INT. J. BIOL. BIOTECH., 22 (1): 39-52, 2025.

GENOME-WIDE ANALYSIS OF THE *YABBY* TRANSCRIPTION FACTOR FAMILY IN *SPINACIA OLERICIA* AND FUNCTIONAL IDENTIFICATION OF *SPOVYABBY* GENES

Ali Abbas[1,3], Attique Ur Rehman[2,3], Muhammad Asrar[7], Qurban Ali[3], Shahbaz Ahmad[4], Muhammad Ali*[4], Muhammad Ashfaq*[3], Shahbnum Shaheen[6], Mukhdoom Shahjahan Bukhari[5], Urooj Mubashar[8] and Sharmin Ashraf[9]

[1]*National Nanfan Research Institute (Sanya), Chinese Academy of Agricultural Sciences, Sanya 572024, China*
[2]*Biotechnology Research Institute, GSCAAS 100081, Beijing, China*
[3]*Department of Plant Breeding and Genetics, University of the Punjab Lahore, Pakistan*
[4]*Department of Entomology, University of the Punjab, Lahore, Pakistan*
[5]*Agricultural Research Station, Bahawalpur, Pakistan;*
[6]*Department of Botany, Lahore College for Women University, Lahore, Pakistan*
[7]*Department of Zoology, Government College University, Faisalabad, Pakistan*
[8]*Government Girls High School No2 Ghakkhar, Gujranwala, Pakistan*
[9]*Barani Agricultural Research Institute, Chakwal, Pakistan*
Corresponding authors email: ashfaq.iags@pu.edu.pk; ali.iags@pu.edu.pk

ABSTRACT

The *YABBY* gene family with DNA binding domain and a plant-specific transcription factor that performs different functions including; the regulation of style, flower length, and polarity development of lateral organs in flowering plants. In this study, the spinach genome (*Spinacia olericia*) was employed as a reference to identify *YABBY* gene family members using computational methods. The research encompassed an analysis of gene structures, chromosome locations, protein motifs, phylogenetic, synteny, transcriptomics, and miRNA targets that revealing various characteristics of the *YABBY* gene family in *Spinacia olericia*. The findings showed that six specific *YABBY* genes were irregularly distributed across all 12 chromosomes, forming five subgroups (Subfamilies *YAB*5, *YAB*3/AFO, INO, CRC, and *YAB*2) based on the established Arabidopsis classification. Segmental duplication was identified as the prevalent mode of gene tandem duplication in spinach. Transcriptomic analysis highlighted the high expression of *SpovYABBY*5 and *SpovYABBY*1 during leaf development, suggesting their involvement in oxalate content formation in spinach. Cis-regulatory elements (CREs) analysis revealed elements responsive to light, ABA hormone, drought induction, cell cycle regulation, seed and meristem expression. Comparing spinach and Arabidopsis *YABBY* genes indicated common characteristics among five subfamilies. This comprehensive genome analysis sheds light on the distribution and role of *YABBY* genes in spinach and provides valuable insights for cloning and functional studies. The study enhances our understanding of the *YABBY* gene family in *Spinacia olericia* in comparison to Arabidopsis thaliana.

Keywords: *YABBY* Transcription Factor, *Spinacia olericia*, *SpovYABBY* genes, oxalate content, Genome wide analysis

INTRODUCTION

The *YAB*BY gene family in plants is a significant transcription factor that influences various aspects of plant development (Sun *et al*., 2021). It regulates the length of floral styles, enhances resistance to environmental stresses, promotes the growth of lateral plant structures, and contributes to the elongation of these structures (Hussain *et al*., 2021; Ma *et al*., 2021; Zhang et al., 2020; Nurani *et al*., 2020). Moreover, it improves the efficiency of plant hormonal responses, leading to better-performing vascular structures and nectaries (Phukela *et al*., 2020). Additionally, it plays a role in increasing seed germination rates and post-germination growth (Romanova *et al*., 2021). The *YAB*BY gene has two main domains in the N and C terminals with zinc finger proteins. The zinc finger region with the N terminal and C-terminal has a *YAB*BY region. There are some amino acid residues in both these domains, and specific DNA binding occurs through these domains (Zhang *et al*., 2019).

In *Arabidopsis thaliana*, six genes are encoded by the *YAB*BY gene family, which is separated into five subfamilies: *YAB*1, *YAB*2, *YAB*5, CRABS CLAW (CRC), and INNER NO OUTER (INO). The *YAB*2 gene is involved in the formation of flowers, CRC genes are involved in carpel orientation and nectary advancement, INO genes are involved in the maturation of the external integument of ovules, and *YAB*5/*YAB*3 are involved in the developmental process of leaves and cotyledons. Additionally, the *YAB*BY transcription factor is crucial for lateral

development and growth of leaves, flower improvement, carpel improvement, ovule improvement, vein formation in leaves, seed breaking in cereals, plant responses to abiotic stress, and fruit ripening.

Hormones such as auxin, abscisic acid (ABA), and gibberellins impact the *YAB*BY gene. In the family Amaranthaceae, spinach (*Spinacia oleracea* L., 2n= 2X=12) is one of the most nutritious vegetables, widely consumed worldwide (Belic *et al*., 2021). It is a leafy vegetable, and beet, quinoa, and amaranth also belong to this group (Ribera *et al*., 2021). Spinach follows an annual plant pattern through vegetative and reproductive phases (Van den Eede *et al*., 2004). The global yield of spinach crossed 26.7 million tons in 2016, with a yielding value of 18 billion USD (FASTAI 2018). Denmark produces spinach seeds due to favorable environmental conditions, while China's harvest yield of *Spinacia* is almost 91.5% (Ribera *et al*., 2020). Spinach has wind pollination, and dioecious as well as monoecious species of plants are also present. (Takahata *et al*., 2016).

In terms of nutritional value, spinach supplies a better amount of minerals and vitamins, particularly ascorbic acid (vitamin C) (Takahata et al., 2016). Transcription factors were used to analyze and evaluate the YABBY genes in the spinach genome using diverse bioinformatics tools. A systematic approach was followed to analyze YABBY genes from the spinach genome. The intron/exon partition arrangement and chromosomal distribution, along with conserved domain regulatory aspects, were also investigated. A correlative investigation of YABBY genes from spinach and Arabidopsis thaliana was employed for orthologous interaction, shedding light on their apparent function. The genome-wide evaluation of the YABBY gene was provided as an opportunity for further exploration, and participants of this gene family can be cloned.

MATERIALS AND METHODS

Pattern recognition and information for databases

DNA interaction *YABBY* region was obtained from (http://pfam.xfam.org/) Pfam, as PF04690. The 164 amino acid *YAB*BY domain sequence was extracted from *Arabidopsis thaliana* (Accession no A0A1P8APE2) (Finn *et al*., 2014). The encoding genes of *YAB*BY were identified using the *YAB*BY proteome database at Phytozome v13 (https://phytozome-next.jgi.doe.gov) via the program BLAST-P (Protein-Basic Local Alignment Search Tool). With default parameters, the amino acid sequences retrieved via BLAST-P were uploaded to the NCBI CDD (Conserved Domain Database) (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi/) (Lu *et al*., 2018). The *YAB*BY conserved domain (PF04690) (https://pfam.xfam.org/family/PF04690) was rejected.

Analyzing the physiochemical characteristics, Subcellular Localization and cis element of the *YABBY* proteins in spinach

The chromosomal location and orientation, gene IDs, and protein and gene IDs were collected using Phytozome (https://phytozome-next.jgi.doe.gov/). The number of amino acids, protein range (amino acid residues), molecular weight, and theoretical pI value of *YAB*BY proteins were retrieved using the ProtParam tool (http://web.expasy.org/protparam/) (Hou *et al*., 2019). The *YAB*BY genes were renamed according to their physical position.

The program WoLF PSORT (https://wolfpsort.hgc.jp/) was used to predict the subcellular localization of HaCCO (Horton *et al*., 2016). For visual inspection, a heat map was constructed in TBTools to study the features of different organelles of a cell. A 1000-bp upstream sequence was obtained from the initiation codon of putative Spov*YAB* genes for promoter analysis. The prediction of cis-regulatory elements was retrieved from the PlantCare database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Rombauts et al., 1999) and validated using the PLACE database (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Higo *et al*., 1998; Higo *et al*., 1999).

Detection of conserved motifs and Exon Intron

To find the maximum number of motifs in protein sequences of *SpovYABBY*, the MEME program was used (https://meme-suite.org/meme/tools/meme). The minimum and maximum widths of the motifs were set to 6 and 50, respectively (Bailey *et al*., 2015).

For the exon/intron pre-arrangement of *SpovYABBY* genes, the Phytozome v13 database was used to retrieve identified coding and genomic sequences. Additionally, Phytozome v13 (http://spinachbase.org) was used to obtain the GFF3 file of the spinach genome. These sequences were used in the Gene Structure Display Server v2.0 (GSDS v2.0, available at http://gsds.cbi.pku.edu.cn/) to draw the gene structure (Hu *et al*., 2015).

Several Genomes connections and phylogenetic evaluation

*YAB*BY amino acid sequences were aligned using Clustal W version 2.1 (http://www.clustal.org/clustal2/), and phylogenetic analysis was performed using MEGA 11 v11.0.10 program with the neighbor-joining (NJ) method and

bootstrapping set at 1000 repetitions with pairwise deletion. Six *YAB*BY protein sequences from Arabidopsis and six from spinach were used for phylogenetic analysis (Godini *et al*., 2019; Lau *et al*., 2022).

Assessment of Potential MicroRNA Target Regions

The PmiREN tool was used for target site identification of the 6 *YAB*BY gene family members in spinach (https://www.pmiren.com/). For visual comparison of the CDS sequences of the genes with the mature miRNA sequences, the PsRNA tool was used for further study (https://www.zhaolab.org/psRNATarget/). For visualizing the connections between target genes and predicted miRNAs, the Cytoscape program was used (Mazhar *et al*., 2023).

Chromosomal Location of Spov*YAB* genes family

Phytozome (https://phytozome-next.jgi.doe.gov/) provided details about the Spov*YAB* genes' locations, including the length of the chromosome and their precise location within the chromosome. Using TB-tools software, an image displaying the location of Spov*YAB* genes on the chromosome was created.

Gene Duplication and Synteny evaluation

The duration of deviation within the spinach *YAB*BY gene group was calculated using Ks and Ka assessments. Protein sequence regions were aligned using Clustal W, and the rates of Ka and Ks interrelation were calculated using Mega 11 software with the Nei-Gojobori model. A gamma distribution was used to describe the rate variation between sites (shape parameter = 1).

The ratio of Ka and Ks parameters was configured in the software manuals. The molecular evolution of each paralogous gene pair was estimated by Ka and Ks ratios. The T = Ks / 2r equation was used for estimating divergence time (T), where r represents the neutral substitution rate (7.0 × 10–9 substitutions per site per year) (Xu et al., 2017). Gene duplication was analyzed with default parameters using MCScanX (Multiple Collinearity Scan Toolkit) (Verhoeven et al., 2013). To show the synteny relationship of spinach *YAB*BY genes, syntenic analysis maps were constructed using the Micro Synteny tool in TBtools software.

Analysis of protein-protein Interaction

The analysis of *SpovYABBY* genes was validated using gene ontology (GO) studies. For this purpose, the ShinyGo v0.741 online program was used to determine the function of *SpovYABBY* genes in spinach (https://bioinformatics.sdstate.edu/go74/

Transcriptome-analysis

For transcriptomic analysis, differentially expressed *SpovYABBY* genes in the roots and leaves of spinach cultivars (Bloomsdale and PI175311) with distinct oxalate contents were studied. Their data were retrieved from the NCBI GEO (https://www.ncbi.nlm.nih.gov/geo/) (GSE146711) database to explore oxalate in the *SpovYABBY* gene family.

RESULTS

Analyzing the physiochemical characteristics, Subcellular Localization and cis element of the *YABBY* proteins in spinach

There were six *SpovYABBY* genes gained in Spinach. The length of amino acid in the *YABBY* genes protein range from 187-259 with molecular weight from 20.39-25.04 kD. The *SpovYABBY* 4 has a minimum protein and the *SpovYABBY* 5 has a maximum protein length. The recognized isoelectric points limit is from 4.54 to 9.4 **(Table 1).**

Subcellular localization analysis revealed distinct patterns for the specified genes. *SpovYABBY*1 showed a predominant nuclear localization with 14 instances, while *SpovYABBY*2 displayed a more diversified distribution, with 2 instances in the nucleus, 4 in the chloroplasts, and 1 each in the extracellular space, cytoplasm, mitochondria, and vacuole. *SpovYABBY*3 predominantly localized to the nucleus, with 13 occurrences, and *SpovYABBY*4 exhibited a similar nuclear preference with 11 occurrences, along with 2 instances in the chloroplasts and 1 in the vacuole. *SpovYABBY*5 had a notable nuclear localization with 13 instances, while *SpovYABBY*6 showed a more even distribution, with 2 instances in the nucleus, 4 in the chloroplasts, 3 in the extracellular space, 2 in the cytoplasm, and 1 in the mitochondria (**Fig.1**).

Table 1. Six *YABBY* genes are found in the spinach genome.

	Accession No.	Chromosome No.	Chromosome position (bp)		Direction	Molecular Weight (KD)	Pi value	Number of amino acids	No. of introns	No. of Exons
			start	end						
SpoYABBY1	Spov3_chr5.01174	05	13785462	13789119	R	25.4	9.4	229	6	7
SpoYABBY2	Spov3_chr5.02270	05	36618488	36621731	R	21.02	7.10	189	6	7
SpoYABBY3	Spov3_chr6.01719	06	25016034	25019047	R	23.74	8.45	211	5	6
SpoYABBY4	Spov3_chr2.00033	02	438633	442032	R	20.39	9.24	187	6	7
SpoYABBY5	Spov3_chr5.03629	05	8132667	8132963	R	28.56	7.07	259	5	6
SpoYABBY6	Spov3_chr4.04704	04	118332792	118334653	R	21.82	4.54	190	4	5

Table 2. Represent targeted MiRNA and their sequences.

miRNA_Acc.	Target_Acc.	miRNA_start	miRNA_end	Target_start	Target_end	miRNA_aligned_fragment	alignment	Target_aligned_fragment	Inhibition	Cleavage
Sol-miRN20	Spo*YABBY6*	1	21	107	127	UUACAGAGACUAUCAAUGAUG	... ...	UGUCAAUGGUGGUGACUGUAA		

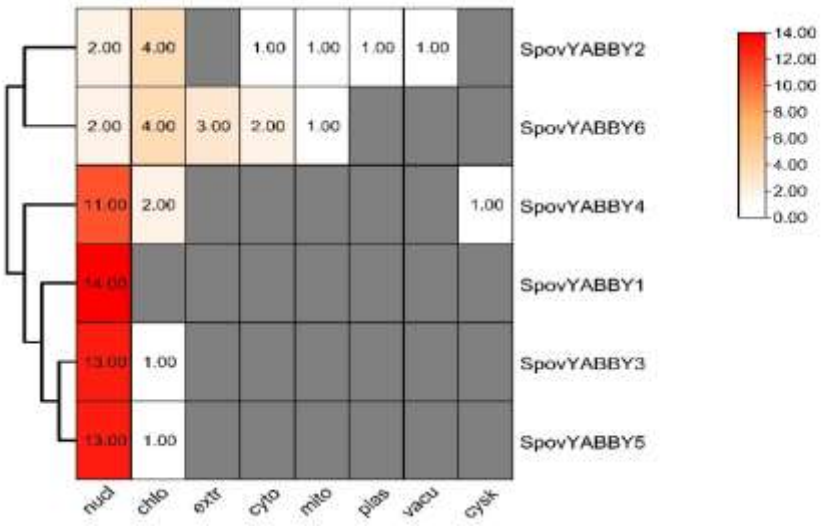

Fig. 1. Heat Map illustrating the sub-cellular localization of all 5 *SpovYABBY* genes to the nucleus, cytoplasm, chloroplast, Golgi-apparatus, mitochondria, plasmid, peroxisomes of the plant cell. White color foretells the smallest functional presence of the relevant gene in the specified region, and red color denotes the highest functional significance of the relevant gene in the indicated region.

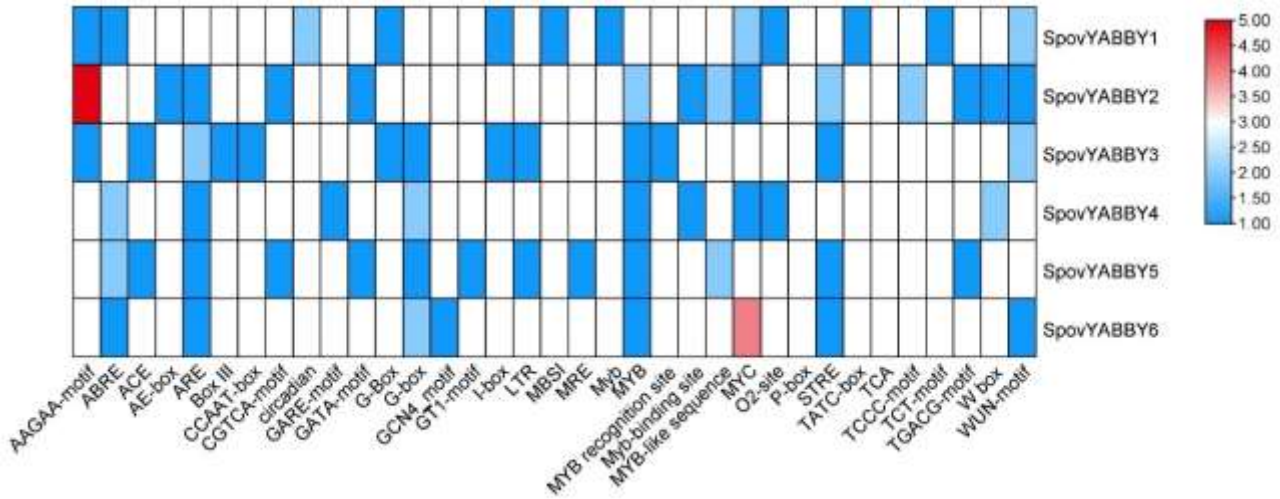

Fig. 2. The visual depiction of the *SpovYABBY* gene's cis regulatory analysis along with the strength of each function. From red (highest) to the white (lowest) intensity is used to describe the biochemical and physiological activities happening in plants.

Cis-regulatory elements with diverse functions include light responsive, abscisic acid, alicyclic acid, gibberellins, anaerobic induction, meristem expression, seed specific regulation, and sensitive regulatory elements. The analysis of cis-elements in the promoter regions of the specified genes reveals specific patterns of regulatory elements associated with each gene. *SpovYABBY*1 was distinguished by the presence of 3 P-box elements, indicating its involvement in gibberellin-responsive transcriptional regulation. *SpovYABBY*2 exhibited a significant abundance of 5 AAGAA-motif elements, hinting at its potential modulation of gene expression through interactions with specific transcription factors or RNA-binding proteins. *SpovYABBY*3 showcased a dominant presence of 3 ABRE elements related to abscisic acid responsiveness, as well as 3 TCA elements associated with salicylic acid responsiveness and 3 MYC elements, emphasizing their likely significance in gene regulation. *SpovYABBY*4 featured a notable number of 3 STRE elements, suggesting its role in stress-responsive gene regulation through interactions with transcription factors. Both *SpovYABBY*5 and *SpovYABBY*6 were marked by multiple MYC elements (3 each), indicating their potential involvement in jasmonic and abscisic acid signaling **(Fig. 2)**.

Detection of conserved motifs and Exon Intron

In each of the *SpovYABBY* proteins, the *YABBY* domain was present. The motif length ranged from a maximum of 11 to a minimum of 5. Among these motifs, motif 1 and motif 2 were conserved. The *YABBY* genes in the identical group were found to have motifs with similar patterns, suggesting that these conserved motifs behaviors are unique to a group or subgroup. All the spinach *YABBY* genes contain the same *YABBY* domain but its arrangement was a little varied in some genes. *YABBY* genes may have arisen as a consequence of gene expansion, according to motifs found in different *YABBY* genes **(Fig. 3)**.

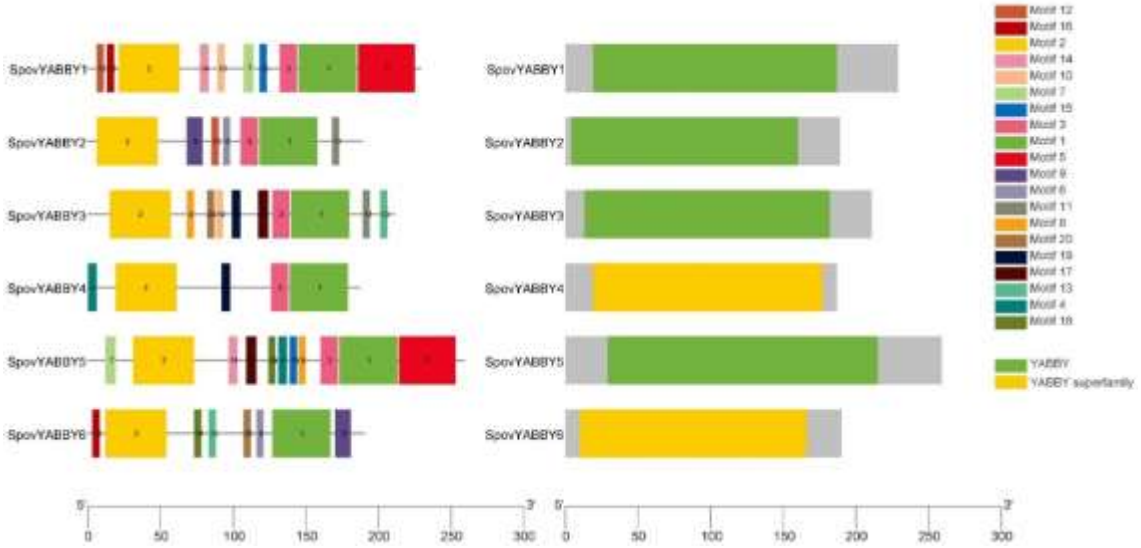

Fig. 3. 20 motifs' locations on 5 Spov*YAB* proteins. By applying MEME version 4.9.0 and integrating it with a phylogenetic tree, we were able to find the motifs in spinach and gain a clear knowledge of their relationship. The bars stand in for themes, with various color codes for various motifs.

All the *SpovYABBY* genes contain introns and exons with a little variation in quantity. *SpovYABBY*1, *SpovYABBY*2, and *SpovYABBY*4 are comprised of 6 introns and 7 exons respectively. *SpovYABBY*6 with 4 introns and exons 5. *SpovYABBY*3 and *SpovYABBY*5 comprised 5 introns and 6 exons respectively. The *SpovYABBY* gene displays a variation in the number of introns, with a maximum of 6 and a minimum of 4, alongside a range in the number of exons, ranging from 7 at the maximum to 5 at the minimum **(Fig. 4)**.

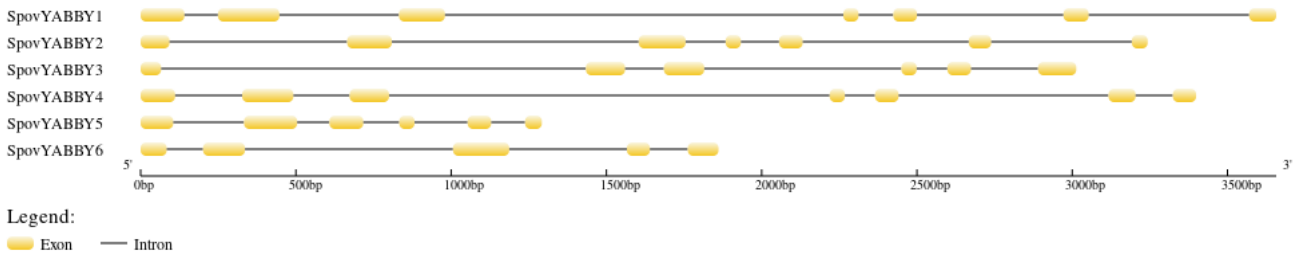

Fig. 4. The intron-exon structure is phylogenetically represented, and it shows the number of introns and exons in several *Spov YABBY* genes has remained constant.

Several Genomes connections and phylogenetic evaluation

The *YABBY* genes of *Spinacia oleracea*, Arabidopsis thaliana, Amianthus chondroid, and Beta vulgaris were examined in terms of their phylogenetic relationships. Phylogenetic analysis revealed that five *YABBY* genes were divided into six subfamilies *YAB*5, *YAB*3/AFO, INO, CRC, and *YAB*2. All *YABBY*. There were 23 *YABBY* genes examined in total. The related *YABBY* proteins within each clade are likely to share similar sequences and functional characteristics due to their close evolutionary relationship **(Fig. 5).**

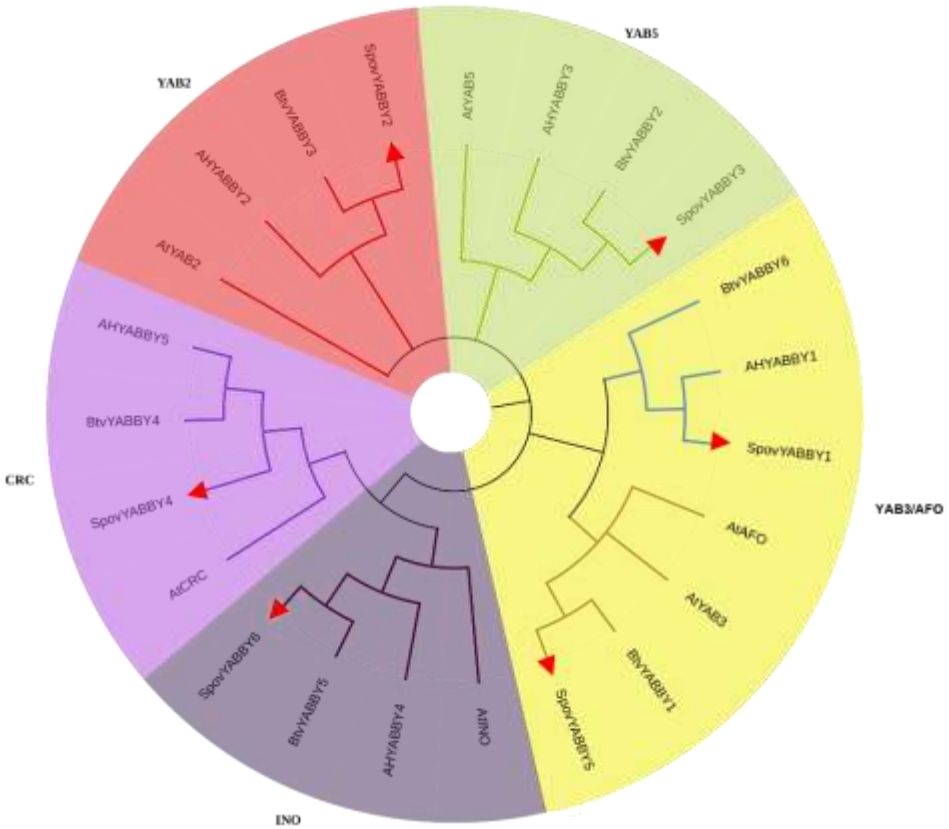

Fig. 5. A red triangle indicates the location of the *Spinacia oleracea* gene. In MEGA 11, the evolutionary analyses were carried out. The NJ technique with 1000 bootstrap replications was used to determine the pattern of evolutionary history.

Assessment of Potential MicroRNA Target Regions

A total of one miRNA were discovered that specifically targeted one of the total of six *SpovYABBY* genes. None of these miRNAs targeted the remaining 5 *SpovYABBY* genes. This miRNA varied in length from 1 to 21 amino acids. This proved that *SpovYABBY* 01 was the sole gene that received the largest amount of functional miRNA targets. Sol-miRN20 play role in signaling axis regulates neural progenitor cell differentiation **Table 2.**

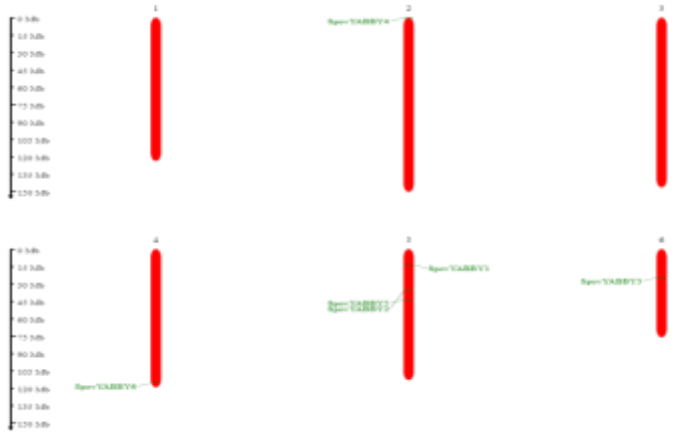

Fig. 6. The *S .oleracea* genome's chromosomal mapping of the *YABBY* genes shows an existence of paralogous copies with plausible locations. During time of selection pressure and chromosomal rearrangement, suffered duplication events. Despite these alterations, the duplication *SpovYABBY* genes were remained to able to keep their original roles and develop stable functional traits in the *S .oleracea* genome.

Chromosomal Location of Spov*YAB* genes family

The *SpovYABBY* genes were distributed across different positions on the *S. oleracea* chromosomes. Notably, *SpovYABBY*1, *SpovYABBY*2, and *SpovYABBY*5 genes existed on chromosome 5, while chromosome 2 contained the *SpovYABBY*4 gene. *SpovYABBY*3 was present on chromosome 6, and *SpovYABBY*6 was located on chromosome 4 **(Fig. 6)**.

Gene Duplication and Synteny evaluation

During the analysis of Ka (nonsynonymous substitutions per site) and Ks (synonymous substitutions per site), genetic divergence and selective pressures acting on gene pairs from the *SpovYABBY* gene family were examined. The Ka/Ks ratio, a critical metric in evolutionary genetics, was calculated for each gene pair to elucidate the nature of selection acting upon them. A Ka/Ks ratio less than 1 is indicative of purifying (negative) selection, equal to 1 suggests neutral evolution, and greater than 1 implies positive selection. The analysis revealed a range of Ka/Ks ratios among the studied gene pairs. The minimum observed Ka/Ks ratio was approximately 0.121, indicating strong purifying selection, which is consistent with the conservation of gene function. On the other end of the spectrum, the maximum Ka/Ks ratio observed was approximately 0.268, suggesting that certain gene pairs may be under a less stringent purifying selection or potentially experiencing positive selection, signifying evolutionary divergence. The divergence time (T) for these gene pairs, which ranges from approximately 1.146 million years to 2.553 million years ago. These divergence times provide valuable insights into the historical evolutionary events and relationships between the *SpovYABBY* genes **(Fig. 7)**.

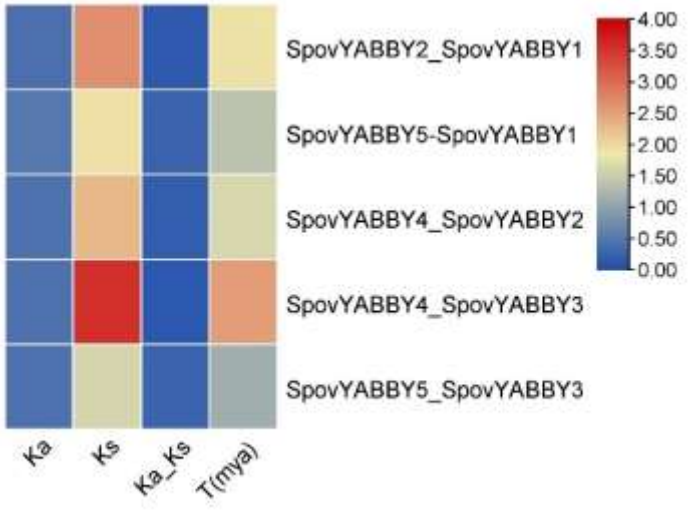

Fig. 7. The Ks (synonymous substitution rate) and Ka (nonsynonymous substitution rate) were estimated using TBTools.

The synteny analysis revealed interesting paralogous relationships within the spinach genome. Specifically, it indicated that the *SpovYABBY*2 gene was found to be paralogous to both the *SpovYABBY*1 and *SpovYABBY*5 genes, all located on chromosome 5. Similarly, the analysis demonstrated that *SpovYABBY*3 was paralogous to *SpovYABBY*4. This information suggested that these *SpovYABBY* genes likely shared a common ancestry and had arisen through gene duplication events in the past. The paralogous relationships observed among these genes implied potential functional similarities or diversification within the *SpovYABBY* gene family **(Figure 8 ab)**.

In a comprehensive dual analysis, the study delved into the chromosomal distribution and intrachromosomal linkages of *YABBY* genes among three distinct plant species: Spinach with *Amianthus chondroid* , *Arabidopsis thaliana*, and *Beta vulgaris*. The findings are visually represented, with red lines indicating the presence of duplicate *YABBY* gene pairs and gray lines denoting the existence of synteny blocks. Notably, chromosome 6 of Spinach exhibited orthology with scaffold-12, chromosome 1 with scaffold-2 and chromosome 2 with scaffold-9 of *A. chondroid*, establishing a clear link between these two species. However, in the case of *A. thaliana*, no apparent orthologue linkage was observed, suggesting a more distant evolutionary relationship. Of particular significance, the highest degree of orthologue linkage was identified between Spinach and Beta vulgaris. Chromosome 6 of Spinach displayed orthologous correspondence with NC-025816.2, chromosome 1 with NC-025817.2, and chromosome 2 with NC-025819.2 of *B. vulgaris*.

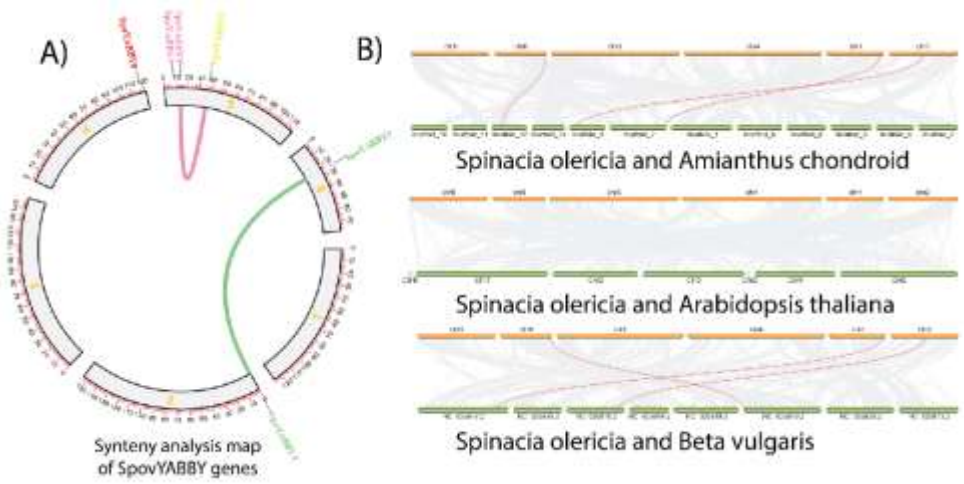

Fig. 8 (a). Distribution of *YABBY* genes on different chromosomes of spinach. **(b)** The collinearity relationship between *S. olericia* with *A. chondroid*, *A. thaliana* and *B. vulgaris*.

Protein-protein Interaction

During a protein interaction study, there were a total of 5 nodes and 2 edges observed (Fig. 9). The average node degree was calculated to be 0.8, while the average local clustering coefficient was found to be 0.4. The expected number of edges in this context was zero, and the p-value for protein-protein interaction enrichment was notably low at 1.13e^-05. This suggests a significant enrichment of interactions. To meet the minimum required interaction score, a low confidence threshold of 0.150 was applied. The *SpovYABBY2* gene that showed protein-protein interaction with respect to highest associations with *SpovYABBY1* and *SpovYABBY5*.

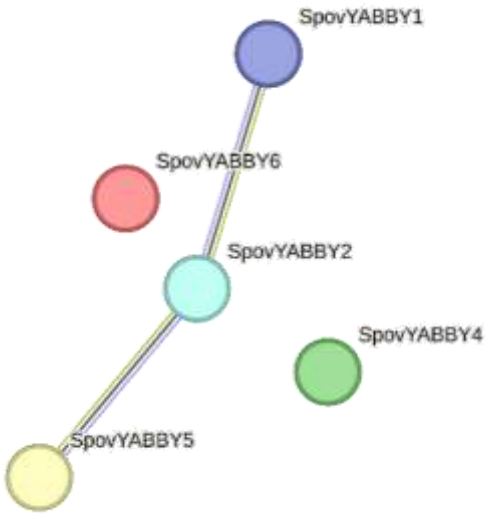

Fig. 9. The *SpovYABBAY2* gene has associations to *SpovYABBAY5* and Spov*YABBAY5* only. These proteins may take part in a specific biological process or have a coordinated functional relationship.

Transcriptomic Data

During a transcriptomic analysis of spinach, variations in oxalate content were observed to be associated with the differential expression of specific genes involved in oxalate content regulation. *SpovYABBY*1 exhibited relatively higher expression levels in the PI175311 cultivar, notably in the leaf (XS1L) with a Fragments Per Kilobase Million (FPKM) value of 12.95, as well as in the root of Bloomsdale (XBDL) with an FPKM of 6.65. Interestingly, its expression was

notably lower in the root of PI175311 and the leaf of Bloomsdale cultivars. *SpovYABBY*2 displayed varying expression levels, with a pronounced upregulation in the root of Bloomsdale (XBDR) where it boasted an FPKM of 14.08. However, this gene exhibited lower expression levels in the remaining samples under investigation. *SpovYABBY*3 showcased significantly high expression in the root of PI175311 (XS1R), yet it failed to express in the leaf of PI175311 (XS1L) and the leaf of Bloomsdale (XBDL). This variation in expression highlights the specificity of this gene in different plant parts and cultivars. Conversely, *SpovYABBY*4 registered very low expression levels across all samples, indicating its limited role in oxalate content regulation in spinach. The most striking observation was made with *SpovYABBY*5, which demonstrated high expression levels in both the leaf and root of Bloomsdale (XBDL and XBDR) with FPKM values of 108.40 and 312.49, respectively. This gene also displayed considerable expression in the leaf (XS1L) and root (XS1R) of PI175311, underlining its vital importance in these specific tissues. These results collectively suggest that the expression levels of these genes differ significantly across various cultivars and plant parts, hinting at the distinct roles they may fulfill in regulating oxalate content in spinach (Fig 10).

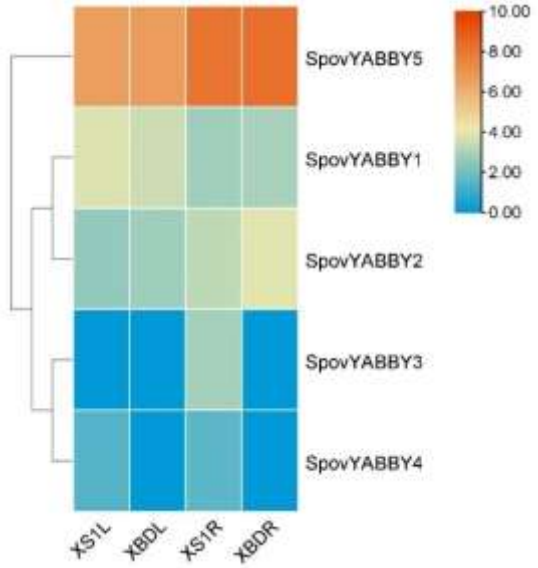

Fig. 10. Illustrate the expression of XS1L (Leaf of PI175311), XBDL (Leaf of Bloomsdale), XS1R (Root of PI175311) and XBDR (Root of Bloomsdale)

DISCUSSION

Plant-specific Transcription Factors (PSTrFs) play a crucial role in guiding various biological and biochemical processes throughout the developmental and growth stages of plants, exerting their influence in a spatiotemporal manner. Among these PSTrFs, *YAB*BY genes in *Spinacia oleracea* and other plant species serve as transcription factors, offering fundamental support throughout the plant's developmental journey (DuPont *et al*., 2003). TFs are essential regulators of molecules and affect gene regulation and networking. If a transcription factor is successfully identified, it improves our understanding of how environmental factors affect plant growth and development (Leghari *et al*., 2016; Kulkarni *et al*., 2020). The *Spinacia oleracea* genome provides insights into traits, aiding trait enhancement through gene identification. Comparative genomics reveals evolutionary patterns, advancing broader plant biology and genetics research.

Phylogenetic analysis of 23 *YABBY* genes revealed their classification into five distinct subfamilies: *YAB*5, *YAB*3/AFO, INO, CRC, and *YAB*2 (**Fig. 3)**. Within each of these subfamilies, *YABBY* proteins are expected to exhibit significant sequence and functional similarities due to their close evolutionary relationships. This categorization provides valuable insights into the potential functional characteristics of these genes and their roles in various plant processes. The quantity of *YABBY* alleles in spinach was less than in all of the crops listed below, including banana (74 Ma*YABBY*), Chinese cabbage (76 BrAT *YABBY*), Rice (30 Os*YABBY*), Arabidopsis (Silva *et al*., 2017; Gupta *et al*., 2015; Ma *et al*., 2015; Yang *et al*., 2006; Hu *et al*., 2015).

Six spinach *YABBY* genes in number, all of which had introns and exons, were discovered by the anticipated exon-intron relationship. *SpovYABBY* genes exhibit varying intron-exon patterns, with a maximum of 6 introns and 7 exons in some, while others have a minimum of 4 introns and 5 exons (Gu *et al*., 2013). The presence of a reduced

and interrelated count of introns and exons in these gene families suggests a trend of purifying selection and evolutionary instability through divergent evolution (Lijavetzky *et al*., 2003). Conversely, a greater number of introns in a plant's genome can be indicative of enhanced evolutionary and genomic stability. The genomic structure and the phylogenetic associations provide a distinct overview of the evolutionary interconnections among different *YABBY* gene families (Bondarenko *et al*., 2016). Identical exon-intron structures have also been found in Arabidopsis, rice, and soybean indicating that these patterns have been retained throughout evolution (Lijavetzky *et al*., 2003; Wang *et al*., 2013).

The presence of the *YABBY* domain in all *SpovYABBY* proteins underscores its conservation, indicating its vital role. The variation in motif length, from 5 to 11, highlights the diversity in these genes. Notably, the conservation of motifs 1 and 2 across these proteins signifies their functional importance. When *YABBY* genes within the same group exhibit similar motif patterns, it suggests unique, shared behaviour within those groups or subgroups. The Spinach *SpovYABBY* genes additionally showed structural conservation in subfamilies and were consistent with other plants like Arabidopsis, Sugar beet, and Peanut (Kong *et al*., 2016; Lu *et al*., 2018; Hamdi *et al*., 2021; Jorrin *et al*., 2007).

The *YABBY* genes' protein lengths vary between 187 and 259 amino acids, with molecular weights ranging from 20.39 to 25.04 kD. Notably, *SpovYABBY* 4 exhibits the shortest protein length, while *SpovYABBY* 5 has the longest. In terms of isoelectric points, the recognized range falls between 4.54 and 9.4**.** These variations in isoelectric points could reflect differences in the charge distribution of the proteins, which may have functional implications. Understanding these protein characteristics, including length, molecular weight, and isoelectric point, is essential for gaining insights into their structural and functional diversity in various biological processes (Weiller *et al*., 2004).

The results of subcellular localization analysis consistently indicated that the majority of these proteins are primarily located in the nucleus and chloroplast. This shared pattern of subcellular distribution suggests that these *YABBY* proteins likely play crucial roles in nuclear functions, such as transcriptional regulation, and in the chloroplast, where they might be involved in photosynthesis-related processes or other chloroplast-specific functions. This consistent localization pattern provides valuable insights into their potential roles within these cellular compartments (Hedden *et al*., 2012).

The analysis of cis-elements in the promoter regions of these specified genes unveils distinct patterns of regulatory elements associated with each one **(Fig. 2)**. *SpovYABBY*2 stands out with a notable abundance of 5 AAGAA-motif elements, which hints at its potential role in gene expression modulation, possibly through interactions with specific transcription factors or RNA-binding proteins (Yin *et al*., 2022). *SpovYABBY*3 highlights a dominant presence of 3 ABRE elements, related to abscisic acid responsiveness, and 3 TCA elements associated with salicylic acid responsiveness, along with 3 MYC elements, emphasizing their potential significance in gene regulation, particularly in response to these signaling molecules. *SpovYABBY*4 features a significant number of 3 STRE elements, suggesting its involvement in stress-responsive gene regulation through interactions with transcription factors. *SpovYABBY*2, *SpovYABBY*3, and *SpovYABBY*4 genes are associated with specific cis-elements that appear to play important roles in responding to various environmental stresses such as heat, cold, and drought. These characteristics indicate that these genes are probably instrumental in controlling how plants react to particular stress conditions, thereby enhancing the plant's capacity to adjust and thrive in challenging environmental situations. The unique patterns of cis-elements within the promoter regions of these genes highlight their specialized roles in regulation and their potential participation in a range of physiological processes or responses to stress in plants (Yamaguchi-Shinozaki and Shinozaki, 1994).

Additionally, *SpovYABBY*2 serves as a key link, forming interactions between *SpovYABBY*1 and *SpovYABBY*5, thereby creating a sub-network within the larger protein-protein interaction network. This phenomenon hints at a potential specific biological process or a coordinated functional relationship among these proteins, emphasizing their importance in a particular context within the network. *SpovYABBY*1 emerges as a central player, receiving numerous miRNA interactions, signifying its intricate involvement in post-transcriptional regulation and possibly influencing a wide array of biological processes. In contrast, the other 5 genes in the family appear to be less affected by miRNA-mediated regulation, suggesting relatively stable and essential functions with limited reliance on such post-transcriptional control Green *et al*., 2016).

The range of Ka/Ks ratios observed in this analysis reveals valuable information about the selective pressures acting on these gene pairs (Liu *et al*., 2022). The minimum Ka/Ks ratio, around 0.121, indicates strong purifying selection, highlighting the conservation of gene function. Conversely, the maximum Ka/Ks ratio, approximately 0.268, suggests that certain gene pairs might be subject to less stringent purifying selection or even experiencing positive selection, indicating evolutionary divergence (Zeeshan Haider *et al*., 2023). In addition to the Ka/Ks ratios, the divergence time (T) for these gene pairs also offers significant insights. Divergence times span from approximately 1.146 million years to 2.553 million years ago, shedding light on historical evolutionary events and

relationships among the *SpovYABBY* genes. These findings collectively contribute to a better understanding of the evolutionary dynamics and functional divergence within this gene family (Kesawat *et al*., 2022). The position of the gene on the chromosome can be used to predict whether a gene will duplicate. This indicates that two or more genes found on the same chromosome may have been created through tandem duplication, whereas genes found on distinct chromosomes may have been created through segmental duplication (Paunchy *et al*., 2016). Despite the relatively dense presence of the Spinach *YABBY* genes on chromosome 5, which is a sign of tandem duplication, synteny analysis also revealed segmental duplication and tandem duplication. In the *YABBY* gene family, segmental duplication was also prevalent in chickpeas and pigeon peas (Nasim et al., 2016; Belic et al., 2021). The primary force behind the expansion of gene families is the phenomena of gene duplication, and the rise in the number of *YABBY* genes in higher plants may be a result of domain duplication during the evolution of eukaryotic plants; (Lau *et al*., 2022).

A significant finding from this research is that the expression levels of the *SpovYABBY* genes vary notably across different spinach cultivars and plant parts. *SpovYABBY*1 and *SpovYABBY*5, in particular, demonstrate distinctive expression patterns. *SpovYABBY*1 is highly expressed in the leaf of the PI175311 cultivar and the root of the Bloomsdale cultivar, indicating its importance in specific plant parts. Conversely, *SpovYABBY*5 exhibits high expression levels in both the leaf and root of Bloomsdale, as well as in the leaf and root of PI175311, suggesting its crucial role in regulating oxalate content. These findings emphasize the tissue- and cultivar-specific roles of these genes in oxalate content regulation in spinach. Downregulating *SpovYABBY*5 and *SpovYABBY*1 in spinach can potentially lead to a decrease in oxalate content. These study findings offer valuable insights into the metabolic pathways and biological processes associated with leaf development and oxalate content regulation in spinach.

CONCLUSION

The investigation into the *YABBY* gene family in *Spinacia olericia* has yielded valuable insights into its structural and functional characteristics. Through an array of computational methods and genomic analyses, six specific *YABBY* genes distributed across the spinach genome were identified and classified into five subgroups according to Arabidopsis criteria. Notably, segmental duplication emerged as the predominant mechanism responsible for gene tandem duplication in spinach, shedding light on the evolutionary history of this gene family. Transcriptomic profiling indicated the high expression of specific *YABBY* genes during leaf development, hinting at their role in oxalate content formation, a trait of significance in spinach. Moreover, the analysis of cis-regulatory elements revealed key elements responsive to various environmental and hormonal cues, providing crucial insights into the regulatory network governing *YABBY* gene expression in spinach. By drawing comparisons between spinach and Arabidopsis *YABBY* genes, common characteristics among subfamilies were identified, enhancing comprehension of the *YABBY* gene family's roles and evolution in *Spinacia olericia*. This study contributes to a deeper understanding of *YABBY* gene functions in spinach, offering a foundation for future research on cloning and functional studies, and facilitating the comparison of *YABBY* genes across different plant species.

REFERENCES

Bailey, T. L., J. Johnson, C. E. Grant and W. S. Noble (2015). The MEME suite. *Nucleic acids research,* 43: W39-W49.

Belić, M. M. (2021). Sinergistički uticaj giberelina i svetlosti na indukciju somatske embriogeneze iz lateralnih korenova spanaća (*Spinacia oleracea* L.) *in vitro*. Ph D Thesis, University of Belgrade Faculty of Biology, Belgrade.

Bondarenko, V. S. and M. S. Gelfand (2016). Evolution of the exon-intron structure in ciliate genomes. *PLoS One,* 11: e0161476.

DuPont, F. M. and S. B. Altenbach (2003). Molecular and biochemical impacts of environmental factors on wheat grain development and protein synthesis. *Journal of cereal science,* 38: 133-146.

Finn, R. D., A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy and A. Heger (2014). Hetherington, K.; Holm, L.; Mistry, J. Pfam: the protein families database. *Nucleic acids research,* 42: D222-D230.

Godini, R. and H. A. Fallahi (2019) brief overview of the concepts, methods and computational tools used in phylogenetic tree construction and gene prediction. *Meta Gene,* 21: 100586.

Green, S. and N. Jones (2016). Constraint-based reasoning for search and explanation: Strategies for understanding variation and patterns in biology. *Dialectica,* 70: 343-374.

Gu, X., Y. Zou, Z. Su, W. Huang, Z. Zhou, Z. Arendsee and Y. Zeng (2013). An update of DIVERGE software for functional divergence analysis of protein family. *Molecular biology and evolution,* 30: 1713-1719.

Gupta, S., N. Malviya, H. Kushwaha, J. Nasim, N. C. Bisht, V. Singh and D. Yadav (2015). Insights into structural and functional diversity of Dof (DNA binding with one finger) transcription factor. *Planta*, 241: 549-562.

Hamdi, J., N. Kmeli and D. Bouktila (2021). Genome-wide survey of sugar beet (Beta vulgaris subsp. vulgaris) Dof transcription factors reveals structural diversity, evolutionary expansion and involvement in taproot development and biotic stress response. *Biologia*, 76: 2421-2436.

Hedden, P. and S. G. Thomas (2012). Gibberellin biosynthesis and its regulation. *Biochemical Journal*, 444: 11-25.

Higo, K., Y. Ugawa, M. Iwamoto and H. Higo (1998). PLACE: a database of plant cis-acting regulatory DNA elements. *Nucleic acids research*, 26: 358-359.

Higo, K., Y. Ugawa, M. Iwamoto and T. Korenaga (1999). Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic acids research*, 27: 297-300.

Horton, P., K. -J. Park, T. Obayashi and K. Nakai (2006). Protein subcellular localization prediction with WoLF PSORT. In: *Proceedings of the 4th Asia-Pacific bioinformatics conference*, 2006; pp. 39-48.

Hou, H., P. Wu, L. Gao, C. Zhang and X. Hou (2019). Characterization and expression profile analysis of *YABBY* family genes in Pak-choi (*Brassica rapa* ssp. chinensis) under abiotic stresses and hormone treatments. *Plant Growth Regulation*, 87: 421-432.

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

Hussain, Q., M. Asim, R. Zhang, R. Khan, S. Farooq and J. Wu (2021). Transcription factors interact with ABA through gene expression and signaling pathways to mitigate drought and salinity stress. *Biomolecules*, 11: 1159.

Jorrín, J. V., A. M. Maldonado and M. A. Castillejo (2007). Plant proteome analysis: a 2006 update. *Proteomics*, 7: 2947-2962.

Kesawat, M.S.;, B. S. Kherawat, A. Singh, P. Dey, S. Routray, C. Mohapatra, D. Saha, C. Ram, K. H. Siddique and A. Kumar (2022). Genome-wide analysis and characterization of the proline-rich extensin-like receptor kinases (PERKs) gene family reveals their role in different developmental stages and stress conditions in wheat (*Triticum aestivum* L.). *Plants*, 11: 496.

Kong, J., Y. Dong, Y. Song, X. Bai, X. Tian, L. Xu, S. Liu and Z. He (2016). Role of exogenous nitric oxide in alleviating iron deficiency stress of peanut seedlings (*Arachis hypogaea* L.). *Journal of plant growth regulation*, 35: 31-43.

Kulkarni, S.R. and K. Vandepoele (2020). Inference of plant gene regulatory networks using data-driven methods: A practical overview. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1863: 194447.

Lau, M. M. L., L. W. K. Lim, H. H. Chung and H. M. Gan (2022). The first draft genome assembly and data analysis of the Malaysian mahseer (Tor tambroides). *Aquaculture and Fisheries*, ???

Leghari, S.J., N. A. Wahocho, G. M. Laghari, A. HafeezLaghari, G. MustafaBhabhan, K. HussainTalpur, T. A. Bhutto, S. A. Wahocho and A. A. Lashari (2016). Role of nitrogen for plant growth and development: A review. *Advances in Environmental Biology*, 10: 209-219.

Lijavetzky, D., P. Carbonero and J. Vicente-Carbajosa (2003). Genome-wide comparative phylogenetic analysis of the rice and Arabidopsis Dof gene families. *BMC evolutionary biology*, 3: 1-11.

Liu, H., H. Ye, J. Wang, S. Chen, M. Li, G. Wang, N. Hou and P. Zhao (2022). Genome-wide identification and characterization of *YABBY* gene family in Juglans regia and Juglans mandshurica. *Agronomy*, 12: 1914.

Lu, Q., X. Niu, M. Zhang, C. Wang, Q. Xu, Y. Feng, Y. Yang, S. Wang, X. Yuan and H. Yu (2028). Genome-wide association study of seed dormancy and the genomic consequences of improvement footprints in rice (*Oryza sativa* L.). *Frontiers in Plant Science*, 2213.

Ma, J., M. -Y. Li, F. Wang, J. Tang and A. -S. Xiong (2015). Genome-wide analysis of Dof family transcription factors and their responses to abiotic stresses in Chinese cabbage. *BMC genomics*, 16: 1-15.

Ma, R., B. Huang, Z. Huang and Z. Zhang (2021). Genome-wide identification and analysis of the *YABBY* gene family in Moso Bamboo (Phyllostachys edulis (Carrière) J. Houz). *PeerJ.*, 9: e11780.

Mazhar, H. S. -U. -D., M. Shafiq, H. Ali, M. Ashfaq, A. Anwar, J. Tabassum, Q. Ali, G. Jilani, M. Awais and R. Sahu (2023). Genome-Wide Identification, and In-Silico Expression Analysis of *YABBY* Gene Family in Response to Biotic and Abiotic Stresses in Potato (*Solanum tuberosum*). *Genes*, 14: 824.

Nasim, J.; N. Malviya, R. Kumar and D. Yadav (2016). Genome-wide bioinformatics analysis of Dof transcription factor gene family of chickpea and its comparative phylogenetic assessment with Arabidopsis and rice. *Plant Systematics and Evolution*, 302: 1009-1026.

Nurani, A. M., Y. Ozawa, T. Furuya, Y. Sakamoto, K. Ebine, S. Matsunaga, T. Ueda, H. Fukuda and Y. Kondo (2020). Deep imaging analysis in VISUAL reveals the role of *YABBY* genes in vascular stem cell fate determination. *Plant and Cell Physiology*, 61: 255-264.

Panchy, N.; M. Lehti-Shiu and Shiu S. -H. Shiu (2016). Evolution of gene duplication in plants. *Plant physiology*, 171: 2294-2316.

Phukela, B., R. Geeta, S. Das and R. Tandon (2020). Ancestral segmental duplication in Solanaceae is responsible for the origin of CRCa–CRCb paralogues in the family. *Molecular Genetics and Genomics*, 295: 563-577.

Ribera, A., R. van Treuren, C. Kik, Y. Bai and A.-M.A. Wolters (2021). On the origin and dispersal of cultivated spinach (Spinacia oleracea L.). *Genetic Resources and Crop Evolution*, 68: 1023-1032.

Ribera, A.; Y. Bai, A. -M. A. Wolters, R. van Treuren and C. Kik (2020). A review on the genetic resources, domestication and breeding history of spinach (*Spinacia oleracea* L.). *Euphytica*, 216: 1-21.

Romanova, M.A., A. I. Maksimova, K. Pawlowski and O. V. Voitsekhovskaja (2021). *YABBY* genes in the development and evolution of land plants. *International Journal of Molecular Sciences*, 22: 4139.

Rombauts, S., P. Déhais, M. Van Montagu and P. Rouzé (1999). PlantCARE, a plant cis-acting regulatory element database. *Nucleic acids research*, 27: 295-296.

Silva, M. D. D. (2017). Identificação e validação de transcritos SuperSAGE de cana-de-açúcar diferencialmente expressos sob déficit hídrico, ????

Sun, L., Y. Q. Wei, K. H. Wu, J. Y. Yan, J. N. Xu, Y. R. Wu, G. X. Li and J. M. Xu (2021). Harberd, N.P.; Ding, Z.J. Restriction of iron loading into developing seeds by a *YABBY* transcription factor safeguards successful reproduction in Arabidopsis. *Molecular Plant*, 14: 1624-1639.

Takahata, S., T. Yago, K. Iwabuchi, H. Hirakawa, Y. Suzuki and Y. Onodera (2016). Comparison of spinach sex chromosomes with sugar beet autosomes reveals extensive synteny and low recombination at the male-determining locus. *Journal of Heredity*, 107: 679-685.

Van den Eede, G., H. Aarts, H. -J. Buhk, G. Corthier, H. J. Flint, W. Hammes, H. Jacobsen, T. Midtvedt, J. Van der Vossen and A. von Wright (2004). The relevance of gene transfer to the safety of food and feed derived from genetically modified (GM) plants. *Food and Chemical Toxicology*, 42, 1127-1156.

Verhoeven, V. J., P. G. Hysi, R. Wojciechowski, Q. Fan, J. A. Guggenheim, R. Höhn, S. MacGregor, A. W. Hewitt, A. Nag and C. -Y. Cheng (2013). Genome-wide meta-analyses of multiancestry cohorts identify multiple new susceptibility loci for refractive error and myopia. *Nature genetics*, 45: 314-318.

Wang, Y., F. M. You, G. R. Lazo, M. -C. Luo, R. Thilmony, S. Gordon, S. F. Kianian and Y. Q. Gu (2013). PIECE: a database for plant gene structure comparison and evolution. *Nucleic acids research*, 41: D1159-D1166.

Weiller, G. F., G. Caraux and N. Sylvester (2004). The modal distribution of protein isoelectric points reflects amino acid properties rather than sequence evolution. *Proteomics*, 4: 943-949.

Xu, C., C. Jiao, H. Sun, X. Cai, X. Wang, C. Ge, Y. Zheng, W. Liu, X. Sun and Y. Xu (2017). Draft genome of spinach and transcriptome diversity of 120 Spinacia accessions. *Nature Communications*, 8: 1-10.

Yamaguchi-Shinozaki, K. and K. Shinozaki (1994). A novel cis-acting element in an Arabidopsis gene is involved in responsiveness to drought, low-temperature, or high-salt stress. *The Plant Cell*, 6: 251-264.

Yang, X., G. A. Tuskan and Z. -M. Cheng (2006). Divergence of the Dof gene families in poplar, Arabidopsis, and rice suggests multiple modes of gene evolution after duplication. *Plant physiology*, 142: 820-830.

Yin, S., S. Li, Y. Gao, E. S. Bartholomew, R. Wang, H. Yang, C. Liu, X. Chen, Y. Wang, and X. Liu (2022). Genome-wide identification of *YABBY* gene family in Cucurbitaceae and expression analysis in Cucumber (*Cucumis sativus* L.). *Genes*, 13: 467.

Zeeshan Haider, M., A. Sami, M. Shafiq, W. Anwar, S. Ali, Q. Ali, M. A. Shahid, D. Ali and S. Alarifi (2023). Genome-Wide Identification and In-Silico Expression Analysis of Carotenoid cleavage oxygenases (CCO) Gene Family CCOCCO Gene Family in Oryza sativaOryza sativa (rice) in response to abiotic stress. *Frontiers in Plant Science*, 14: 1269995.

Zhang, S., L. Wang, X. Sun, Y. Li, J. Yao, S. V. Nocker and X. Wang (2029). Genome-wide analysis of the *YABBY* gene family in grapevine and functional characterization of Vv*YABBY*4. *Frontiers in Plant Science*, 10: 1207.

Zhang, T., C. Li, D. Li, Y. Liu and X. Yang (2020). Roles of *YABBY* transcription factors in the modulation of morphogenesis, development, and phytohormone and stress responses in plants. *Journal of Plant Research*. 133: 751-763.

(*Accepted for publication January 2025*)