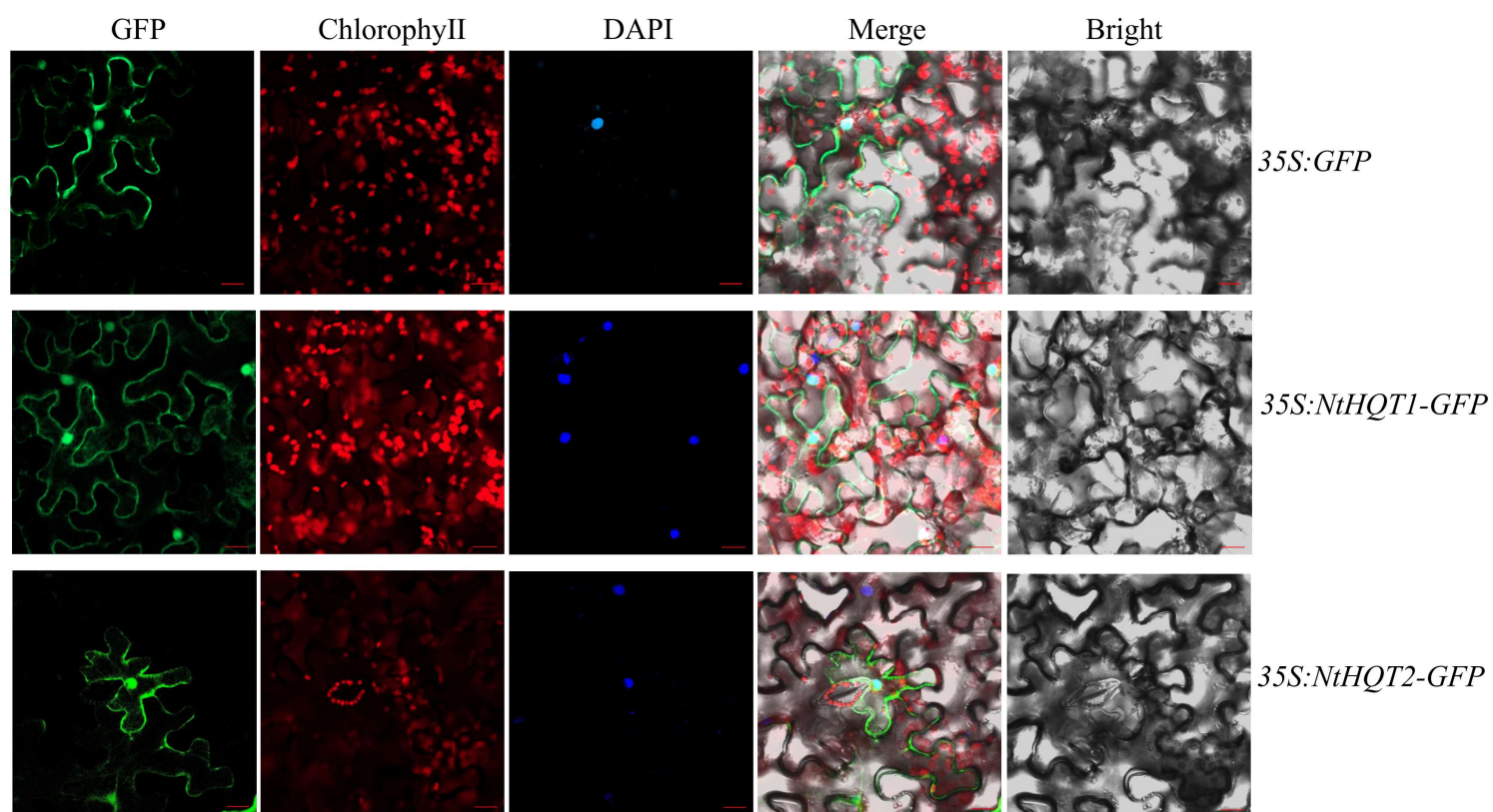


Figure 6

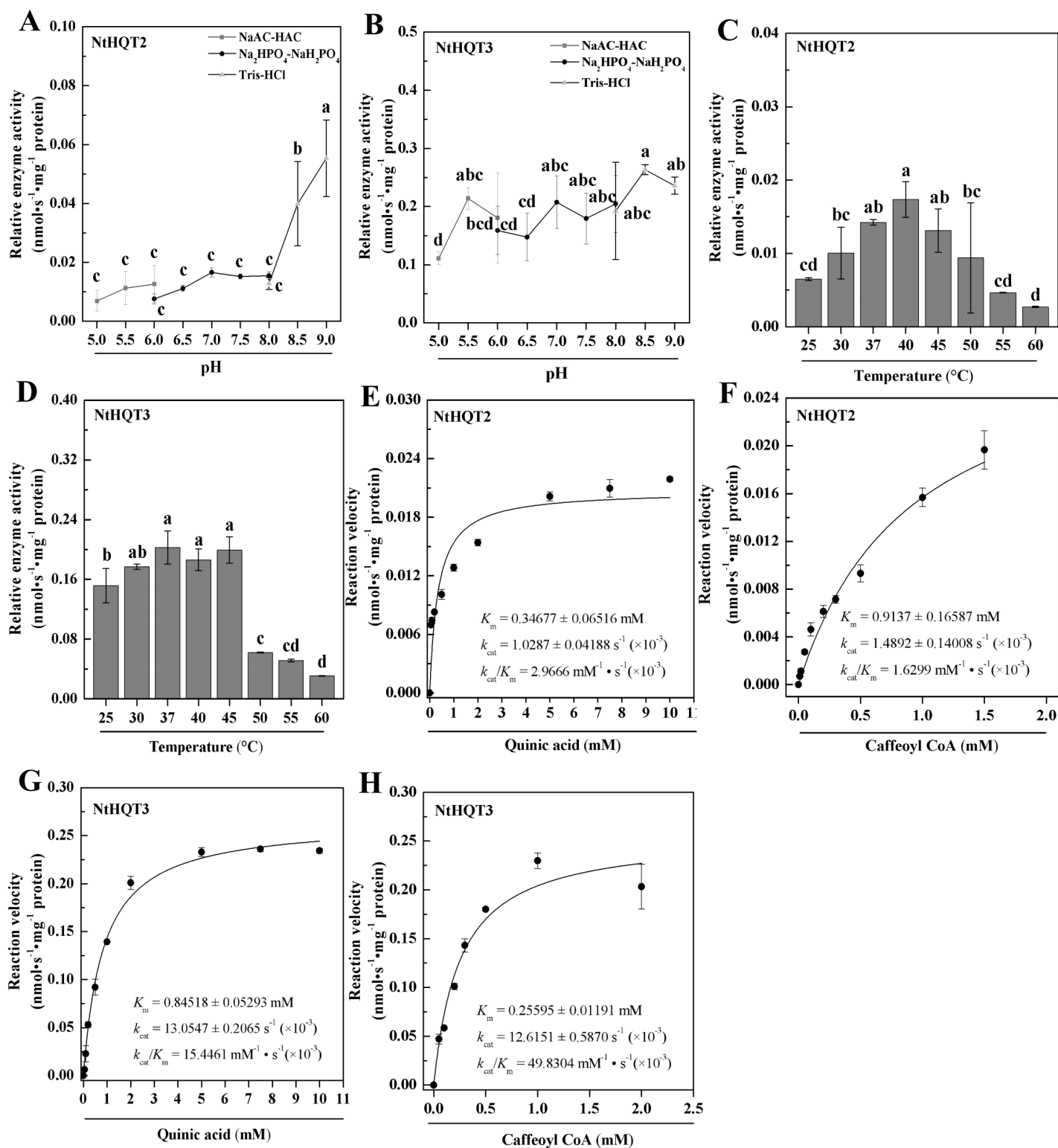


Figure 7

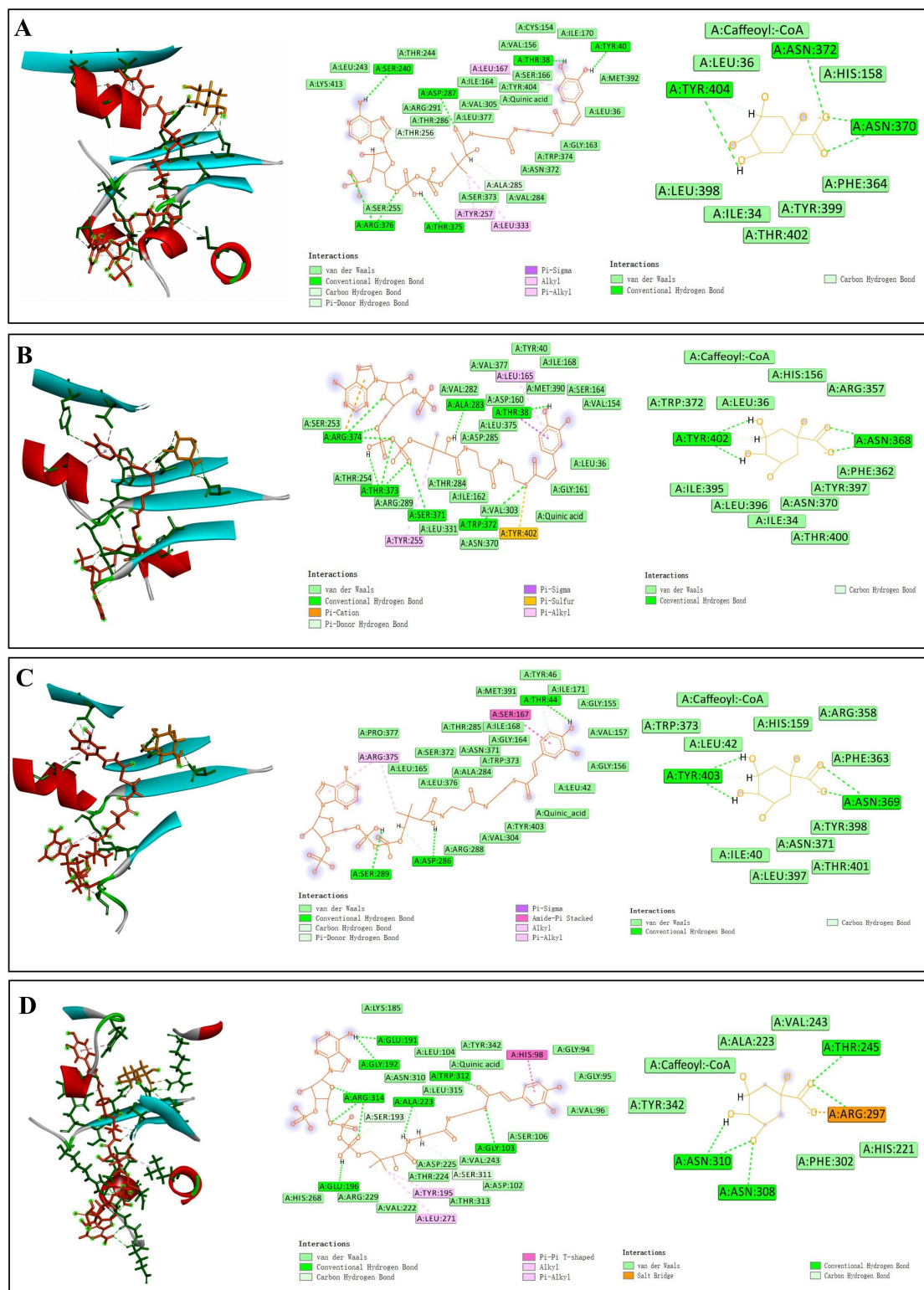


Figure 8

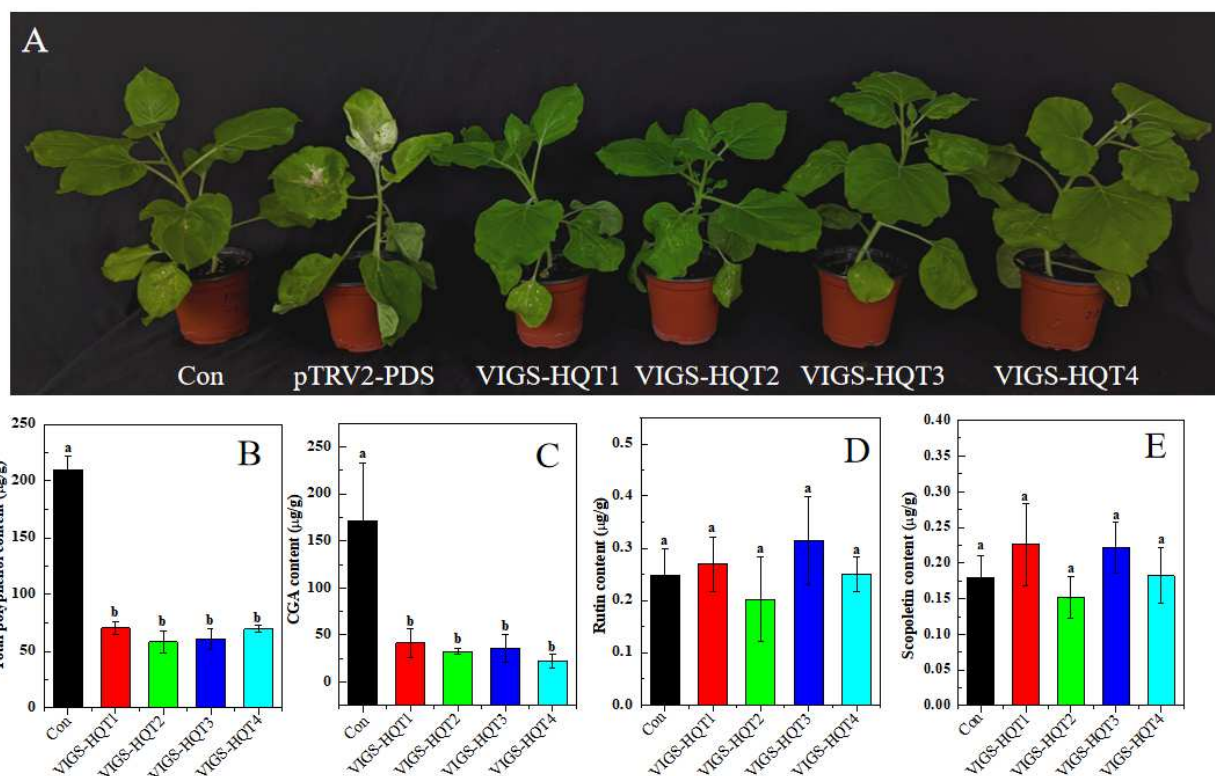


Figure 9

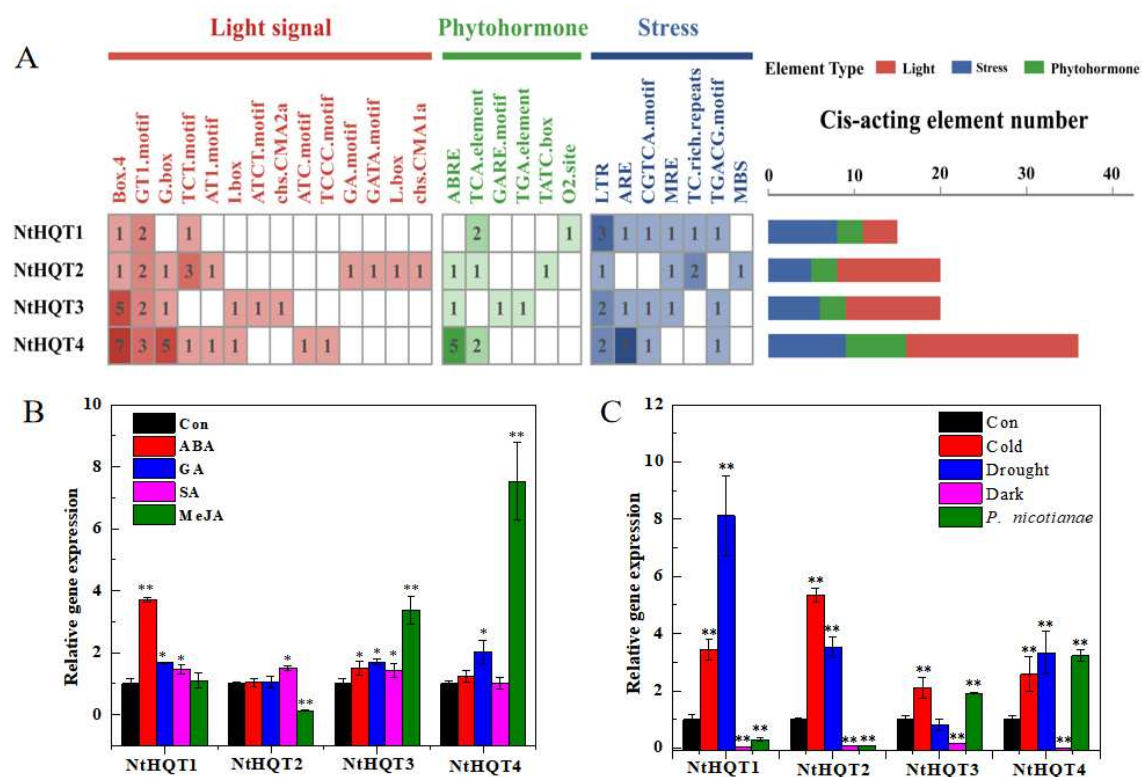


Figure 10

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Table1.docx](#)
- [Table2.docx](#)
- [SupplementaryMaterial.docx](#)
- [supplementarydata1.xlsx](#)