Article

Not peer-reviewed version

Genome-Wide Identification and Analyses of the AHL Gene Family in Pepper (Capsicum annuum L.)

Xiao-Yan Sui , Yan-Long Li , Xi Wang , Yi Zhong , Qing-Zhi Cui , Yin Luo , Bing-Qian Tang , Feng Liu * ,
Xue-Xiao Zou *

Posted Date: 6 February 2025

doi: 10.20944/preprints202502.0476.v1

Keywords: pepper; AT-hook motif nuclear-localized (AHL) gene family; gene expression; genic male sterility

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 February 2025 doi:10.20944/preprints202502.0476.v1

Article

Genome-Wide Identification and Analyses of the *AHL* Gene Family in Pepper (*Capsicum annuum* L.)

Xiao-Yan Sui [1,2], Yan-Long Li [1,2], Xi Wang [1,2], Yi Zhong [1,2], Qing-Zhi Cui [1,2], Yin Luo [2], Bing-Qian Tang [1,2], Feng Liu [1,2,*] and Xue-Xiao Zou [1,2,*]

[1] Engineering Research Center of Education, Ministry for Germplasm Innovation and Breeding New Varieties of Horticultural Crops, Key Laboratory for Vegetable Biology of Hunan Province, College of Horticulture, Hunan Agricultural University, Changsha 410128, China

[2] Yuelushan Laboratory, Changsha 410128, China

* Correspondence: liufengrich@126.com (F.L.); zou_xuexiao@163.com (X.-X.Z.)

Abstract: The *AT-hook motif nuclear-localized* (*AHL*) genes play crucial roles in chromatin remodeling and regulation of gene transcription, significantly influencing biological processes such as plant growth, development, and stress responses. Although *AHL* genes have been extensively studied in various plants, their biological roles in pepper (*Capsicum annuum* L.) remain unknown. In this study, 45 *AHL* genes were identified in the pepper genome through bioinformatics approaches. These genes were analyzed for their chromosomal distribution, phylogenetic relationships, and the structural and functional characteristics of their encoded proteins. Phylogenetic analysis categorized the CaAHL proteins into four subgroups. Transcriptome data revealed that *CaAHL* genes are expressed across various tissues, including roots, stems, flowers, seeds, and fruits, at different growth stages. Subcellular localization experiments confirmed that CaAHL23 predominantly localizes to the cell nucleus, consistent with the localization of its homolog TEK in Arabidopsis thaliana. A yeast two-hybrid assay revealed that CaAHL23 binds to the promoter of *CaCYP703A2*, which is the homolog of the genic male sterility gene *SlCYP703A2* in tomato. These findings indicate a potential link between *CaAHL23* and genic male fertility in pepper. In summary, this study provides valuable insights that establish a foundation for future functional studies of *CaAHL* genes in pepper.

Keywords: pepper; *AT-hook motif nuclear-localized* (*AHL*) gene family; gene expression; genic male sterility

1. Introduction

Capsicum annuum L., commonly known as chili pepper, is a globally significant crop renowned for its distinctive pungency and rich nutritional profile. Their economic importance is underscored by their extensive cultivation and the diverse products derived from them, including fresh produce, spices, and pharmaceuticals. Advancements in genomic research, notably the sequencing of the pepper genome, have paved the way for in-depth studies into gene families that influence key agronomic traits, thereby enhancing breeding programs aimed at improving yield, disease resistance, and stress tolerance [1–3].

The *AT-hook motif nuclear-localized* (*AHL*) gene family is defined by the presence of an AT-hook motif, a small DNA-binding domain that specifically interacts with AT-rich DNA regions, as well as a conserved plant and prokaryote conserved (PPC) domain. Over the past decade, the role of AHLs in regulating various plant growth and developmental processes has been increasingly recognized. These include the elongation of hypocotyls [4–6], the formation of pollen walls and flower development [7,8], root growth [9,10], petiole elongation [11], and the modulation of phytohormone signaling [11]. Furthermore, *AHL* genes have been implicated in plant responses to pathogen infections, as well as to environmental stresses such as salt and drought [12–15]. Recent studies have also broadened our understanding of the *AHL* gene family in important crops like cotton, soybean, maize, and rice [16–19]. These findings emphasize that diverse roles of *AHLs* in plant development, offering new strategies for crop improvement.

In recent years, significant progress has been made in understanding the roles of specific *AHL* genes in plant flower development and male sterility. In *Arabidopsis thaliana*, *AHL22* was identified as a critical regulator of flowering time by modulating *FLOWERING LOCUS T* (*FT*) chromatin state [20]. Similarly, *AHL16* suppressed transposon activation and ensured timely flowering [21]. *TEK* (*TRANSPOSABLE ELEMENT SILENCING VIA AT-HOOK*) was found essential for male fertility. *TEK* regulates *AGPs* in anthers critical for pollen wall formation. Mutations caused microspore development defects and sterility [22]. *AHL15* promoted somatic embryogenesis, highlighting its biotechnological potential [23]. In rice, *PERSISTENT TAPETAL CELL2* (*PTC2*) regulated tapetal programmed cell death and pollen wall patterning, with its loss leading to male sterility [7]. However, despite these advancements, comprehensive investigations into the AHL gene family in chili pepper remain scarce.

Given the economic importance of chili pepper and the potential roles of *AHL* genes in plant development and stress responses, a genome-wide analysis of the *AHL* gene family in chili pepper is warranted. This study aims to identify and characterize the *CaAHL* genes, analyze their expression profiles across different tissues and developmental stages, and explore their potential functions in relation to genic male sterility. The findings from this research will provide valuable insights into the molecular mechanisms governing chili pepper development and facilitate the development of improved cultivars through targeted breeding strategies.

2. Results

2.1. Identification and Basic Information About the CaAHL Gene Family

The Hidden Markov Model (HMM) profile of the PPC domain (PF03479) was used as a query sequence to identify AHL proteins in the Zhangshugang genome database. An initial search yielded 47 *AHL* candidates from the reference genome. After verification, partial *AHL* genes with incomplete PPC domains, considered pseudogenes, were excluded. Ultimately, 45 AHL proteins were confirmed for further analysis. To facilitate subsequent studies, these pepper *AHL* genes were designated as CaAHL1 to CaAHL45 based on their positions across the 12 chromosomes in the Zhangshugang reference genome. Comprehensive details, including gene names, chromosomal locations, protein lengths, molecular weights, theoretical isoelectric points (pI), instability indices, grand average of hydropathicity, and predicted subcellular localizations, are provided in Table 1.

The CaAHLs exhibited significant variation in protein length and physicochemical properties. The protein lengths ranged from 111 amino acids (CaAHL24) to 578 amino acids (CaAHL39). Correspondingly, the molecular weights (MW) varied from 11.71 kDa (CaAHL40) to 60.51 kDa (CaAHL39). The theoretical isoelectric points (pI) spanned from 4.44 (CaAHL34) to 11.25 (CaAHL40), while the instability indices ranged from 30.43 (CaAHL24) to 63.62 (CaAHL31). The grand average of hydropathicity (GRAVY) values ranged from -0.76 (CaAHL4) to 0.18 (CaAHL27). Subcellular localization predictions for the 45 CaAHL proteins revealed diverse distribution patterns: 20 localized to the nucleus, 17 to the chloroplast, five to the cytoplasm, and one each to the plasma membrane, vacuole membrane, and endoplasmic reticulum (ER) (Table 1).

Table 1. Basic information of AHL genes identified in pepper.

Name	Gene ID	Chromosome location	Protein length	Molecular weight (kDa)	Theoretical pI	Instability index	Grand average of hydropathicity	Subcellular localization¹
CaAHL1	Caz01g03310.1	Chr01:5691094-5697294	347	35.27	9.62	38.59	-0.33	nucl
CaAHL2	Caz01g04670.1	Chr01:9114966-9116603	295	31.17	5.79	52.67	-0.54	nucl
CaAHL3	Caz01g04810.1	Chr01:9383031-9386809	322	33.18	8.74	50.79	-0.17	chlo
CaAHL4	Caz01g08370.1	Chr01:20063956-20073612	418	43.66	8.82	60.17	-0.76	nucl
CaAHL5	Caz01g08400.1	Chr01:20111560-20112474	304	31.69	5.75	51.83	-0.27	nucl
CaAHL6	Caz01g13980.1	Chr01:55893363-55894391	314	33.10	6.83	53.54	-0.60	nucl
CaAHL7	Caz01g14080.1	Chr01:56075436-56079692	324	33.72	9.57	47.83	-0.25	chlo
CaAHL8	Caz01g31040.1	Chr01:251924071-251924907	278	29.43	6.54	50.25	-0.47	chlo
CaAHL9	Caz01g35190.1	Chr01:276409484-276418256	123	13.45	5.87	47.09	-0.20	cyto
CaAHL10	Caz01g40510.1	Chr01:326409380-326414186	265	28.25	5.96	44.58	-0.44	nucl
CaAHL11	Caz01g40520.1	Chr01:326434617-326442536	331	35.20	9.51	49.49	-0.46	chlo
CaAHL12	Caz01g40530.1	Chr01:326492721-326618469	316	33.80	10.04	51.86	-0.37	nucl
CaAHL13	Caz01g40540.1	Chr01:326621531-326626277	207	22.30	6.96	44.75	-0.13	nucl
CaAHL14	Caz01g40550.1	Chr01:326687022-326691770	207	22.34	7.84	44.38	-0.14	chlo
CaAHL15	Caz01g40640.1	Chr01:327024609-327031508	329	34.72	9.65	48.09	-0.47	chlo
CaAHL16	Caz01g40680.1	Chr01:327253871-327258557	288	30.70	8.76	47.03	-0.45	chlo
CaAHL17	Caz01g40710.1	Chr01:327378313-327383011	273	29.34	7.79	47.92	-0.53	chlo
CaAHL18	Caz01g40720.1	Chr01:327501110-327505809	273	29.30	7.79	51.24	-0.53	chlo
CaAHL19	Caz01g41290.1	Chr01:329720440-329727085	332	34.13	9.54	44.92	-0.10	chlo
CaAHL20	Caz01g41410.1	Chr01:329900501-329901727	316	33.79	6.05	59.8	-0.63	chlo
CaAHL21	Caz01g41740.1	Chr01:331181444-331186914	352	35.75	9.44	49.49	-0.27	nucl
CaAHL22	Caz02g02600.1	Chr02:44359225-44368231	345	34.69	8.97	55.44	-0.11	E.R.
CaAHL23	Caz02g20690.1	Chr02:156842236-156848762	331	33.33	9.99	50.43	-0.23	nucl
CaAHL24	Caz03g00210.1	Chr03:551605-558146	111	12.02	6.39	30.43	0.01	chlo
CaAHL25	Caz03g21400.1	Chr03:72579177-72587401	346	36.17	6.33	45.59	-0.38	nucl

CaAHL26	Caz03g34730.1	Chr03:267518820-267519650	276	28.85	5.45	57.84	-0.47	cyto
CaAHL27	Caz03g36660.1	Chr03:273808903-273810928	122	12.87	5.19	34.02	0.18	cyto
CaAHL28	Caz03g36670.1	Chr03:273824211-273825851	113	12.01	6.01	47.05	-0.25	nucl
CaAHL29	Caz03g36680.1	Chr03:273829352-273850851	273	28.76	9.64	49.08	-0.30	vacu
CaAHL30	Caz04g00390.1	Chr04:713251-718475	267	26.36	6.42	48.45	-0.13	nucl
CaAHL31	Caz04g08200.1	Chr04:20375417-20376564	352	37.79	7.05	63.62	-0.66	nucl
CaAHL32	Caz05g17580.1	Chr05:235719249-235723745	294	31.56	5.35	52.68	-0.61	nucl
CaAHL33	Caz06g17080.1	Chr06:50064634-50076684	349	36.50	9.34	53.47	-0.51	nucl
CaAHL34	Caz06g17990.1	Chr06:57415662-57416120	152	15.59	4.44	51.44	-0.03	nucl
CaAHL35	Caz06g24080.1	Chr06:182665460-182672828	341	35.70	7.02	40.07	-0.30	plas
CaAHL36	Caz07g19270.1	Chr07:251735483-251736292	269	28.24	9.24	33.67	-0.13	cyto
CaAHL37	Caz08g07460.1	Chr08:137932240-137938844	341	34.99	10.26	58.12	-0.22	nucl
CaAHL38	Caz09g21720.1	Chr09:274809109-274813725	438	45.01	9.38	50.94	-0.34	nucl
CaAHL39	Caz12g05920.1	Chr12:13682382-13697371	578	60.51	7.72	52.36	-0.30	chlo
CaAHL40	Caz12g05950.1	Chr12:13811174-13813071	115	11.71	11.25	59.27	-0.01	chlo
CaAHL41	Caz12g06070.1	Chr12:14037114-14041618	177	18.10	4.89	59.53	-0.20	chlo
CaAHL42	Caz12g06080.1	Chr12:14056732-14062760	438	47.47	9.27	50.27	-0.21	chlo
CaAHL43	Caz12g08880.1	Chr12:32039622-32040395	257	28.05	7.83	49.44	-0.42	nucl
CaAHL44	Caz12g18220.1	Chr12:219722263-219725994	293	29.60	6.16	46.17	-0.33	nucl
CaAHL45	Caz12g18510.1	Chr12:221530027-221540499	358	37.13	9.57	50.03	-0.37	cyto

[1] Note: nucl, nucleus; chlo, chloroplast; cyto, cytoplasm; plas, plasma membrane; vacu, vacuole membrane; E.R., endoplasmic reticulum.

2.2. Chromosome Distribution of the CaAHL Gene Family

The 45 *CaAHLs* were unevenly distributed across 9 of the 12 chromosomes in the Zhangshugang genome (Figure 1a). Most *CaAHLs* were concentrated near the distal ends of the chromosomes, with only a few located in central regions. Chromosome 1 had the highest number of *CaAHLs*, with a total of 21, followed by chromosomes 12 and 3, which contained 7 and 6 genes, respectively. In contrast, no *CaAHLs* were identified on chromosomes 10 and 11 (Figure 1b).

Interestingly, the *CaAHLs* were enriched in four specific regions of the genome: the proximal and distal ends of chromosome 1, the distal end of chromosome 3, and the proximal end of chromosome 12. These regions were designated as cluster 1, cluster 2, cluster 3, and cluster 4, respectively. Cluster 1 comprises five genes (*CaAHL1* to *CaAHL5*), cluster 2 contains 12 genes (*CaAHL10* to *CaAHL21*), cluster 3 includes three genes (*CaAHL27* to *CaAHL29*), and cluster 4 consists of four genes (*CaAHL39* to *CaAHL42*) (Figure. 1a). The uneven distribution and clustering of *CaAHLs* suggest a possible accumulation of genes through tandem duplication events in specific genomic regions.

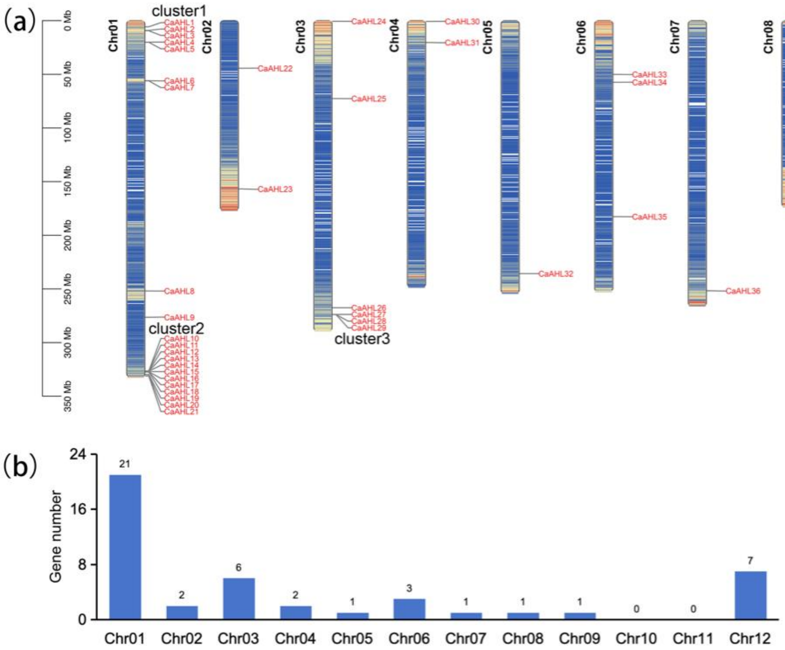

Figure 1. Chromosome mapping and gene number of *CaAHLs* in Zhangshugang genome. (a) Chromosome mapping of *CaAHLs*. Chromosome numbers are represented on the top, and the scale is shown on the left. (b) Statistical analysis for *CaAHLs* on 12 chromosomes.

2.3. Phylogenetic Analysis of the CaAHL Gene Family

We subsequently performed a phylogenetic analysis of the 45 *CaAHLs*, using their protein sequences to construct an unrooted tree for classification of the *CaAHL* gene family (Figure 2). The phylogenetic analysis revealed that the *CaAHL* gene family is divided into four main branches: branch A, containing 15 genes; branch B, containing 9 genes; branch C, containing 6 genes; and branch D, containing 15 genes. This classification provides insight into the evolutionary relationships within the *CaAHL* gene family. Combined with chromosome localization analysis, gene clusters were mostly distributed in the branch D, suggesting that the differentiation of gene function may gradually occur with gene duplication.

Next, we analyzed the tandem duplications within the *CaAHL* gene family and highlighted the genes in the four *CaAHL* gene clusters that may exhibit tandem repeats by using different colors in the phylogenetic tree (Figure 2). Overall, we observed that the five genes in cluster 1 grouped together with the five genes in cluster 2 in the phylogenetic tree, while the remaining three genes in cluster 1 clustered with the three genes in cluster 3 (Figure 2). We speculate that this pattern may result from the duplication of entire gene clusters on chromosomes during evolutionary processes. From the perspective of gene clusters, we found that five genes in cluster 1 (*CaAHL13*, *CaAHL14*, *CaAHL16*, *CaAHL17*, and *CaAHL18*) and four genes in cluster 4 (*CaAHL39*, *CaAHL40*, *CaAHL41*, and *CaAHL42*) were clustered together in the phylogenetic tree, respectively, suggesting the occurrence of tandem duplications within these clusters. In summary, the *CaAHLs* may undergo both inter-chromosomal replication and intra-cluster duplication.

doi:10.20944/preprints202502.0476.v1

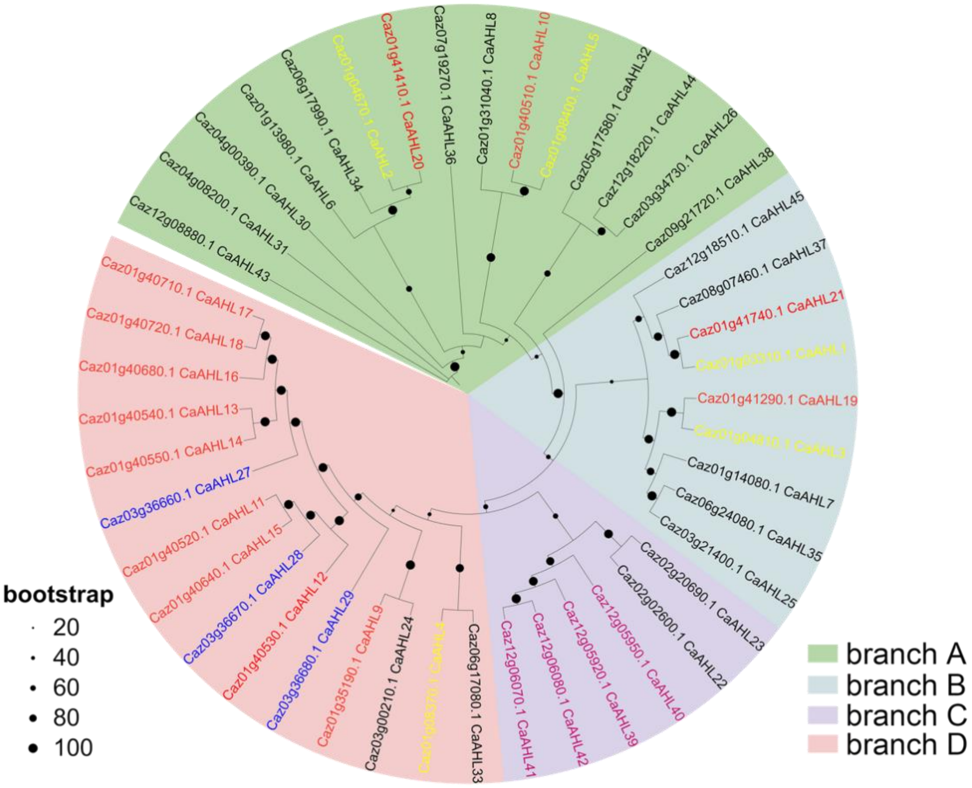

Figure 2. Phylogenetic relationships among *CaAHLs* in the Zhangshugang genome. Branches are shaded in different colors to distinguish groups. Black dots indicate bootstrap support values, with only those greater than 0.5 displayed. Predicted tandem duplicate gene pairs are highlighted using various font colors: yellow for genes in cluster 1, red for cluster 2, blue for cluster 3, and purple for cluster 4.

2.4. Analysis of CaAHLs Conserved Motifs

We conducted an analysis of the phylogenetic relationships and conservative motifs of CaAHL proteins in pepper. The results indicated that most CaAHLs contained motifs 1, 2, 3, and 4. Nearly all CaAHLs included motif 1 and motif 3, except for CaAHL41, which lacked motif 1, and CaAHL17 and CaAHL18, which lacked motif 3. Additionally, 82.22% and 73.33% of CaAHLs contained motif 4 and motif 2, respectively (Figure 3 and Table S1). Some CaAHLs contained multiple identical motifs. For example, CaAHL11, CaAHL12, and CaAHL15 each contained two motif 4 elements, while CaAHL31 and CaAHL4 contained two and three motif 10 elements, respectively. Additionally, we found that CaAHL11 and CaAHL15, CaAHL13 and CaAHL14, as well as CaAHL23 and CaAHL24 exhibited highly similar motif structures, which is consistent with their clustering in the phylogenetic analysis (Figure 3 and Table S1). Overall, CaAHLs in the same subgroup in phylogenetic tree had similar structures and conserved motif distributions (Figure 3 and Table S1), indicating that CaAHLs contain highly conserved amino acid residues and that CaAHLs in the same cluster may have similar roles.

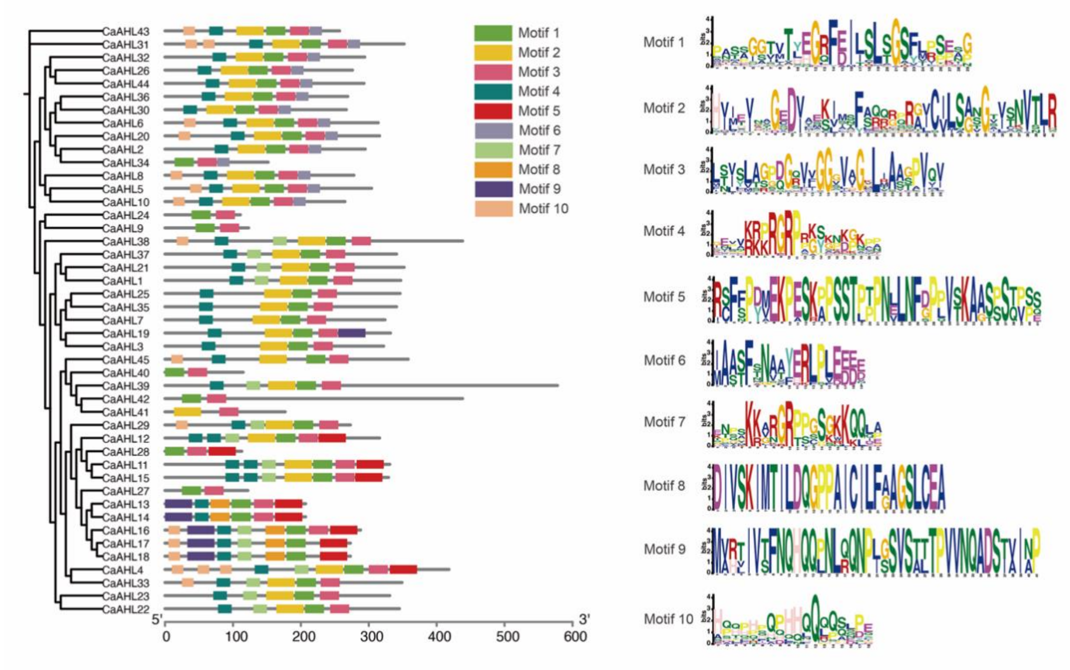

Figure 3. Conserved motif analysis of CaAHLs. The evolutionary tree on the left is constructed based on the protein sequences of CaAHLs.

2.5. Cis-Regulatory Element Analysis of the CaAHLs Promoter

Since *CaAHLs* play important roles in response to plant growth, development, and stress responses, we utilized PlantCARE to analyze the cis-regulatory elements in the pre-2000 bp region of the *CaAHLs* promoter to explore their possible functions. For the promoter of the *CaAHLs*, we screened cis-regulatory elements related to growth, development, stress, and hormone responses. Using TBtools, these elements were categorized into 18 distinct classes. (Figure 4 and 5). The analysis revealed that the promoters of *CaAHL22* and *CaAHL24* are enriched with abscisic acid-responsive cis-regulatory elements, containing 5 and 4 elements, respectively. Promoters of *CaAHL7*, *CaAHL1*, and *CaAHL34* exhibit a higher abundance of anaerobic-responsive elements, with 5, 4, and 4 elements, respectively. The promoter of *CaAHL28* shows a significant presence of gibberellin-responsive elements. Similarly, the promoters of *CaAHL23* and *CaAHL4* are enriched with jasmonic acid-responsive elements, containing 7 and 5 elements, respectively. *CaAHL28* and *CaAHL29* exhibit higher numbers of MYB binding site-related elements, with 6 and 5 elements, respectively, while the promoter of *CaAHL21* contains 4 salicylic acid-responsive elements (Figure 4 and 5). These distinct distributions of cis-regulatory elements suggest potential functional differentiation among the *CaAHLs*.

doi:10.20944/preprints202502.0476.v1

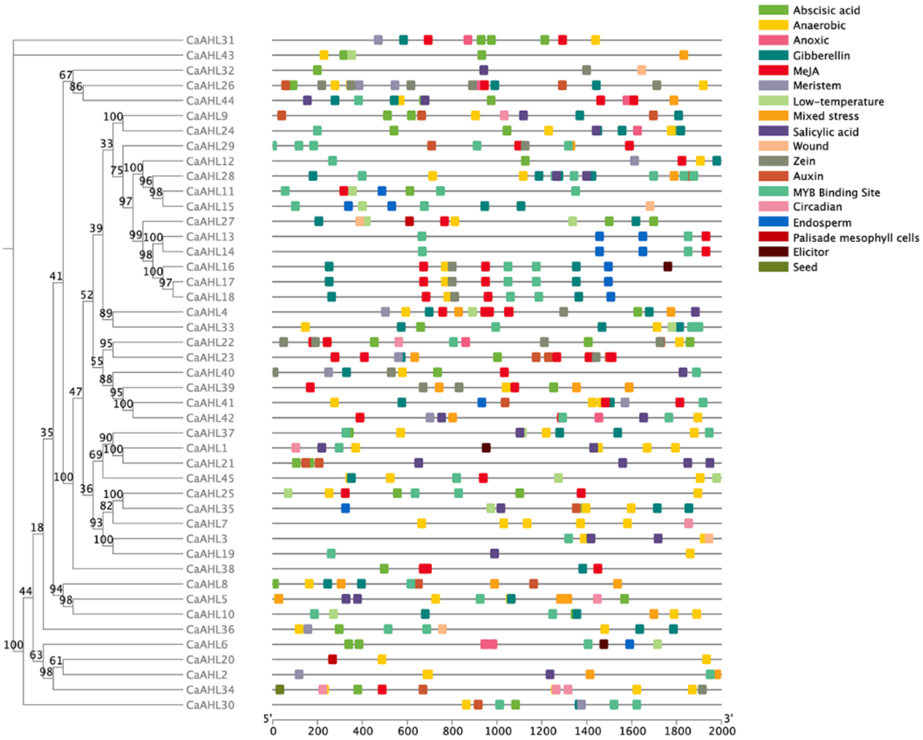

Figure 4. The distribution of cis-regulatory elements predicted in the *CaAHLs* promoter. Different colored boxes represent different cis-regulatory elements. The evolutionary tree on the left is constructed based on the protein sequences of *CaAHLs*.

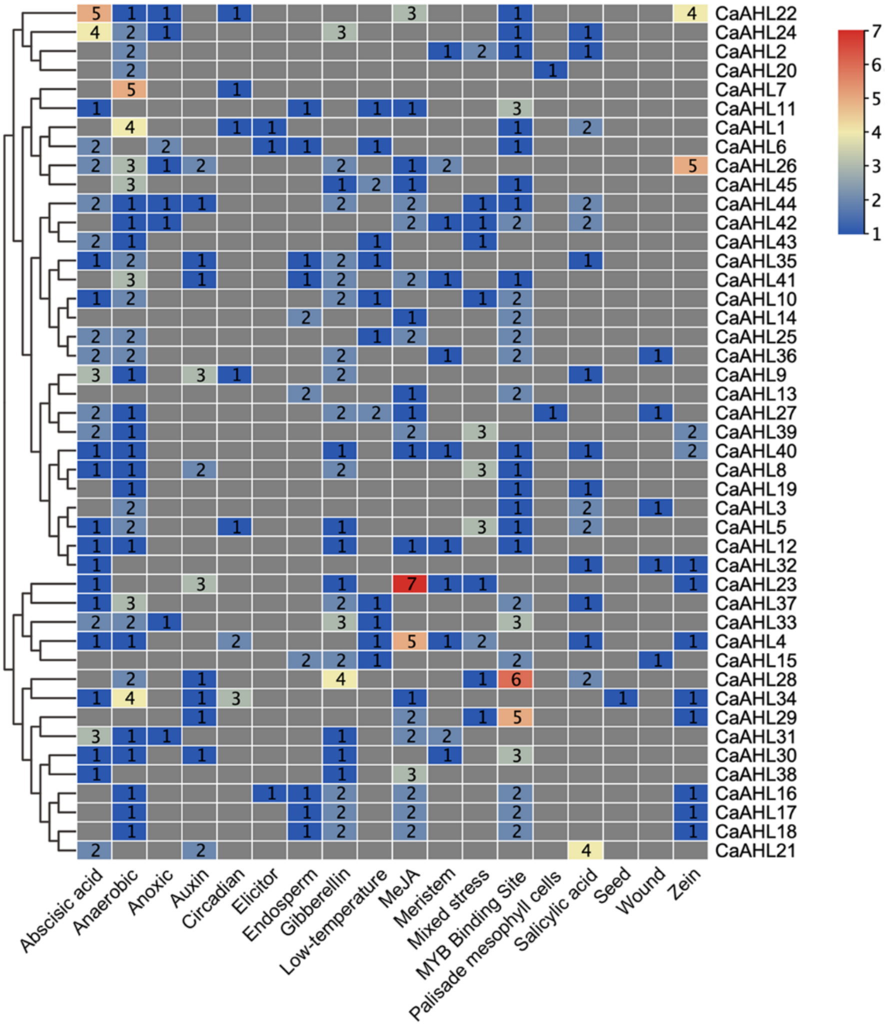

Figure 5. Number statistics of cis-regulatory elements in the *CaAHLs* promoter regions. The evolutionary tree on the left is based on the analysis of the number and types of cis-regulatory elements.

2.6. Tissue-Specific Expression Profiles of the CaAHLs in Peppers

To explore whether the *CaAHLs* plays a role in the tissue development of pepper, the expression profiles of *CaAHLs* in the roots, stems, leaves, pericarp, and placenta were observed using RNA-seq data from the pepper line 6421 (Figure 6). The results revealed that several genes, including *CaAHL1*, *CaAHL26*, *CaAHL5*, *CaAHL8*, *CaAHL10*, *CaAHL44*, *CaAHL2*, and *CaAHL20*, exhibited root- and stem-specific expression (Figure 6). Certain genes, such as *CaAHL9*, *CaAHL20*, *CaAHL27*, *CaAHL28*, and *CaAHL11* to *CaAHL18*, showed higher expression levels in seeds at 20 and 25 days after flowering (DAF). Some genes were specifically expressed in certain tissues or developmental stages. For instance, *CaAHL29* displayed high expression in seeds at 60 DAF, suggesting a potential role in seed maturation. *CaAHL36*, *CaAHL23*, *CaAHL12*, and *CaAHL16* to *CaAHL18* exhibited elevated expression during various stages of bud development, indicating their involvement in floral development. *CaAHL21* showed specific expression in the placenta at 35 DAF, implying a role in placenta development (Figure 6). These findings suggest that most *CaAHL* genes are specifically expressed in roots, stems, and seeds at 20 and 25 DAF, highlighting their significant roles in root, stem, and seed development. The tissue-specific expression patterns of certain genes indicate functional differentiation among the *CaAHL* family members.

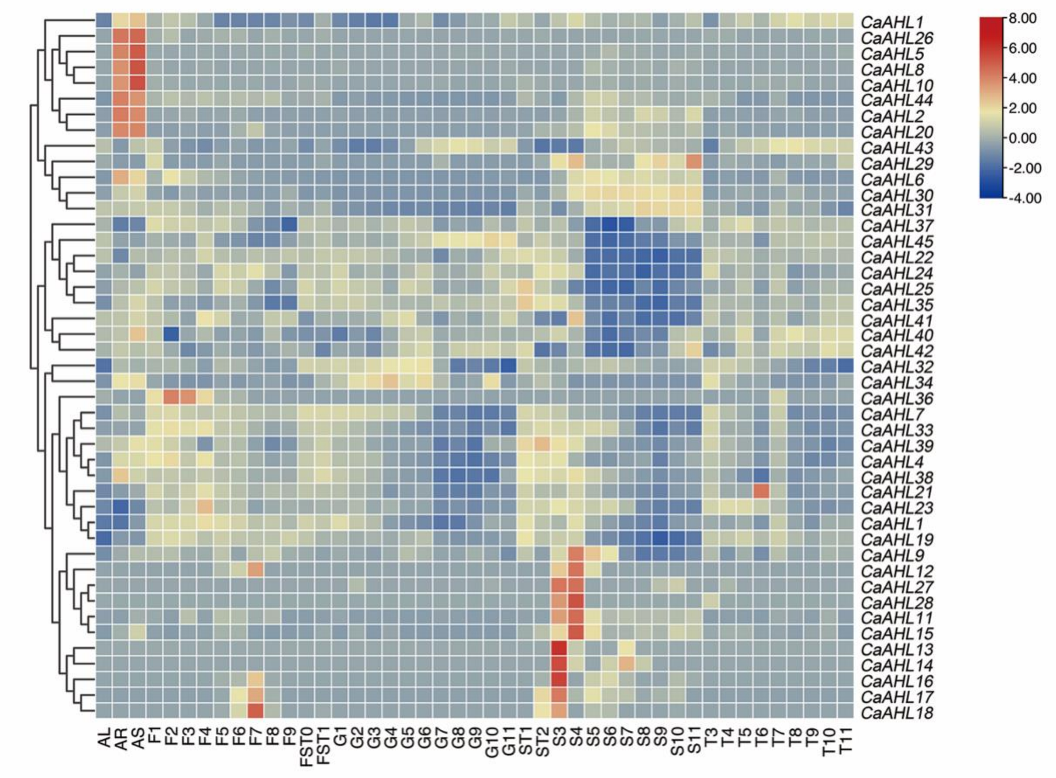

Figure 6. Expression profile analysis of *CaAHLs* in various tissues and organs of pepper. Expression levels were determined in the following tissues and stages: leaf tissues were sampled at 60 days after emergence and marked correspondingly as AL; Stems and roots were marked AS and AR, respectively. Floral buds were sampled at 0.25, 0.35, 0.5, 0.8, 1.0, 1.2, and 1.7 cm, and marked correspondingly as F1, F2, F3, F4, F5, F6, F7, F8, and F9; fruits were collected at 3, 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, and 60 days after flowering (DAF), and marked correspondingly as FST0, FST1, G1, G2, G3, G4, G5, G6, G7, G8, G9, G10, G11; seed samples were collected at 10, 15, 20, 25, 30, 35, 40, 45, 50, 55 and 60 DAF, and marked correspondingly as ST1, ST2, S3, S4, S5, S6, S7, S8, S9, S10, S11; placenta samples were collected at 20, 25, 30, 35, 40, 45, 50, 55 and 60 DAF, and marked correspondingly as T3, T4, T5, T6, T7, T8, T9, T10, and T11. The FPKM values were log2-transformed, and a heatmap was generated using TBtools-II (v2.149) software. The evolutionary tree on the left is constructed based on gene expression levels. Expression values on the right are shown as a color gradient from low expression (blue) to high expression (red).

2.7. *CaAHL23 as a Potential Regulator in Pepper Male Sterility*

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn. To investigate the relationship between *CaAHL* genes and male sterility, we focused on genes specifically expressed during flower development. Through transcriptome data and real-time quantitative data (Figure S1), we found that *CaAHL23* is specifically expressed in buds and exhibits a gene expression pattern highly similar to that of the well-characterized *Arabidopsis TEK* gene [22], suggesting a potential role in genic male sterility. To explore this hypothesis, we first conducted a subcellular localization analysis of CaAHL23, which confirmed its nuclear localization (Figure 7a), consistent with the previously reported localization of TEK. Subsequently, yeast one-hybrid assays were performed to identify downstream genes potentially regulated by *CaAHL23*. These assays demonstrated that CaAHL23 binds directly to the promoter region of *CaCYP703A2* (Figure 7b). Given the high sequence homology between *CaCYP703A2* and the tomato sterility-related gene *SlCYP703A2* (Figure S2), we hypothesize that CaAHL23 may regulate pepper fertility by modulating the expression of *CaCYP703A2*. Subsequently, we detected the expression of *CaCYP703A2* during flower

doi:10.20944/preprints202502.0476.v1

development in the sterile line material WX2A and its fertile line WX2B, and expression profile data in the wild type (Figure 8). The results showed that compared with fertile lines, the expression of *CaCYP703A2* was significantly reduced in sterile lines during flower development. These findings provide new insights into the genetic mechanisms underlying male sterility in pepper and highlights *CaAHL23* as a candidate gene for further functional studies.

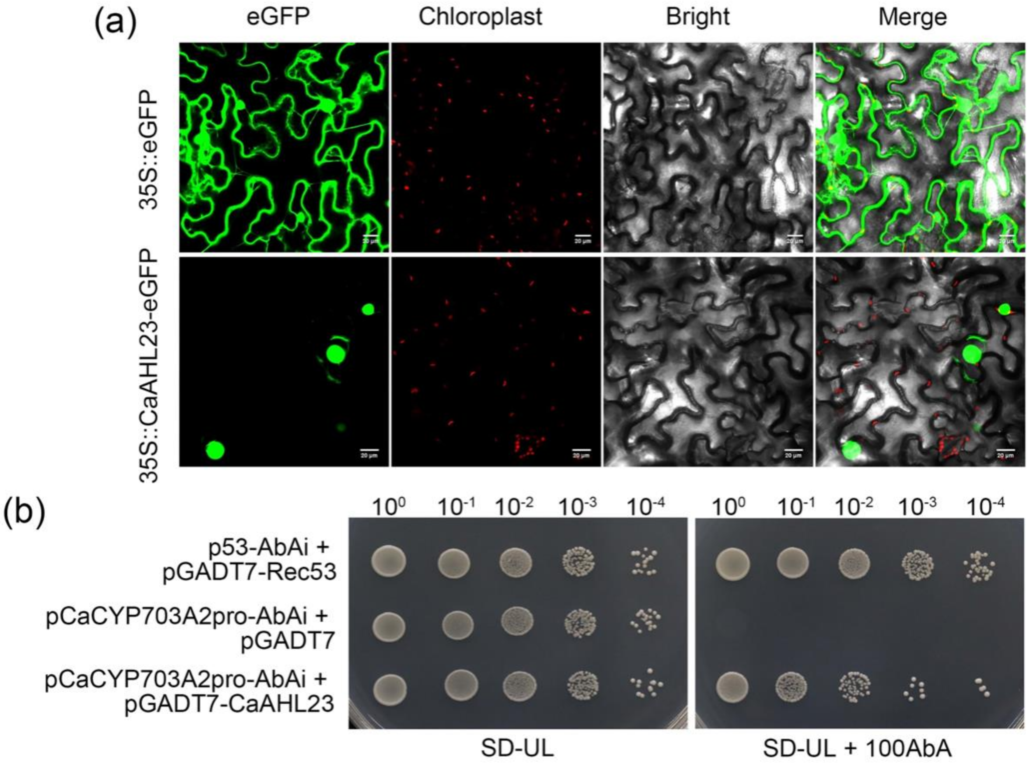

Figure 7. Subcellular localization (a) and yeast one-hybrid assay (b) of CaAHL23.

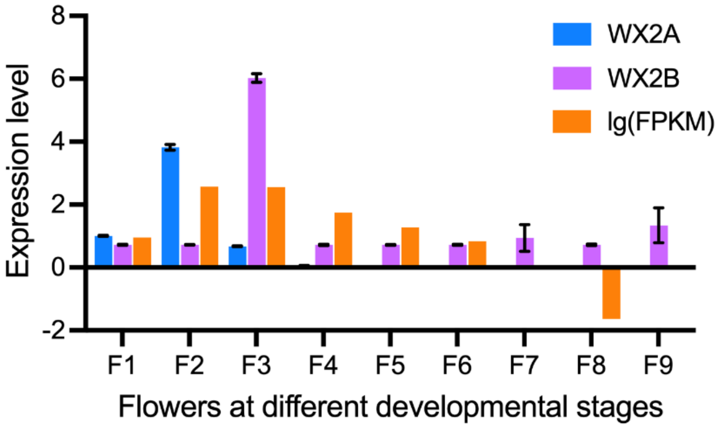

Figure 8. Expression level analysis of *CaCYP703A2* during flower development. Floral buds were sampled at 0.25, 0.35, 0.5, 0.8, 1.0, 1.2, and 1.7 cm, and marked correspondingly as F1, F2, F3, F4, F5, F6, F7, F8, and F9.

3. Discussion

The comprehensive analysis of the *CaAHL* gene family presented in this study offers valuable insights into the structure, evolutionary relationships, and functional roles of these genes in pepper. The identification and characterization of 45 *CaAHL* genes provide a foundational understanding of

doi:10.20944/preprints202502.0476.v1

their molecular properties and potential regulatory functions in plant development and fertility. The identification of 45 *CaAHL* genes and detailed analysis of their protein characteristics, including length, molecular weight, theoretical isoelectric point, instability index, and predicted subcellular localization, reflect the molecular diversity of the *CaAHL* gene family (Table 1). These findings are consistent with previous studies that demonstrated functional diversification within gene families related to plant development [24–26]. Chromosomal distribution analysis revealed four gene clusters (Figure 1), suggesting that tandem duplications have contributed to the expansion of the *CaAHL* gene family. This observation aligns with studies in other species, where tandem duplications play a critical role in the evolution and functional diversification of gene families [27–29].

Phylogenetic analysis grouped the 45 *CaAHL* genes into four distinct subgroups and highlighted both inter-chromosomal replication and intra-cluster duplication as mechanisms driving gene family expansion (Figure 2). The conservation of motif distributions among members of the same subgroup indicates functional similarities and evolutionary conservation (Figure 3), consistent with observations in other plant species [30,31]. The presence of highly conserved amino acid residues underscores the evolutionary importance of these motifs in maintaining gene function.

Cis-regulatory element analysis revealed elements related to stress response, hormone signaling, and developmental processes, suggesting that *CaAHL* genes may be involved in diverse regulatory networks (Figure 4 and 5). The tissue-specific expression analysis showed that many *CaAHL* genes are highly expressed in roots, stems, and seeds at 20 and 25 days after flowering (DAF) (Figure 6), highlighting their potential roles in organ development. Notably, the functional differentiation indicated by distinct expression patterns among family members suggests specialization in different developmental or stress response pathways, as similarly reported in studies of *AHL* family genes in *Arabidopsis* [6,8,22,32,33].

A critical finding of this study is the potential regulatory role of *CaAHL23* in pepper male sterility. Subcellular localization experiments confirmed its nuclear localization (Figure 7a), consistent with the *Arabidopsis TEK* gene, known for its role in pollen wall development and male sterility [22]. Yeast one-hybrid assays demonstrated that CaAHL23 binds directly to the promoter region of *CaCYP703A2* (Figure 7b), a gene whose homologs in *Arabidopsis*, cotton, and rice are essential for pollen wall formation. Knockout studies of these homologs have resulted in male sterility [34,35]. At the same time, the expression of *CaCYP703A2* gene in the sterile line was significantly reduced during flower development compared with fertile lines (Figure 8). Therefore, we hypothesize that *CaAHL23* regulates pepper fertility by modulating *CaCYP703A2* expression, making it a promising target for further studies on fertility regulation and hybrid breeding.

Our findings provide a comprehensive view of the *CaAHL* gene family's structure, evolution, and functional roles in pepper. The evidence for the involvement of *CaAHL23* in male sterility opens new avenues for exploring its regulatory mechanisms and potential applications in hybrid breeding. Future research should focus on functional validation of *CaAHL23* and its regulatory targets through genome-editing technologies and field trials to assess its role in fertility and crop improvement.

4. Materials and Methods

4.1. Retrieval and Identification of AHL Genes in Pepper

In this study, the candidate AHL proteins were retrieved as follows: first, the protein, nucleotide, and genome sequences of Zhangshugang genomes were downloaded from the Pepper Genomics Database (http://ted.bti.cornell.edu/cgi-bin/pepper/search) [36]. Second, the HMM of AHL protein (PF03479) was downloaded from the Pfam database (http://pfam-legacy.xfam.org/) [37]. Finally, the TBtools-II (v2.149) software [38] was used to search the predicted AHL proteins using the cut-off value of the default parameter.

4.2. *Sequence Analysis and Structural Characteristics*

Protein lengths, molecular weights, theoretical isoelectric points (pI), instability indices, and grand average of hydropathicity of CaAHL proteins were analyzed using the TBtools-II (v2.149) software [38]. The supposed subcellular localizations of CaAHL proteins were predicted using the online tool WoLF PSORT (https://wolfpsort.hgc.jp) [39]. The protein sequences were submitted to the MEME program (https://meme-suite.org/meme/tools/meme) [40] to assess conserved motifs.

4.3. *Chromosome Localization, Tandem Duplication, and Synteny Analysis*

Chromosome locations and gene position in pepper were obtained by searching the Sol Genomics Network. Chromosome mapping of the *CaAHL* gene family was visualized using MG2C (http://mg2c.iask.in/mg2c_v2.1) [41]. Tandem duplication events were further confirmed using the following criteria: (1) the alignment length had a coverage rate of more than 70% of the full length of the *CaAHL* genes; (2) the identity of the aligned region was over 70%, (3) and an array of two or more genes was less than 100 kb distance. TBtools-II (v2.149) software [38] was used to analyze the synteny of *CaAHL* genes among the three pepper genomes.

4.4. *Phylogenetic Analysis*

The phylogenetic tree was generated in the following three steps: first, the CaAHL protein sequences were imported into Clustal X to produce a multiple sequence alignment file. Second, the alignment result was used to build an unrooted tree using MEGA11 with a bootstrap of 1000 replicates and neighbor-joining (NJ) methods [42]. Third, the newly produced phylogenetic tree was visualized using the Interactive Tree of Life online website (https://itol.embl.de/) [43].

4.5. *RNA-Seq Analysis of CaAHL Genes*

Transcriptome sequencing (RNA-seq) data of development was used to explore the distribution of gene expression in pepper (the elite Capsicum line 6421) [44] to gain insight into the expression profiles of the *CaAHL* gene family in different tissues across periods. The treatment methods of all samples were based on those published by Liu et al.[44]. All data of *AHL* genes were normalized (log2(FPKM+1)), and a heatmap was drawn using TBtools-II (v2.149) software [38].

4.6. *RNA Extraction and RT-qPCR Analysis*

Capsicum line 6421 was cultivated under controlled conditions at 27°C during the day and 22°C at night, with a 16-hour light and 8-hour dark photoperiod. Total RNA was isolated using the RNAprep Pure Plant Kit (DP421, TIANGEN, China) following the manufacturer's protocol and subsequently reverse-transcribed into cDNA using the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific, China). Quantitative real-time PCR (RT-qPCR) was performed on a QuantStudio 3 Real-Time PCR System (ABI Life, USA) using 2× ChamQ Universal SYBR qPCR Master Mix (TransGen, China) in accordance with the provided instructions. The CaActin7 gene was used as the reference gene, with the forward primer sequence 5′-CTCGAGCAGTGTTTCCCAGT-3′ and the reverse primer sequence 5′-AGCTTCATCACCCACATAGGC-3′. Gene-specific primers were designed for *CaAHL23* and *CaCYP703A2* as follows: *CaAHL23* forward primer, 5′-GAGTCTAGCGGTGGACCTAT-3′, and reverse primer, 5′-AGGTTCTCTCATCTCCAGGG-3′; *CaCYP703A2* forward primer, 5′-TCAATCGACTCCCTCCTGGT-3′, and reverse primer, 5′-CGGAGATAGACCAATGGCCC-3′. RT-qPCR experiments were conducted with three biological replicates and three technical replicates per sample. Relative gene expression levels were determined using the $2^{-\Delta\Delta Ct}$ method.

4.7. Subcellular Localization

The full-length ORF sequences of *CaAHL23* without the termination codon were cloned into the pCAMBIA1300-GFP vector and transformed into *Agrobacterium tumefaciens* GV3101. *CaAHL23* fusion constructs was transformed into tobacco (*Nicotiana benthamiana*) leaves. After three days, the fluorescence signals were observed and captured using a confocal laser scanning microscope (Zeiss LSM510 META, Germany).

4.8. Y1H Assays

Pepper floral buds (F1-F9) were collected for cDNA library construction via a Matchmaker GAL4 One-Hybrid System (Clontech) following the manufacturer's instructions. A 244-bp fragment (2769-3012 upstream of ATG) from the CaCYP703A2 promoter was synthesized by Tsingke Tech (Beijing, China) and subcloned and inserted into a pAbAi vector, and the pCaCYP703A2pro-AbAi bait was transformed into the yeast strain Y1HGold. Y1H library screening assays were performed against the cDNA expression library using the Matchmaker GAL4 One Hybrid System (Clontech). Competent Y1HGold yeast cells containing pCaCYP703A2pro-AbAi as bait were transformed with the contents of the cDNA library, plated onto selective SD/-Leu media supplemented with 100 ng/mL AbA, and incubated at 30℃. The p53-AbAi Control Vector and pGADT7-Rec53 Control Insert from Clontech were used as the positive controls. Yeast plasmids from positive colonies were extracted and transformed into Escherichia coli DH5α competent cells. The prey fragments from the positive colonies were identified by Sanger sequencing. To verify the interaction of CaAHL23 with the bait, the full-length cDNA sequences of CaAHL23 were amplified and fused in frame with the GAL4 activation domain of pGADT7rec (Clontech), forming pGADT7-CaAHL23.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org. Figure S1: Expression level analysis of CaAHL23 during flower development; Figure S2: CDS and protein sequence alignment of CaCYP703A2 and SlCYP703A2; Table S1: Number statistics of conserved motifs in the CaAHLs.

Author Contributions: Conceptualization, XY.S. and YL.L.; methodology, X.W.; software, XY.S.; validation, YL.L., Y.Z. and XY.S.; formal analysis, QZ.C. and Y.L.; investigation, BQ.T.; resources, F.L. and XX.Z.; data curation, XY.S. and YL.L.; writing—original draft preparation, XY.S.; writing—review and editing, YL.L. and F.L. and XX.Z.; supervision, YL.L.; project administration, F.L. and XX.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by China Agriculture Research System of MOF and MARA（CARS-24-A-15）.

Data Availability Statement: Data is contained within the article and Supplementary Materials.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kim, S.; Park, M.; Yeom, S.-I.; Kim, Y.-M.; Lee, J.M.; Lee, H.-A.; Seo, E.; Choi, J.; Cheong, K.; Kim, K.-T.; et al. Genome Sequence of the Hot Pepper Provides Insights into the Evolution of Pungency in Capsicum Species. *Nat Genet* 2014, *46*, 270–278, doi:10.1038/ng.2877.
2. Mahmood, U.; Li, X.; Fan, Y.; Chang, W.; Niu, Y.; Li, J.; Qu, C.; Lu, K. Multi-Omics Revolution to Promote Plant Breeding Efficiency. *Front Plant Sci* 2022, *13*.
3. Islam, K.; Momo, J.; Rawoof, A.; Vijay, A.; Anusree, V.K.; Kumar, A.; Ramchiary, N. Integrated Use of Molecular and Omics Approaches for Breeding High Yield and Stress Resistance Chili Peppers BT - Smart Plant Breeding for Vegetable Crops in Post-Genomics Era. In; Singh, S., Sharma, D., Sharma, S.K., Singh, R., Eds.; Springer Nature Singapore: Singapore, 2023; pp. 279–335 ISBN 978-981-19-5367-5.

4. Xu, L.; Zheng, S.; Witzel, K.; Van De Slijke, E.; Baekelandt, A.; Mylle, E.; Van Damme, D.; Cheng, J.; De Jaeger, G.; Inzé, D.; et al. Chromatin Attachment to the Nuclear Matrix Represses Hypocotyl Elongation in Arabidopsis Thaliana. *Nat Commun* 2024, *15*, 1286, doi:10.1038/s41467-024-45577-5.
5. Street, I.H.; Shah, P.K.; Smith, A.M.; Avery, N.; Neff, M.M. The AT-Hook-Containing Proteins SOB3/AHL29 and ESC/AHL27 Are Negative Modulators of Hypocotyl Growth in Arabidopsis. *The Plant Journal* 2008, *54*, 1–14, doi:https://doi.org/10.1111/j.1365-313X.2007.03393.x.
6. Xiao, C.; Chen, F.; Yu, X.; Lin, C.; Fu, Y.-F. Over-Expression of an AT-Hook Gene, AHL22, Delays Flowering and Inhibits the Elongation of the Hypocotyl in Arabidopsis Thaliana. *Plant Mol Biol* 2009, *71*, 39–50, doi:10.1007/s11103-009-9507-9.
7. Uzair, M.; Xu, D.; Schreiber, L.; Shi, J.; Liang, W.; Jung, K.-H.; Chen, M.; Luo, Z.; Zhang, Y.; Yu, J.; et al. PERSISTENT TAPETAL CELL2 Is Required for Normal Tapetal Programmed Cell Death and Pollen Wall Patterning1 [OPEN]. *Plant Physiol* 2020, *182*, 962–976, doi:10.1104/pp.19.00688.
8. Tayengwa, R.; Sharma Koirala, P.; Pierce, C.F.; Werner, B.E.; Neff, M.M. Overexpression of AtAHL20 Causes Delayed Flowering in Arabidopsis via Repression of FT Expression. *BMC Plant Biol* 2020, *20*, 559, doi:10.1186/s12870-020-02733-5.
9. Machaj, G.; Grzebelus, D. Characteristics of the AT-Hook Motif Containing Nuclear Localized (AHL) Genes in Carrot Provides Insight into Their Role in Plant Growth and Storage Root Development. *Genes (Basel)* 2021, *12*.
10. Feng, H.; Yang, H.; Sun, H.; Lu, P.; Yan, P.; Zhao, W.; Zeng, L.; Li, Z.; Zhang, L.; Hou, W. Characterization of AHL Transcription Factors and Functional Analysis of IbAHL10 in Storage Root Development in Sweetpotato. *Sci Hortic* 2024, *338*, 113718, doi:https://doi.org/10.1016/j.scienta.2024.113718.
11. Favero, D.S.; Kawamura, A.; Shibata, M.; Takebayashi, A.; Jung, J.-H.; Suzuki, T.; Jaeger, K.E.; Ishida, T.; Iwase, A.; Wigge, P.A.; et al. AT-Hook Transcription Factors Restrict Petiole Growth by Antagonizing PIFs. *Current Biology* 2020, *30*, 1454-1466.e6, doi:10.1016/j.cub.2020.02.017.
12. Jeong, H.-N.; Sun, H.-J.; Zuo, Z.-F.; Lee, D.H.; Song, P.-S.; Kang, H.-G.; Lee, H.-Y. Overexpression of ATHG1/AHL23 and ATPG3/AHL20, Arabidopsis AT-Hook Motif Nuclear-Localized Genes, Confers Salt Tolerance in Transgenic Zoysia Japonica. *Plant Biotechnol Rep* 2020, *14*, 351–361, doi:10.1007/s11816-020-00606-9.
13. Howden, A.J.M.; Stam, R.; Martinez Heredia, V.; Motion, G.B.; ten Have, S.; Hodge, K.; Marques Monteiro Amaro, T.M.; Huitema, E. Quantitative Analysis of the Tomato Nuclear Proteome during Phytophthora Capsici Infection Unveils Regulators of Immunity. *New Phytologist* 2017, *215*, 309–322, doi:https://doi.org/10.1111/nph.14540.
14. Rayapuram, N.; Jarad, M.; Alhoraibi, H.M.; Bigeard, J.; Abulfaraj, A.A.; Völz, R.; Mariappan, K.G.; Almeida-Trapp, M.; Schlöffel, M.; Lastrucci, E.; et al. Chromatin Phosphoproteomics Unravels a Function for AT-Hook Motif Nuclear Localized Protein AHL13 in PAMP-Triggered Immunity. *Proceedings of the National Academy of Sciences* 2021, *118*, e2004670118, doi:10.1073/pnas.2004670118.
15. Zhou, L.; Liu, Z.; Liu, Y.; Kong, D.; Li, T.; Yu, S.; Mei, H.; Xu, X.; Liu, H.; Chen, L.; et al. A Novel Gene OsAHL1 Improves Both Drought Avoidance and Drought Tolerance in Rice. *Sci Rep* 2016, *6*, 30264, doi:10.1038/srep30264.
16. Zhao, L.; Lü, Y.; Chen, W.; Yao, J.; Li, Y.; Li, Q.; Pan, J.; Fang, S.; Sun, J.; Zhang, Y. Genome-Wide Identification and Analyses of the AHL Gene Family in Cotton (Gossypium). *BMC Genomics* 2020, *21*, 69, doi:10.1186/s12864-019-6406-6.
17. Wang, M.; Chen, B.; Zhou, W.; Xie, L.; Wang, L.; Zhang, Y.; Zhang, Q. Genome-Wide Identification and Expression Analysis of the AT-Hook Motif Nuclear Localized Gene Family in Soybean. *BMC Genomics* 2021, *22*, 361, doi:10.1186/s12864-021-07687-y.
18. Kumar, A.; Singh, S.; Mishra, A. Genome-Wide Identification and Analyses of the AHL Gene Family in Rice (Oryza Sativa). *3 Biotech* 2023, *13*, 248, doi:10.1007/s13205-023-03666-0.
19. Bishop, E.H.; Kumar, R.; Luo, F.; Saski, C.; Sekhon, R.S. Genome-Wide Identification, Expression Profiling, and Network Analysis of AT-Hook Gene Family in Maize. *Genomics* 2020, *112*, 1233–1244, doi:https://doi.org/10.1016/j.ygeno.2019.07.009.

20. Yun, J.; Kim, Y.-S.; Jung, J.-H.; Seo, P.J.; Park, C.-M. The AT-Hook Motif-Containing Protein AHL22 Regulates Flowering Initiation by Modifying FLOWERING LOCUS T Chromatin in Arabidopsis*. *Journal of Biological Chemistry* 2012, *287*, 15307–15316, doi:https://doi.org/10.1074/jbc.M111.318477.
21. Xu, Y.; Gan, E.-S.; Ito, T. The AT-Hook/PPC Domain Protein TEK Negatively Regulates Floral Repressors Including MAF4 and MAF5. *Plant Signal Behav* 2013, *8*, e25006, doi:10.4161/psb.25006.
22. Lou, Y.; Xu, X.-F.; Zhu, J.; Gu, J.-N.; Blackmore, S.; Yang, Z.-N. The Tapetal AHL Family Protein TEK Determines Nexine Formation in the Pollen Wall. *Nat Commun* 2014, *5*, 3855, doi:10.1038/ncomms4855.
23. Karami, O.; Rahimi, A.; Mak, P.; Horstman, A.; Boutilier, K.; Compier, M.; van der Zaal, B.; Offringa, R. An Arabidopsis AT-Hook Motif Nuclear Protein Mediates Somatic Embryogenesis and Coinciding Genome Duplication. *Nat Commun* 2021, *12*, 2508, doi:10.1038/s41467-021-22815-8.
24. 24. Karlgren, A.; Gyllenstrand, N.; Källman, T.; Sundström, J.F.; Moore, D.; Lascoux, M.; Lagercrantz, U. Evolution of the PEBP Gene Family in Plants: Functional Diversification in Seed Plant Evolution . *Plant Physiol* 2011, *156*, 1967–1977, doi:10.1104/pp.111.176206.
25. Ahmad, Z.; Tian, D.; Li, Y.; Aminu, I.M.; Tabusam, J.; Zhang, Y.; Zhu, S. Characterization, Evolution, Expression and Functional Divergence of the DMP Gene Family in Plants. *Int J Mol Sci* 2024, *25*.
26. Hofer, J.; Ellis, N. Conservation and Diversification of Gene Function in Plant Development. *Curr Opin Plant Biol* 2002, *5*, 56–61, doi:https://doi.org/10.1016/S1369-5266(01)00228-X.
27. Feng, L.-Y.; Lin, P.-F.; Xu, R.-J.; Kang, H.-Q.; Gao, L.-Z. Comparative Genomic Analysis of Asian Cultivated Rice and Its Wild Progenitor (Oryza Rufipogon) Has Revealed Evolutionary Innovation of the Pentatricopeptide Repeat Gene Family through Gene Duplication. *Int J Mol Sci* 2023, *24*.
28. Qanmber, G.; Liu, J.; Yu, D.; Liu, Z.; Lu, L.; Mo, H.; Ma, S.; Wang, Z.; Yang, Z. Genome-Wide Identification and Characterization of the PERK Gene Family in Gossypium Hirsutum Reveals Gene Duplication and Functional Divergence. *Int J Mol Sci* 2019, *20*.
29. Qiao, X.; Li, Q.; Yin, H.; Qi, K.; Li, L.; Wang, R.; Zhang, S.; Paterson, A.H. Gene Duplication and Evolution in Recurring Polyploidization–Diploidization Cycles in Plants. *Genome Biol* 2019, *20*, 38, doi:10.1186/s13059-019-1650-2.
30. Chen, G.; He, W.; Guo, X.; Pan, J. Genome-Wide Identification, Classification and Expression Analysis of the MYB Transcription Factor Family in Petunia. *Int J Mol Sci* 2021, *22*.
31. Cao, L.; Lu, X.; Zhang, P.; Wang, G.; Wei, L.; Wang, T. Systematic Analysis of Differentially Expressed Maize ZmbZIP Genes between Drought and Rewatering Transcriptome Reveals BZIP Family Members Involved in Abiotic Stress Responses. *Int J Mol Sci* 2019, *20*.
32. Zhou, Y.; Zhang, X.; Chen, J.; Guo, X.; Wang, H.; Zhen, W.; Zhang, J.; Hu, Z.; Zhang, X.; Botella, J.R.; et al. Overexpression of AHL9 Accelerates Leaf Senescence in Arabidopsis Thaliana. *BMC Plant Biol* 2022, *22*, 248, doi:10.1186/s12870-022-03622-9.
33. Zhao, J.; Favero, D.S.; Peng, H.; Neff, M.M. Arabidopsis Thaliana AHL Family Modulates Hypocotyl Growth Redundantly by Interacting with Each Other via the PPC/DUF296 Domain. *Proceedings of the National Academy of Sciences* 2013, *110*, E4688–E4697, doi:10.1073/pnas.1219277110.
34. Yang, X.; Wu, D.; Shi, J.; He, Y.; Pinot, F.; Grausem, B.; Yin, C.; Zhu, L.; Chen, M.; Luo, Z.; et al. Rice CYP703A3, a Cytochrome P450 Hydroxylase, Is Essential for Development of Anther Cuticle and Pollen Exine. *J Integr Plant Biol* 2014, *56*, 979–994, doi:https://doi.org/10.1111/jipb.12212.
35. Ma, H.; Wu, Y.; Lv, R.; Chi, H.; Zhao, Y.; Li, Y.; Liu, H.; Ma, Y.; Zhu, L.; Guo, X.; et al. Cytochrome P450 Mono-Oxygenase CYP703A2 Plays a Central Role in Sporopollenin Formation and Ms5ms6 Fertility in Cotton. *J Integr Plant Biol* 2022, *64*, 2009–2025, doi:https://doi.org/10.1111/jipb.13340.
36. Liu, F.; Zhao, J.; Sun, H.; Xiong, C.; Sun, X.; Wang, X.; Wang, Z.; Jarret, R.; Wang, J.; Tang, B.; et al. Genomes of Cultivated and Wild Capsicum Species Provide Insights into Pepper Domestication and Population Differentiation. *Nat Commun* 2023, *14*, 5487, doi:10.1038/s41467-023-41251-4.
37. Mistry, J.; Chuguransky, S.; Williams, L.; Qureshi, M.; Salazar, G.A.; Sonnhammer, E.L.L.; Tosatto, S.C.E.; Paladin, L.; Raj, S.; Richardson, L.J.; et al. Pfam: The Protein Families Database in 2021. *Nucleic Acids Res* 2021, *49*, D412–D419, doi:10.1093/nar/gkaa913.

38. Chen, C.; Wu, Y.; Li, J.; Wang, X.; Zeng, Z.; Xu, J.; Liu, Y.; Feng, J.; Chen, H.; He, Y.; et al. TBtools-II: A "One for All, All for One" Bioinformatics Platform for Biological Big-Data Mining. *Mol Plant* 2023, *16*, 1733–1742, doi:10.1016/j.molp.2023.09.010.
39. Horton, P.; Park, K.-J.; Obayashi, T.; Fujita, N.; Harada, H.; Adams-Collier, C.J.; Nakai, K. WoLF PSORT: Protein Localization Predictor. *Nucleic Acids Res* 2007, *35*, W585–W587, doi:10.1093/nar/gkm259.
40. Bailey, T.L.; Elkan, C. Fitting a Mixture Model by Expectation Maximization to Discover Motifs in Bipolymers. *Proc Int Conf Intell Syst Mol Biol* 1994, *2*, 28–36.
41. Chao, J.-T.; Kong, Y.-Z.; Wang, Q.; Sun, Y.-H.; Gong, D.-P.; Lv, J.; Liu, G.-S. MapGene2Chrom, a Tool to Draw Gene Physical Map Based on Perl and SVG Languages. *Yi Chuan* 2015, *37*, 91–97, doi:10.16288/j.yczz.2015.01.013.
42. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol* 2021, *38*, 3022–3027, doi:10.1093/molbev/msab120.
43. Letunic, I.; Bork, P. Interactive Tree Of Life (ITOL) v5: An Online Tool for Phylogenetic Tree Display and Annotation. *Nucleic Acids Res* 2021, *49*, W293–W296, doi:10.1093/nar/gkab301.
44. Liu, F.; Yu, H.; Deng, Y.; Zheng, J.; Liu, M.; Ou, L.; Yang, B.; Dai, X.; Ma, Y.; Feng, S.; et al. PepperHub, an Informatics Hub for the Chili Pepper Research Community. *Mol Plant* 2017, *10*, 1129–1132, doi:10.1016/j.molp.2017.03.005.