Erzincan Üniversitesi
Fen Bilimleri Enstitüsü Dergisi
2024,17(1), 147-163
ISSN: 1307-9085, e-ISSN: 2149-4584
Araştırma Makalesi

Erzincan University
Journal of Science and Technology
2024,17(1), 147-163
DOI: 10.18185/erzifbed.1375233
Research Article

Genome-Wide Analysis and Characterization of the *AHL* Gene Family in Common Beans (*Phaseolus vulgaris* L.)

Yaren Bozkurt[1], Merve Yüce[2], Esra Yaprak[1], Ayşe Gül Kasapoğlu[1], Emre İlhan[1], Murat Turan[1], Murat Aydın[3], Ertan Yıldırım[2*]

[1]Erzurum Technical University, Faculty of Science, Department of Molecular Biology and Genetics, Erzurum/Türkiye
[2]Ataturk University, Faculty of Agriculture, Department of Horticulture, Erzurum/Türkiye
[3]Ataturk University, Faculty of Agriculture, Department of Agricultural Biotechnology, Erzurum/Türkiye

Received: 12/10/2023, **Revised:** 18/11/2023, **Accepted:** 18/11/2023, **Published:** 28/03/2024

Abstract

The AT-hook motif nuclear localized family (*AHL*) is defined as a small DNA-binding protein motif, functioning as a transcription factor. This transcription factor family plays a significant role in influencing plant growth, development processes, and the response mechanism to environmental stresses. In *Phaseolus vulgaris* (common bean) genome, 41 *AHL* genes have been identified. Using in silico bioinformatics tools, the characteristic features of AHL genes, their exon-intron structures, chromosomal locations of *AHL* genes, conserved motifs, promoter regions, duplication relationships, protein features of AHL proteins, protein-protein interactions, and the expression levels of *AHL* genes against drought and salinity stresses have been studied. Additionally, a phylogenetic comparison of *P. vulgaris* with *Arabidopsis thaliana* and *Glycine max* has been conducted. The amino acid lengths of these proteins vary between 167 and 422, with the molecular weights of the gene family ranging from 18.19 kDa to 45.13 kDa. The isoelectric points (pI) of the AHL proteins vary between 4.5 and 10.3. *AHL* genes are dispersed across all chromosomes of the bean, with the highest number of genes found on chromosomes 6 and 8. In *P. vulgaris*, 14 duplicated genes have been identified, and subsequent Ka/Ks analysis has revealed that all are subject to purifying selection. The findings from this research will aid future functional studies in better understanding the functions of *AHL* genes in beans.

Keywords: *AHL* motif, DNA-binding protein, PPC/DUF296, Genome-Wide Analysis

*Corresponding Author: ertanyil@atauni.edu.tr
Yaren BOZKURT, https://orcid.org/0009-0003-0213-5398
Merve YÜCE, https://orcid.org/0000-0003-1091-0609
Esra YAPRAK, https://orcid.org/0000-0002-8753-494X
Ayşe Gül KASAPOĞLU, https://orcid.org/0000-0002-6447-4921
Emre İLHAN, https://orcid.org/00000-0002-8404-7900
Murat TURAN, https://orcid.org/0000-0003-2900-1755
Murat AYDIN, https://orcid.org/0000-0003-1091-0609
Ertan YILDIRIM, https://orcid.org/0000-0003-3369-0645

Fasulye'de (*Phaseolus vulgaris* L.) *AHL* Gen Ailesinin Genom Çaplı Analizi ve Karakterizasyonu

Öz

AT-hook motifi nükleer lokalize ailesi (*AHL*), küçük bir DNA bağlayıcı protein motifi olarak tanımlanmış ve bir transkripsiyon faktörü olarak görev yapmaktadır. Bu transkripsiyon faktörü ailesi, bitki büyüme ve gelişme süreçlerini ve çevresel streslere verilen yanıt mekanizmasını etkilemede önemli bir rol oynamaktadır. *Phaseolus vulgaris* (fasulye) genomunda 41 *AHL* geni tanımlanmıştır. İn silico biyoinformatik araçlar kullanılarak, AHL genlerinin karakteristik özellikleri, ekzon-intron yapıları, *AHL* genlerinin kromozomlardaki yerleşimleri, korunmuş motifleri, promotor bölgeleri, duplikasyon ilişkileri, AHL proteinlerinin protein özellikleri, protein-protein etkileşimleri ve kuraklık ile tuzluluk streslerine karşı *AHL* genlerinin ifade düzeyleri incelenmiştir. Ayrıca, *P. vulgaris*'in *Arabidopsis thaliana* ve *Glycine max* ile filogenetik ilişkisi belirlenmiştir. Bu proteinlerin amino asit uzunlukları 167 ile 422 ve moleküler ağırlıkları ise 18.19 kDa ile 45.13 kDa arasında değişmektedir. AHL proteinlerinin izoelektrik noktaları (pI) 4,5 ile 10,3 arasında değişmektedir. Pvul-*AHL* genleri fasulyenin bütün kromozomlarına yayılmış olup, en fazla gen 6 ve 8 numaralı kromozomlarda bulunmaktadır. *P. vulgaris*'te 14 duplike olmuş gen tespit edilmiş ve ardından yapılan Ka/Ks analizi, tümünün arındırıcı seleksiyona tabi olduğunu göstermiştir. Bu araştırmadan elde edilen bulgular, gelecekte yapılacak fonksiyonel çalışmaların fasulyedeki *AHL* genlerinin işlevlerini daha iyi anlamamıza yardımcı olacaktır.

Anahtar Kelimeler: *AHL* motif, DNA bağlanma proteini, PPC/DUF296, Genome-Wide Analysis

1. Introduction

P. vulgaris are acknowledged as the primary food supply for millions of people in underdeveloped nations and a significant source of total protein, nutrients, and energy on a global scale. 100 grams of *P. vulgaris* provide 60 grams of carbs, 20 to 25 grams of protein, and 0.7 to 1.5 grams of fat. The *P. vulgaris* structure also includes these macromolecules as well as varying levels of other vitamins, minerals, and phytochemicals [1, 2].

Plant genomes have expanded and diversified rapidly from their common ancestor throughout the evolutionary process, beginning with the colonization of land plants. In this process, some genes that regulate important growth and development processes of plant species increased in number and gradually became multi-member gene families [3, 4, 5]. One of these gene families, the AT-hook motif nuclear localized (*AHL*) family, a small DNA binding protein motif, serves as a transcription factor and is found in all ordered dicot and monocot land plants. AT-hook motif and the plants and Prokaryotes Conserved Domain (*PPC/DUF296*) found in *AHL* proteins are two highly conserved domains [5].

The DNA-binding protein AHL was first discovered in the high mobility group (HMG-I/Y) of mammalian nonhistone chromosomes, and later discovered in prokaryotes and plants [5, 6, 7]. According to conserved amino acid sequences, AHL falls into two main categories: Type-I motifs with the consensus sequences Gly-Ser-Lys-Asn-Lys and Type-II motifs with the Arg-Lys-Tyr. The Arg-Gly-Arg-Pro sequence is preserved in both type 1 and type 2 motifs. The primary distinction between the conserved Arg-Gly-Arg sequence and other transcription factors is its ability to bind to the AT-rich minor groove of B-form DNA. About 120 amino acids make up the PPC domain, which is found at the carboxyl terminus of the AT-hook motif and is in charge of the nucleus localization of the AHL protein [8]. Their nuclear localization and interactions with other proteins are controlled by the PPC domain. They can also interact with some transcription factors and form homo- or hetero-oligomers with other AHLs [9].

AHL gene family members play important roles in different growth and development processes such as pollen development [10], flower development [11], regulation of flowering time [12], hypocotyl elongation [5], leaf aging [13], differentiation of vascular tissues [5], homeostasis of gibberellins [14], and regulation of auxin-related gene expression with jasmonic acid [15]. In addition, *AHL*s are involved in the response to biotic and abiotic stress [16]. In a study conducted in rice, the OsAHL1 gene was reported to positively regulate drought-affected genes to increase tolerance and resistance to abiotic stimuli such as salt and cold [17]. The AHL genes are known to have significant implications in several aspects of plant growth and development, as well as in the plant's ability to respond to both biotic and abiotic stressors. These genes achieve this by modulating the expression of target genes or by engaging in protein-protein interactions with other molecules [18].

So far, varying numbers of AHL gene family members have been identified in different plants: 29 in *Arabidopsis thaliana* (L.) Heynh. [14], 20 in *Oryza sativa* L. [18], 22 in *Sorghum bicolor* (L.) Moench [5], 37 in *Zea mays* L. [16], 48 in *Gossypium raimondii* Ulbr., 51 in *Gossypium arboreum* L., 99 in *Gossypium hirsutum* L. [20], 47 in cultivated carrot (*Daucus carota* subsp. *sativus*) [21], 63 in *Glycine max* L. Merr [22], 37 in *Populus trichocarpa* [8], 14 in *Vitis vinifera* L. [23] and 122 in *Brassica napus* L. [24], 42 in *Brassica rapa* L. [25] and 22 genes in *peach (Prunus persica)* [26]. However, the *AHL* gene family has not been characterized in the *Phaseolus vulgaris*. Thus, the aim of this study was to determine the *AHL* gene family in *P. vulgaris* and to characterize it genome-wide by in silico methods. This study possesses a high level of originality as it is the first in the literature to focus on the *AHL* genes of *P. vulgaris*.

2. Material and Method

Determination of *AHL* Genes

The sequences of the *AHL* gene family in *P. vulgaris* [27], *A. thaliana* [28] and *G. max* [29] genomes were obtained from the Phytozome database v13 [30] using the Pfam Accession Number (PF02178) [31]. In addition, The SMART database (Simple Modular Architecture Research Tool) was used to confirm the presence of AHL proteins [32].

Sequence Alignment and Phylogenetic Analyses

The protein sequences of the members of the *AHL* gene family in the genomes of *P. vulgaris*, *A. thaliana* and *G. max* species were aligned using the Multiple Sequence Alignment by CLUSTALW tool [33] and the phylogenetic tree was constructed using MEGA 11 program by the neighbor-joining (NJ) method with 1000 bootstrap replicates. Subsequently, the aligned protein sequences were used to construct a phylogenetic tree using the ITOL (Interactive Tree of Life) tool [34].

Identification of AHL Proteins in the *P. vulgaris* Genome

The amino acid count, molecular weight (kDa), theoretical isoelectric point (pI), and stability of AHL proteins in *P. vulgaris* were determined using the "ProtParam tool" [35].

Determination of the structure and chromosome location and promoter region analysis

The exon and intron regions of the *P. vulgaris* AHL proteins were determined using the GSDS (Gene Structure Display Server v2.0) [36]. The chromosomal locations and sizes of *P. vulgaris AHL* genes were determined via the Phytozome database v13 and mapped with the MapChart program [37]. Each member of the *P. vulgaris AHL* gene family had roughly 2000 bp of DNA extracted from the 5` upstream region, and cis acting element analysis was done using the PlantCARE database [38].

Gene duplications of *AHL* genes and identification of conserved motifs

Multiple EM for Motif Elimination (MEME) tool [39] was used to detect conserved motifs of *P. vulgaris AHL* genes. The maximum number of motifs was determined as 10 and the width range was at least 2 and maximum 50. Gene duplications were determined using the MCScanX (The Multiple Collinearity Scan Toolkit) tool in TBtools program. Synonym ratios (Ks), nonsynonymous ratios (Ka), and evolutionary strains (Ka/Ks) between binary pairs of genes were calculated using the TBtools program [40].

Synteny analysis

The duplications of *AHL* genes in *P. vulgaris*, *A. thaliana*, and *G. max* were identified using the MCScanX tool (The Multiple Collinearity Scan Toolkit) and were visualized using the TBtools program [40].

Protein-protein interactions (PPI)

Pvul-AHL protein-protein interactions (PPI) were identified at the physical, functional, and experimental levels using the STRING database [41].

In silico gene expression analysis

Illumina RNA-seq data were obtained from the Sequence Read Archive (SRA) data bank in the National Center for Biotechnology Information (NCBI) database. We used accession numbers for salt and drought conditions to locate pertinent RNA-seq data. The following leaf types were used: leaf under salt stress (SRR957668), leaf under drought stress (SRR8284481), and leaf under drought control (SRR8284480) [42]. Gene expression levels were normalized using the Read per Kilobase (RPKM) technique [43]. The CIMMiner tool was used to create a heatmap [44].

3. Results and Discussion

Determination and Characteristics of *AHL* Genes

AHL genes corresponding to *P. vulgaris*, *A. thaliana*, and *G. max* were extracted from the Phytozome database using the Pfam Accession Number. In this process, 41 genes from *P. vulgaris*, 32 genes from *A. thaliana*, and 65 genes from *G. max* were identified. These genes were then systematically numbered based on their chromosomal sequences. To enhance clarity in the manuscript, abbreviations were used for the nomenclature of the *AHL* genes specific to *P. vulgaris*. The details of these genes, including their abbreviations, Phytozome Transcript IDs, chromosomal locations, start and end points, and strand orientations, have been presented in Table 1. Upon examining the AHL proteins of *P. vulgaris*, it was observed that the amino acid counts of the genes ranged from 167 to 422. The protein with the shortest amino acid sequence was identified as *Pvul-AHL*-19 (167 amino acids), while the one with the longest sequence was *Pvul-AHL*-15 (422 amino acids). A correlation was noted between their molecular weights and amino acid lengths. Among these proteins, 21 were found to exhibit acidic or near-neutral acidic characteristics, and 22 proteins exhibited basic or near-neutral basic characteristics. In the analysis of the 41 AHL proteins, 6 were determined to be stable, while the remaining were identified as unstable (Table 1).

Table 1. Information about *Pvul-AHL* genes

Gene ID	Phytozome Transcript ID	Chr	Start	End	Strand	aa	MW (kDa)	pI	Stabitily
Pvul-AHL-1	Phvul.001G081632.1	Chr01	12159318	12160837	reverse	278	28.80	5.54	unstable
Pvul-AHL-2	Phvul.001G234600.1	Chr01	48832896	48834455	reverse	287	30.43	5.55	unstable
Pvul-AHL-3	Phvul.002G006700.1	Chr02	698485	703324	forward	351	36.74	8.34	unstable
Pvul-AHL-4	Phvul.002G006900.1	Chr02	722290	723601	reverse	272	28.61	5.8	unstable
Pvul-AHL-5	Phvul.002G118500.1	Chr02	25245881	25247521	reverse	208	23.14	4.86	unstable
Pvul-AHL-6	Phvul.002G151600.1	Chr02	30512278	30513948	forward	299	31.19	6.63	unstable
Pvul-AHL-7	Phvul.002G157300.1	Chr02	31183111	31187446	reverse	355	36.29	9.94	unstable
Pvul-AHL-8	Phvul.002G288900.1	Chr02	45766457	45770953	reverse	358	36.51	9.71	unstable
Pvul-AHL-9	Phvul.003G195100.1	Chr03	41924534	41925104	forward	189	20.24	5.61	unstable
Pvul-AHL-10	Phvul.003G216000.1	Chr03	44318578	44323527	reverse	326	33.28	9.08	unstable
Pvul-AHL-11	Phvul.003G216200.1	Chr03	44377183	44379047	forward	284	29.84	6.11	unstable
Pvul-AHL-12	Phvul.003G230500.1	Chr03	46212036	46217232	forward	368	37.36	9.24	unstable
Pvul-AHL-13	Phvul.003G230600.1	Chr03	46229664	46230522	reverse	285	30.07	5.87	unstable
Pvul-AHL-14	Phvul.004G127200.1	Chr04	42321598	42327649	forward	334	34.77	5.87	unstable
Pvul-AHL-15	Phvul.005G127900.1	Chr05	36564911	36567434	reverse	422	45.13	4.87	unstable
Pvul-AHL-16	Phvul.006G007200.1	Chr06	1056593	1057256	reverse	220	23.49	7.13	unstable
Pvul-AHL-17	Phvul.006G007300.1	Chr06	782344	782962	forward	205	22.76	4.5	unstable
Pvul-AHL-18	Phvul.006G013800.1	Chr06	6443800	6444580	reverse	259	28.12	8.9	unstable
Pvul-AHL-19	Phvul.006G016300.1	Chr06	7617513	7620976	reverse	167	18.19	10.3	stable
Pvul-AHL-20	Phvul.006G031000.1	Chr06	11868121	11868781	forward	219	23.69	7.71	stable
Pvul-AHL-21	Phvul.006G031100.1	Chr06	11882688	11883287	forward	188	20.22	5.67	stable
Pvul-AHL-22	Phvul.006G031200.1	Chr06	11927780	11928392	forward	203	21.90	9.75	stable
Pvul-AHL-23	Phvul.006G106500.2	Chr06	21587154	21592842	reverse	363	38.46	9.34	unstable
Pvul-AHL-24	Phvul.007G129400.1	Chr07	14122140	14123194	reverse	310	31.76	6.26	unstable
Pvul-AHL-25	Phvul.007G111600.1	Chr07	20569440	20576763	reverse	331	34.96	8.37	unstable
Pvul-AHL-26	Phvul.007G223900.1	Chr07	34761172	34765536	forward	377	39.11	6.85	stable
Pvul-AHL-27	Phvul.007G271800.1	Chr07	39308747	39309554	reverse	268	27.17	5.3	unstable
Pvul-AHL-28	Phvul.008G050100.1	Chr08	4352524	4353889	forward	271	27.99	5.83	unstable
Pvul-AHL-29	Phvul.008G065900.1	Chr08	6037682	6044398	forward	341	35.16	9.86	unstable
Pvul-AHL-30	Phvul.008G076100.1	Chr08	7332658	7337374	forward	332	34.47	9.4	unstable
Pvul-AHL-31	Phvul.008G181100.1	Chr08	51225879	51226665	reverse	261	27.29	7.72	unstable
Pvul-AHL-32	Phvul.008G192600.1	Chr08	53291708	53292560	reverse	283	29.51	4.79	unstable
Pvul-AHL-33	Phvul.008G231500.1	Chr08	53537135	53538349	reverse	248	26.45	9.14	unstable
Pvul-AHL-34	Phvul.008G200900.1	Chr08	54741078	54741993	forward	253	27.48	5.1	unstable
Pvul-AHL-35	Phvul.008G264600.1	Chr08	61064853	61065423	reverse	189	20.39	5.86	unstable
Pvul-AHL-36	Phvul.009G008300.1	Chr09	1270665	1274490	reverse	369	37.70	10	unstable

Pvul-AHL-37	Phvul.009G144300.1	Chr09	21607296	21608142	reverse	281	29.32	6.79	unstable
Pvul-AHL-38	Phvul.010G085300.1	Chr10	23954897	23956568	reverse	310	31.92	7.2	unstable
Pvul-AHL-39	Phvul.010G084600.1	Chr10	24191274	24197814	forward	356	36.56	9.08	unstable
Pvul-AHL-40	Phvul.010G097000.1	Chr10	36212885	36218337	reverse	340	34.87	9.87	unstable
Pvul-AHL-41	Phvul.011G089400.1	Chr11	8654261	8657512	forward	373	38.93	5.74	stable

*:aa: aminoacid length, pI: theoretical isoelectric points, Chr: Chromosome.

Phylogenetic Analysis of AHL Genes

The AHL genes' protein sequences from *P. vulgaris*, *A. thaliana*, and *G. max* were aligned in ClustalW, and a phylogenetic tree was constructed using the Mega11 program. The ITOL online tool was used to colorize the phylogenetic tree branching. According to this analysis, AHL proteins are basically divided into two classes: Clade A and Clade B. Clade A contains Type I AHL proteins, while Clade B contains Type II and Type III AHL proteins. There are 16 bean, 16 *A. thaliana*, and 31 soybean AHL proteins in the Clade A group. In the Clade B group, there are 25 bean, 16 *A. thaliana*, and 34 soybean AHL proteins. When the phylogenetic tree was examined, Type I *AHL*s were clearly separated from the others, while Type II and Type III *AHL*s showed mixed branching (Figure 1). A similar branching pattern was observed in the phylogenetic trees generated from maize, rice, sorghum, and *A. thaliana* [45] and *P. trichocarpa*, *A. thaliana*, and *O. sativa* [8].

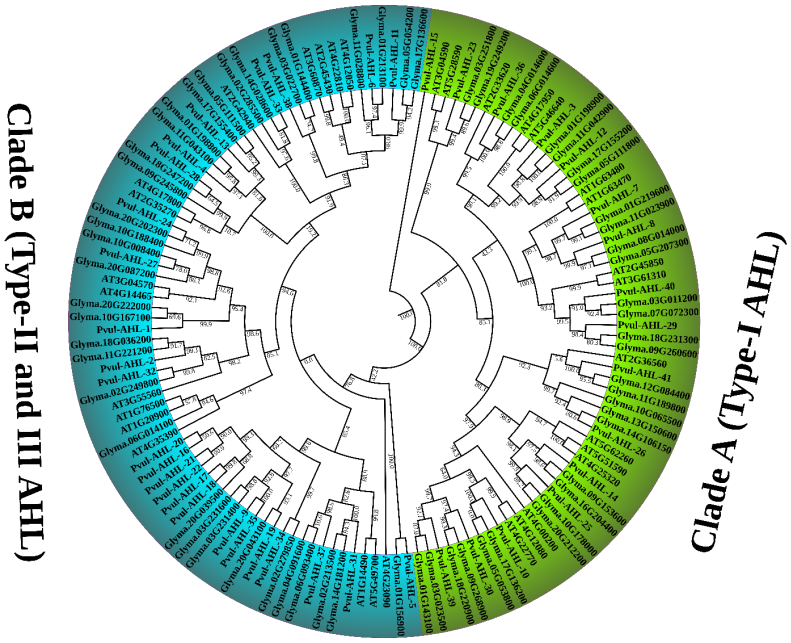

Figure 1. Phylogenetic tree illustrating the relationship of *AHL* family in *P. vulgaris*, *A. thaliana* and *G. max*. The phylogenetic tree was drawn with AHL proteins from P. vulgaris, A. thaliana and G. max. AHL full-length amino acid sequences from P. vulgaris and 2 other species are aligned by ClustalW and the phylogenetic tree was constructed using MEGA 11 program by the neighbor-joining (NJ) method with 1000 bootstrap replicates. AHL proteins are basically divided into two classes: Clade A (green) and Clade B (blue).

Structure of *AHL* Genes

The similarity in the intron/exon structures of homologous and paralogous genes across different species can be utilized as an indicator to assess the evolutionary proximity or distance between these species [46]. It has been discovered that some genes of *P. vulgaris* contain only one exon region, while others have multiple. The highest number of exons (6) and introns (5) were observed in the *Pvul-AHL-23* gene. Variability in the structures and numbers of exon and intron regions was observed to be dependent on the branches in the phylogenetic tree of *P. vulgaris*. In 19 genes, no intron regions were present, containing only a single exon region. Additionally, 16 genes were found to have 5 exon regions and 4 intron regions. It was determined that the identified *Pvul-AHL* genes are located on all chromosomes of *P. vulgaris* (Figure 2). The 6th and 8th chromosomes were found to have the highest number of *AHL* genes, with 8 each (Figure 3).

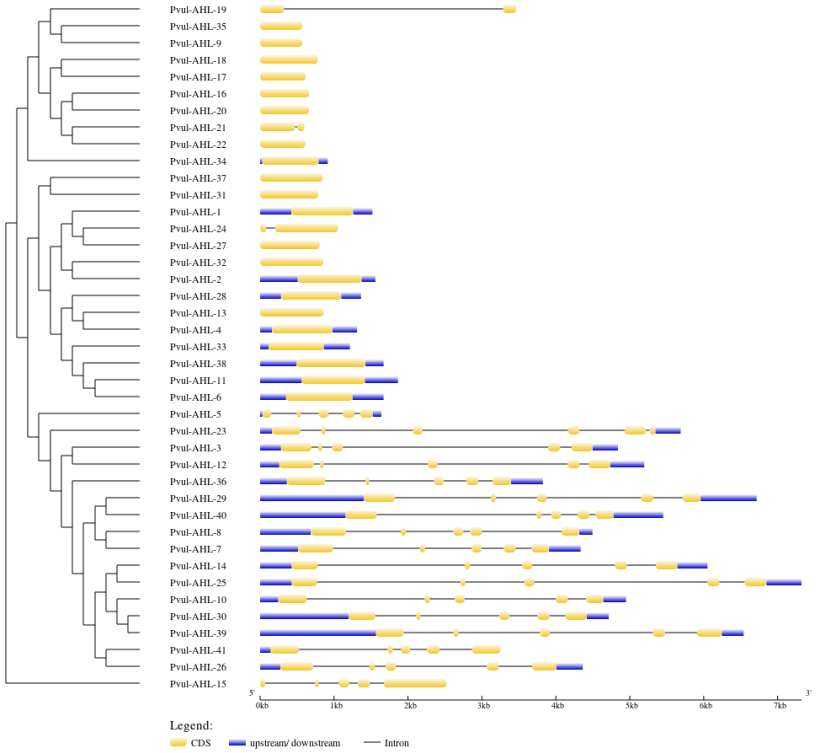

Figure 2. The length and position of exons and introns in the Pvul-AHL genes. The blue lines represent 5′-UTR or 3′-UTR, yellow boxes indicate exons, and black lines exhibit introns.

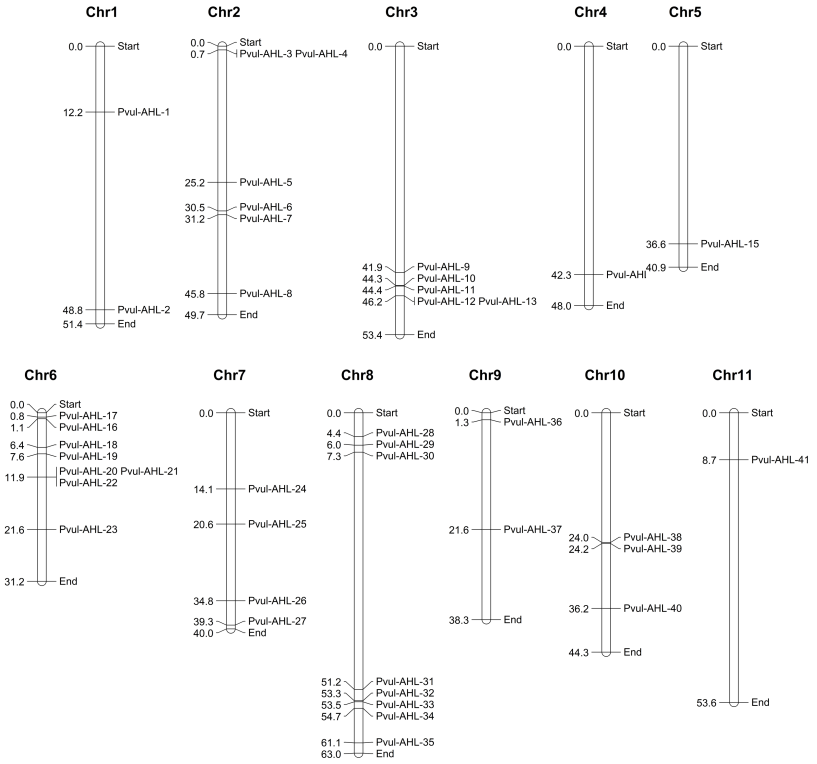

Figure 3. Chromosomal locations of *P. vulgaris* AHL genes. Distribution of the AHL genes on *P. vulgaris* chromosomes according to the linkage map.

Conserved Motifs of *AHL* Genes

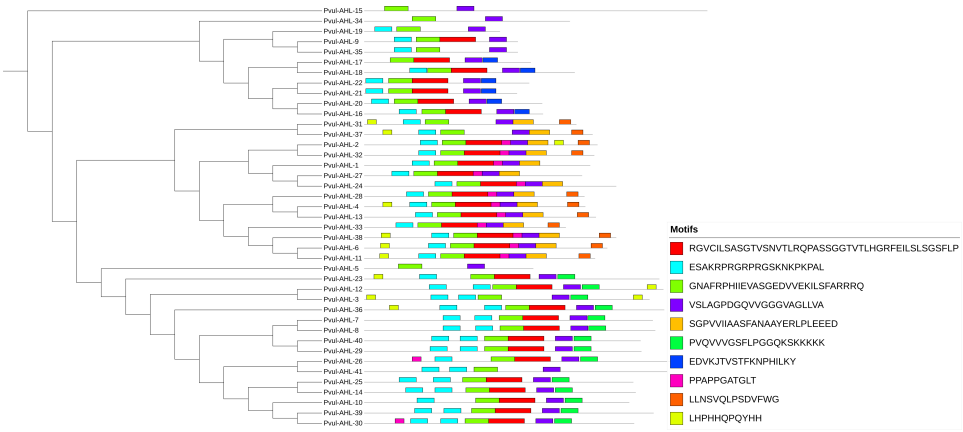

Figure 4. Distribution of predicted motifs in *AHL* genes in *P. vulgaris.* Predicted motif distribution in Pvul-AHL proteins identified MEME suite program. The motifs are exhibited with a specific color.

The conserved motifs within the *AHL* gene family were delineated using MEME-Suite. In the course of this analysis, a total of ten distinct conserved motif patterns were successfully identified, facilitating further examination and review of the *AHL* gene family. The range of

identified motifs varied between 2 and 8. It was noted that *Pvul-AHL-15* and *Pvul-AHL-34* exhibited the least number of conserved motifs, with only 2 each, whereas *Pvul-AHL-2*, *Pvul-AHL-4*, *Pvul-AHL-6* and *Pvul-AHL-11* were characterized by the highest number, each containing 8 conserved motifs. Motif 3 and Motif 7 were detected in all *Pvul-AHL* proteins. Motif 1, which was determined to be *AHL* domain, was detected in all other genes except *Pvul-AHL-5* gene (Figure 4).

Promoter region analysis of *AHL* Genes

Sequences retrieved from 2000 bp upstream in the 5′ upstream region of *AHL* genes have been examined, and it has been determined that promoter regions in these genes are influential in plant development, molecular response to abiotic stresses, and adaptation to environmental factors. Cis-acting elements found in sequences of *Pvul-AHL* genes have been identified through analyses in the PlantCARE database and visualized using the TBTools program. A comprehensive identification of cis-acting elements has been achieved across all Pvul-*AHL* genes. Key elements associated with abiotic and biotic stresses, such as MYB (in all except *Pvul-AHL-12*), TC-rich repeats (in *Pvul-AHL-2, Pvul-AHL-6, Pvul-AHL-8, Pvul-AHL-10, Pvul-AHL-13, Pvul-AHL-15, Pvul-AHL-16, Pvul-AHL-21, Pvul-AHL-24, Pvul-AHL-26, Pvul-AHL-29, Pvul-AHL-30, Pvul-AHL-36, Pvul-AHL-38, Pvul-AHL-41*), LTR (in *Pvul-AHL-2, Pvul-AHL-20, Pvul-AHL-21, Pvul-AHL-24, Pvul-AHL-29, Pvul-AHL-30, Pvul-AHL-31, Pvul-AHL-33*), W box (in *Pvul-AHL-1, Pvul-AHL-4, Pvul-AHL-5, Pvul-AHL-11, Pvul-AHL-12, Pvul-AHL-14, Pvul-AHL-16, Pvul-AHL-19, Pvul-AHL-20, Pvul-AHL-21, Pvul-AHL-22, Pvul-AHL-25, Pvul-AHL-27, Pvul-AHL-31, Pvul-AHL-38, Pvul-AHL-41*) have been observed in *Pvul-AHL* genes (Figure 5).

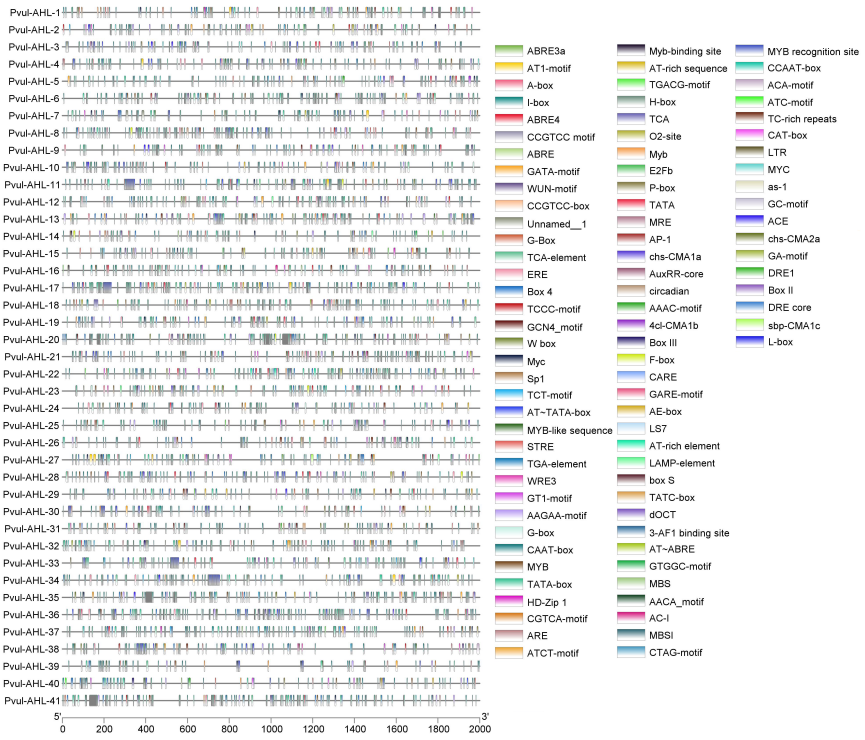

Figure 5. Promoter regions of *Pvul-AHL* genes. PlantCARE analyzed promoter sequences (−2000 bp) of 41 *Pvul-AHL* genes. The scale indicates the upstream length through the translation codon. Different color rectangles indicate different cis elements.

Synteny Analysis of *AHL* Genes

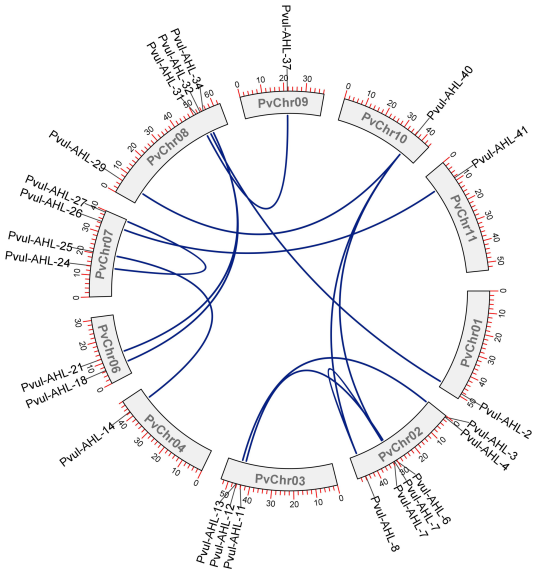

Figure 6. Self-synteny of the AHL genes in the *P. vulgaris* genome. The AHL genes in *P. vulgaris* were mapped to different chromosomes. Gene pairs of the AHL with a syntenic relationship are joined by a blue line.

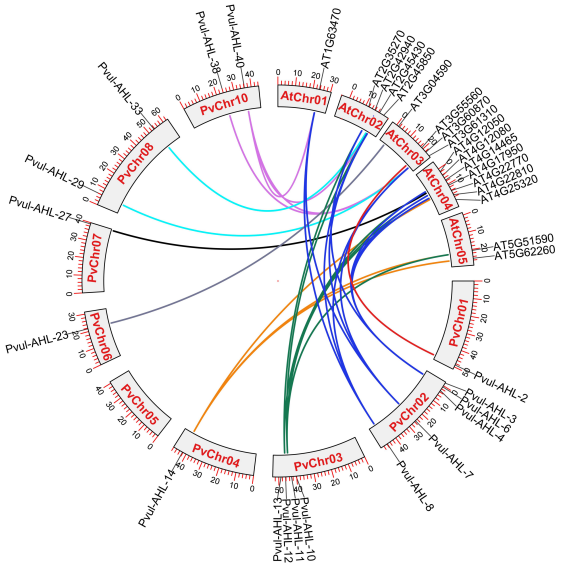

Figure 7. Gene duplication and synteny analysis of AHL genes between *P. vulgaris* and *A. thaliana.* It has been indicated that arrows originating from all chromosomes of *P. vulgaris* are represented in different colors.

In the synteny analysis conducted using MCScanX, duplicated genes among *P. vulgaris* genes, their interconnections, and the resulting selection types have been examined. Additionally, duplicated genes between *P. vulgaris* and *A. thaliana* have been studied. Within *P. vulgaris*, 14 duplicated genes have been identified, and subsequent Ka/Ks analysis has revealed that all are subject to purifying selection (Table 2). In the synteny analysis of *AHL* genes within *P. vulgaris*, it has been determined that duplicated genes exist in all chromosomes except for *Pvul-AHL-5* (Figure 6). When examining duplicated genes between *P. vulgaris* and *A. thaliana*, duplicated genes have been found in all chromosomes of *A. thaliana* (Figure 7).

Table 2. Gene duplications of the *Pvul-AHL* genes

Gene 1	Gene 2	Ka/Ks	Selection Type
Pvul-AHL-2	*Pvul-AHL-32*	0.22	Purifying selection
Pvul-AHL-3	*Pvul-AHL-12*	0.24	Purifying selection
Pvul-AHL-4	*Pvul-AHL-13*	0.11	Purifying selection
Pvul-AHL-6	*Pvul-AHL-11*	0.07	Purifying selection
Pvul-AHL-7	*Pvul-AHL-40*	0.17	Purifying selection
Pvul-AHL-7	*Pvul-AHL-8*	0.30	Purifying selection
Pvul-AHL-8	*Pvul-AHL-40*	0.20	Purifying selection
Pvul-AHL-14	*Pvul-AHL-25*	0.28	Purifying selection
Pvul-AHL-18	*Pvul-AHL-34*	0.34	Purifying selection
Pvul-AHL-21	*Pvul-AHL-34*	0.38	Purifying selection
Pvul-AHL-24	*Pvul-AHL-27*	0.07	Purifying selection
Pvul-AHL-26	*Pvul-AHL-41*	0.31	Purifying selection
Pvul-AHL-29	*Pvul-AHL-40*	0.15	Purifying selection
Pvul-AHL-31	*Pvul-AHL-37*	0.22	Purifying selection

Protein-protein interaction (PPI) Analysis of *AHL* Genes

The protein-protein interactions of *Pvul-AHL* proteins have been analyzed and visualized using the STRING database. It was found that *Pvul-AHL*-9, *Pvul-AHL*-17, *Pvul-AHL*-20, *Pvul-AHL*-34 are interrelated, as are *Pvul-AHL*-15, *Pvul-AHL*-21, *Pvul-AHL*-39 with each other. No relationships have been found among all the other genes (Figure 8).

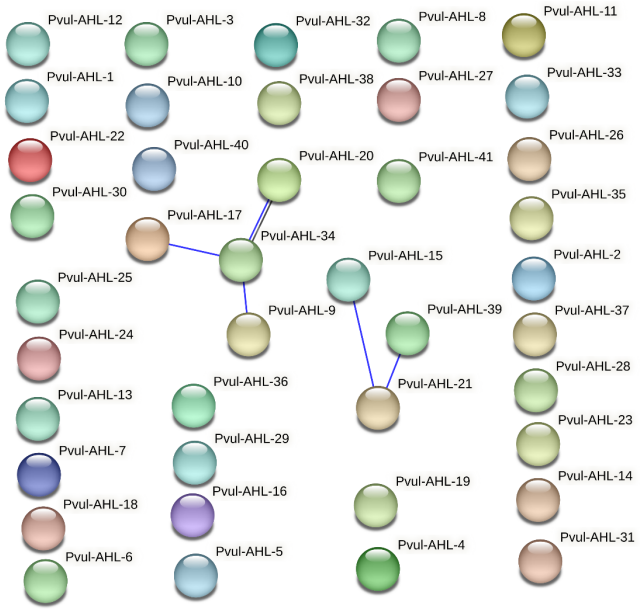

Figure 8. Protein–protein interactions of identified AHL proteins. Protein-protein interactions of *Pvul-AHL* proteins have been analyzed and visualized using the STRING database. Related genes are shown with blue line.

Expression Analysis of *AHL* Genes

The expression profiles of *AHL* genes in leaf tissues of *P. vulgaris* under salt and drought stress were examined using information from the NCBI SRA repository. In drought stress, the expression levels of *Pvul-AHL-19*, *Pvul-AHL-22*, *Pvul-AHL-34* and *Pvul-AHL-40* genes increased compared to the control. In salt stress, the expression levels of *Pvul-AHL-23*, *Pvul-AHL-29*, *Pvul-AHL-36*, *Pvul-AHL-40* and *Pvul-AHL-41* genes increased compared to the control. No change was observed in the expression levels of *Pvul-AHL-5*, *Pvul-AHL-17*, *Pvul-AHL-18*, *Pvul-AHL-19*, *Pvul-AHL-20*, *Pvul-AHL-21*, *Pvul-AHL-22*, *Pvul-AHL-34* and *Pvul-AHL-35* genes under salt stress (Figure 9).

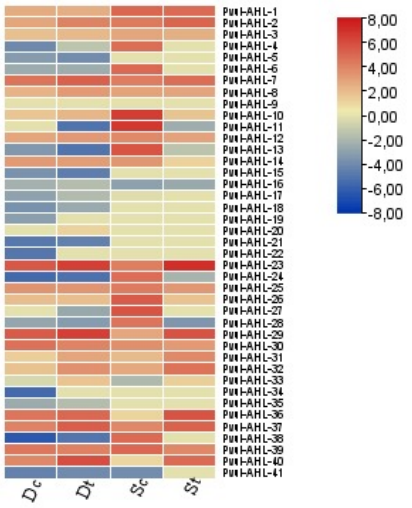

Figure 9. Heat map of the *Pvul-AHL* genes that are differentially expressed in leaf tissue under salt and drought stress. C: control, T: treatment, S: salt, D: drought. The expression levels are represented according to the color bar.

4. Conclusions

An essential transcription factor in plants, the AHL protein controls a variety of biological processes. *AHL* have been the subject of ongoing research into their evolution and role ever since their discovery. The expression differences of *Pvul-AHL* genes were determined in the gene expression analyzes performed in silico in leaf tissues under salt and drought stress. This evaluation will serve as a starting point for additional functional research and crop breeding. The results of this study will shed light on future studies with *P. vulgaris*.

Author Contributions

All authors contributed equally to the writing of this manuscript.

Acknowledgements

This study was presented at the "The 6th Symposium on EuroAsian Biodiversity (SEAB) 2023" and published in the Symposium Book as an abstract.

References

[1] Petry, N., Boy, E., Wirth, J.P., and Hurrell, R.F., (2015) Review: The potential of the common bean (*Phaseolus vulgaris*) as a vehicle for iron biofortification, *Nutrients*, 7, 1144–1173. doi: 10.3390/nu7021144. Physiol., 129, 500–15.

[2] Pereira, W. J., Melo, A. T. D. O., Coelho, A. S. G., Rodrigues, F. A., Mamidi, S., Alencar, S. A. D., Vianello, R. P., (2020) Genome-wide analysis of the transcriptional response to drought stress in root and leaf of common bean, *Genetics and Molecular Biology*, *43*.

[3] Lynch, M., Conery, J. S., (2003) The origins of genome complexity, *Science*, *302*(5649), 1401-1404.

[4] Bowman, J. L., Floyd, S. K., Sakakibara, K., (2007) Green genes—comparative genomics of the green branch of life, *Cell*, *129*(2), 229-234.

[5] Zhao, J., Favero, D. S., Peng, H., Neff, M. M., (2013) 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*, *110*(48), E4688-E4697.

[6] Goodwin, G. H., Sanders, C., Johns, E. W., (1973) A new group of chromatin-associated proteins with a high content of acidic and basic amino acids, *European Journal of Biochemistry*, *38*(1), 14-19.

[7] Eckner, R., Birnstiel, M. L., (1989) Cloning of cDNAs coding for human HMG I and HMG Y proteins: both are capable of binding to the octamer sequence motif, *Nucleic acids research*, *17*(15), 5947-5959.

[8] Wang, H., Leng, X., Yang, J., Zhang, M., Zeng, M., Xu, X., Li, C., (2021) Comprehensive analysis of AHL gene family and their expression under drought stress and ABA treatment in *Populus trichocarpa*, *PeerJ*, *9*, e10932.

[9] Širl, M., Šnajdrová, T., Gutiérrez-Alanís, D., Dubrovsky, J. G., Vielle-Calzada, J. P., Kulich, I., Soukup, A., (2020) At-hook motif nuclear localised protein 18 as a novel modulator of root system architecture, *International journal of molecular sciences*, *21*(5), 1886.

[10] Lou, Y., Xu, X. F., Zhu, J., Gu, J. N., Blackmore, S., Yang, Z. N., (2014) The tapetal AHL family protein TEK determines nexine formation in the pollen wall, *Nature Communications*, *5*(1), 1-9.

[11] Jin, Y., Luo, Q., Tong, H., Wang, A., Cheng, Z., Tang, J., Zhu, L., (2011) An AT-hook gene is required for palea formation and floral organ number control in rice, *Developmental biology*, *359*(2), 277-288.

[12] Xu, Y., Gan, E. S., Ito, T., (2013) The AT-hook/PPC domain protein TEK negatively regulates floral repressors including MAF4 and MAF5, *Plant signaling & behavior*, *8*(8), e25006.

[13] Lim, P. O., Kim, Y., Breeze, E., Koo, J. C., Woo, H. R., Ryu, J. S., Nam, H. G., (2007) Overexpression of a chromatin architecture-controlling AT-hook protein extends leaf longevity and increases the post-harvest storage life of plants, *The Plant Journal, 52*(6), 1140-1153.

[14] Matsushita, A., Furumoto, T., Ishida, S., Takahashi, Y., (2007) AGF1, an AT-hook protein, is necessary for the negative feedback of AtGA3ox1 encoding GA 3-oxidase, *Plant physiology*, *143*(3), 1152-1162.

[15] Rashotte, A. M., Carson, S. D., To, J. P., Kieber, J. J., (2003) Expression profiling of cytokinin action in *Arabidopsis*, *Plant Physiology*, *132*(4), 1998-2011.

[16] Bishop, E. H., Kumar, R., Luo, F., Saski, C., Sekhon, R. S., (2020) Genome-wide identification, expression profiling, and network analysis of AT-hook gene family in maize, *Genomics*, *112*(2), 1233-1244.

[17] Zhou, L., Liu, Z., Liu, Y., Kong, D., Li, T., Yu, S., Mei, H., Xu, X., Liu, H., Chen, L., & Luo, L. (2016). A novel gene OsAHL1 improves both drought avoidance and drought tolerance in rice. Scientific reports, 6, 30264. https://doi.org/10.1038/srep30264.

[18] Zhou, J., Wang, X., Lee, J. Y., Lee, J. Y., (2013) Cell-to-cell movement of two interacting AT-hook factors in Arabidopsis root vascular tissue patterning, *The Plant Cell*, *25*(1), 187-201.

[19] Kumar, A., Singh, S., Mishra, A., (2023) Genome-wide identification and analyses of the AHL gene family in rice (*Oryza sativa*), *3 Biotech*, *13*(7), 248.

[20] Zhao, L., Lü, Y., Chen, W., Yao, J., Li, Y., Li, Q., Zhang, Y., (2020) Genome-wide identification and analyses of the AHL gene family in cotton (*Gossypium*), *BMC genomics*, *21*(1), 1-14.

[21] Machaj, G., Grzebelus, D., (2021) 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*, *12*(5), 764.

[22] Wang, M., Chen, B., Zhou, W., Xie, L., Wang, L., Zhang, Y., Zhang, Q. (2021). Genome-wide identification and expression analysis of the AT-hook Motif Nuclear Localized gene family in soybean. *BMC Genomics, 22*(1), 361.

[23] Li, X., He, H., Wang, H., Wu, X., Wang, H., Mao, J., (2021) Identification and expression analysis of the AHL gene family in grape (*Vitis vinifera*), *Plant Gene*, *26*, 100285.

[24] Zhang, W. M., Fang, D., Cheng, X. Z., Cao, J., Tan, X. L., (2021) Insights Into the Molecular Evolution of AT-Hook Motif Nuclear Localization Genes in *Brassica napus, Frontiers in Plant Science*, *12*.

[25] Zhang, X., Li, J., Cao, Y., Huang, J., Duan, Q., (2023) Genome-Wide Identification and Expression Analysis under Abiotic Stress of BrAHL Genes in *Brassica rapa*, *International Journal of Molecular Sciences*, *24*(15), 12447.

[26] Zhao, J., Xu, E., Wang, Q., (2023) Dissection of AT-Hook Motif Nuclear-Localized Genes and Their Potential Functions in Peach Growth and Development, *Forests*, *14*(7), 1404.

[27] Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., Jackson, S. A., (2014) A reference genome for common bean and genome-wide analysis of dual domestications, Nature genetics, 46(7), 707-713.

[28] Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Huala, E. (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools, *Nucleic Acids Research*, *40*(D1), D1202-D1210.

[29] Zhao, J., Favero, D. S., Qiu, J., Roalson, E. H., Neff, M. M., (2014) Insights into the evolution and diversification of the AT-hook Motif Nuclear Localized gene family in land plants, *BMC plant biology*, *14*(1), 1-19.

[30] Goodstein, D.M., Shu S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., Rokhsar, D.S. (2012) Phytozome: a comparative platform for green plant genomics, *Nucleic Acids Research*. *40*(D1), D1178-D1186.

[31] 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., Finn, R.D., Bateman A. (2020) Pfam: The protein families database in 2021, *Nucleic Acids Research*, *49,* D412-D419.

[32] Letunic, I., Khedkar, S., Bork, P. (2021) SMART: recent updates, new developments and status in 2020", *Nucleic Acids Research*, *49*, D458–D460.

[33] Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, *Nucleic Acids Research*, *25*, 4876-4882.

[34] Letunic, I., Bork, P. (2021) Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation, *Nucleic Acids Research*, *49*, W293–W296.

[35] Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A. (2005) Protein Identification and Analysis Tools on the Expasy Server, The Proteomics Protocols Handbook, Humana Press.

[36] Hu, B., Jin, J.¸ Guo, A.Y., Zhang, H., Luo, J., Gao, G. (2015) GSDS 2.0: an upgraded gene feature visualization server, *Bioinformatics*, *31*, 1296-1297.

[37] Voorrips, R.E. (2002) MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs. *Journal of Heredity*, *93*(1), 77–78.

[38] Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rombauts, S. (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences, Nucleic acids research, 30(1), 325-327.

[39] Bailey, T.L., Williams, N., Misleh, C., Li, W.W. (2006) MEME: discovering and analyzing DNA and protein sequence motifs, *Nucleic Acids Research*, *34*, W369-W373.

[40] Chen, C., Chen, H.¸ Zhang, Y., Thomas, H.R., Frank, M.H., He, Y.¸ Xia, R. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data, *Molecular Plant*, *13*(8), 1194-1202, 2020.

[41] Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., Jensen, L.J., Mering, C.V. (2019) STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, *Nucleic Acids Research*, *47*(D1), D607–D613.

[42] Hiz, M. C., Canher, B., Niron, H., Turet, M., (2014) Transcriptome analysis of salt tolerant common bean (*Phaseolus vulgaris* L.) under saline conditions, *PloS one*, *9*(3), e92598.

[43] Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., Wold, B., (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq, Nature methods, 5(7), 621-628.

[44] Luna, A., Elloumi, F., Varma, S., Wang, Y., Rajapakse, V.N., Aladjem, M.I., Robert, J., Sander, C., Pommier, Y., Reinhold, W.C. (2021) CellMiner Cross-Database (CellMinerCDB) version 1.2: Exploration of patient-derived cancer cell line pharmacogenomics. *Nucleic Acids Research*, *49*(D1), D1083-D1093.

[45] Bishop, E. H., Kumar, R., Luo, F., Saski, C., & Sekhon, R. S. (2020) Genome-wide identification, expression profiling, and network analysis of AT-hook gene family in maize. *Genomics*, *112*(2), 1233-1244.

[46] Rakhimzhanova, A., Kasapoğlu, A.G., Sapakova, A., İlhan, E., Zharmukhametova, R., Turan, M., Zekenova, L., Muslu, S., Kazhygeldiyeva, L., Aydın, M., Çiltaş, A. (2023) Expression analysis and characterization of the CPP gene family of Melatonin-treated common bean cultivars under different abiotic stresses, *South African Journal of Botany*, *160*, 282-294.