HAYATI
Journal of Biosciences

Vol. 32 No. 5, September 2025 1135-1146
DOI:10.4308/hjb.32.5.1135-1146
ISSN: 1978-3019
EISSN: 2086-4094

Research Article

Check for updates

OPEN ACCESS

Genome-wide Analysis of CONSTANS-like (*CqCOL*) Transcription Factors in Quinoa (*Chenopodium quinoa*): Structural Diversity, Phylogeny, and Stress-Responsive Expression

Huyen Thi Thanh Tran[1], Gioi Huy Dong[2], Ha Duc Chu[3], Phi Bang Cao[4*]

[1]*Faculty of Biology, Hanoi National University of Education, Hanoi City 122300, Vietnam*
[2]*Faculty of Biotechnology, Vietnam National University of Agriculture, Hanoi City 122300, Vietnam*
[3]*Faculty of Agricultural Technology, University of Engineering and Technology, Vietnam National University Hanoi, Xuan Thuy Road, Cau Giay District, Hanoi City 122300, Vietnam*
[4]*Faculty of Natural Sciences, Hung Vuong University, Phu Tho Province 35000, Vietnam*

ARTICLE INFO

Article history:
Received October 29, 2024
Received in revised form December 20, 2024
Accepted January 20, 2025

KEYWORDS:
CONSTANS,
transcription factor,
quinoa,
expression pattern,
cis-regulatory element

ABSTRACT

Quinoa (*Chenopodium quinoa*) is an ancient grain renowned for its remarkable nutritional value and remarkable adaptability to diverse environmental conditions, making it a valuable crop for enhancing food security. Understanding the molecular mechanisms triggering its development and stress responses is crucial for crop improvement. This study conducted a comprehensive analysis of the CONSTANS-like (*CqCOL*) transcription factors in quinoa, which play a pivotal role in photoperiodic flowering regulation. We identified and characterized 20 *CqCOL* genes, analyzing their physicochemical properties, phylogenetic relationships, gene structures, and promoter regions. Our findings revealed significant diversity among the *CqCOL* proteins and suggested potential functional specialization within the family. Promoter analysis uncovered various stress-responsive and phytohormone-responsive *cis*-regulatory elements, revealing that *CqCOL* genes may be associated with stress adaptation and hormonal signaling pathways. Transcriptomic analyses under different conditions supported these insights, highlighting the importance of *CqCOL* genes in quinoa's developmental processes and stress responses. Specifically, most *CqCOL* genes exhibited stable expression under heat stress, except *CqCOL02* and *CqCOL12*, which were induced in roots by 1.85- and 1.91-fold, respectively. Under normal conditions, *CqCOL01*, *CqCOL11*, and *CqCOL18* showed organ-specific expression, particularly in flowers and leaves, with no expression detected in roots. This study enhances our understanding of the CqCOL transcription factor family. It provides a foundation for future functional studies and breeding strategies aimed at improving stress tolerance and optimizing flowering time in quinoa.

1. Introduction

Quinoa (*Chenopodium quinoa*) is an important pseudocereal that originates from the Andean region of South America, where it has been cultivated for over 5,000 years (Angeli & Silva 2020). Indigenous populations in South America have long relied on quinoa as a staple food due to its remarkable adaptability to harsh environmental conditions (Bazile *et al*. 2016; Hinojosa *et al*. 2018). In recent decades, quinoa

* Corresponding Author
E-mail Address: phibang.cao@hvu.edu.vn

has garnered global attention due to its exceptional nutritional profile, which includes a complete amino acid composition, high protein content, essential vitamins, minerals, and dietary fiber (Turcios *et al*. 2016; Pereira *et al*. 2019; Pathan & Siddiqui 2022). Its gluten-free nature further enhances its appeal, making it a suitable option for those with gluten intolerance or celiac disease. The crop's resilience and nutritional richness position it as a valuable asset in addressing food security challenges, especially in the context of a growing global population and climate change (Guo *et al*. 2022; Nguyen *et al*. 2022). Recent studies demonstrated that regulating flowering time in quinoa plants is essential for optimizing reproductive success and maximizing crop yields (Patiranage *et al*. 2021; Oustani *et al*. 2023). This regulation relies on the plant's physiological response to the duration of light and darkness, synchronizing flowering with favorable environmental conditions. CONSTANS (*CO*) transcription factors (TFs) play a central role in photoperiodic flowering regulation across various plant species, including quinoa. While no studies have specifically explored the *CO* TF family in quinoa, its interaction with genes such as FLOWERING LOCUS T (FT) and other flowering-related regulators in other species suggests a broader network of genetic control affecting reproductive development. Understanding how these genes interact to modulate reproductive success under diverse environmental conditions remains a critical area for further research.

Of interest to us is the *CO* TF, which is a critical regulator of photoperiodic flowering in numerous plant species (Suarez-Lopez *et al*. 2001; Kim *et al*. 2008). Structurally, it features two specific B-box zinc finger domains at the N-terminal, which mediate protein interactions and DNA binding, and a CCT domain at the C-terminal, essential for nuclear localization and interactions with regulatory proteins (Robson *et al*. 2001; Wenkel *et al*. 2006). Functionally, *CO* functions as a transcriptional activator, linking environmental light cues and the circadian clock to regulate key flowering genes, particularly FT (Suarez-Lopez *et al*. 2001; Cai *et al*. 2007). Under favorable photoperiodic conditions, *CO* accumulates in the nucleus, binds target gene promoters, and activates FT, initiating the transition from vegetative growth to flowering. *CO* activity is finely regulated through transcriptional control, post-translational modifications, and interactions with light signaling proteins. Understanding *CO*'s structure and function is essential for elucidating photoperiodic flowering mechanisms and holds potential for optimizing flowering time to improve crop yields. Recently, the *CO* TF families have been investigated in a large number of higher plant species, including barley (Griffiths *et al*. 2003), rice (Griffiths *et al*. 2003), *Arabidopsis* (Griffiths *et al*. 2003), mango (Liu *et al*. 2022), Chinese white pear (Wang *et al*. 2017), grapevine (Wang *et al*. 2019), sugar beet (Chia *et al*. 2008), chrysanth (Fu *et al*. 2015), and gingko (Yan *et al*. 2017). A previous study provided insights into the genetic diversity of flowering time regulators in quinoa, specifically *FT* and *CO* genes, highlighting the association of sequence variations and haplotypes with flowering responses under different photoperiods and their correlation with geographical distribution (Patiranage *et al*. 2021). However, no comprehensive study has concerned the *CO* TF family in quinoa.

This current study aimed to perform a comprehensive identification, characterization, and expression analysis of the *CO* TF family in quinoa based on computational approaches. By examining their gene structures, evolutionary relationships, physicochemical properties, *cis*-regulatory elements, and expression patterns across diverse tissues and environmental conditions, the current research seeks to explain the functional roles of *CO* genes in photoperiodic flowering and stress responses. The anticipated findings are expected to deepen our understanding of the *CO* TF family within the quinoa and contribute to future strategies for crop improvement.

2. Materials and Methods

2.1. Search of *CO* Transcription Factors in Quinoa

To seek the *CO* TFs within the quinoa genome (Jarvis *et al*. 2017), we utilized a homology-based approach anchored on well-characterized *CO* genes from *Arabidopsis* (Griffiths *et al*. 2003) as recently reported (Chu *et al*. 2018; Niu *et al*. 2020). First, *CO* protein sequences were obtained from a recent study (Griffiths *et al*. 2003) to serve as reference queries. These sequences were employed in BLAST searches against the quinoa assembly (Jarvis *et al*. 2017) to detect homologous proteins. The resulting candidate proteins were then analyzed using HMMER software (Potter *et al*. 2018), employing conserved *CO* domain profiles (Robson *et al*. 2001) sourced from the Pfam database (Mistry *et*

al. 2021) to verify the presence of characteristic *CO* domains. An E-value threshold of < 1E-10 was applied to ensure domain detection's reliability. Proteins that exhibited the conserved *CO* domains were designated as putative *CO* TFs in quinoa and were selected for subsequent characterization and analysis.

2.2. Analysis of Properties of *CO* Transcription Factors in Quinoa

To analyze the physicochemical properties of the *CO* TFs in quinoa, the ProtParam tool available on the ExPASy website (Gasteiger *et al*. 2003) was applied as previously described (Niu *et al*. 2020). Initially, we obtained the full-length amino acid (aa) sequences of all *CO* proteins in quinoa (Jarvis *et al*. 2017). Each sequence was individually input into ProtParam, which computes various physicochemical parameters based solely on the protein's amino acid composition. The parameters assessed included molecular weight (mW), aliphatic index (AI), theoretical isoelectric point (pI), and grand average of hydropathicity (GRAVY).

2.3. Generation of The Phylogenetic Tree of *CO* Transcription Factors in Quinoa

To elucidate the relationships among the *CO* TFs in quinoa, we generated a phylogenetic tree by using the MEGA tool (Kumar *et al*. 2018) as recently guided (La *et al*. 2022; Chu *et al*. 2024). Initially, the full-length aa sequences of all identified *CO* proteins were retrieved from the quinoa proteome database. Well-characterized *CO* TF proteins from *Arabidopsis* and sugar beet were also obtained from previous studies (Griffiths *et al*. 2003; Chia *et al*. 2008) to serve as reference sequences for analysis. Multiple sequence alignments were carried out by using the ClustalW algorithm integrated into MEGA software. Following alignment, an unrooted phylogenetic tree was generated by employing the Maximum Likelihood method. To assess the statistical robustness of the inferred phylogenetic relationships, a bootstrap analysis with 1,000 replicates was applied.

2.4. Analysis of Gene Structure of *CO* in Quinoa

We utilized the GSDS tool (Hu *et al*. 2015) to conduct the structural analysis of *CO* genes in quinoa, as previously reported (Niu *et al*. 2020). Initially, the genomic and corresponding coding sequences of each *CO* gene were retrieved from the quinoa assembly. These sequences were then uploaded to the GSDS platform (Hu *et al*. 2015), which allows for the visualization and analysis of gene structures. The tool aligns the genomic DNA with the coding DNA sequence to generate a schematic representation of exons, introns, and untranslated regions.

2.5. Promoter Analysis of *CO* Genes in Quinoa

To analyze the *cis*-regulatory elements (CREs) in the promoter sequences of the *CO* genes in quinoa, we first collected the 2 kb upstream of the transcription start site of the *CO* gene. These sequences, representing the promoter regions, were extracted from the quinoa genome database (Jarvis *et al*. 2017) and are accessible in the Phytozome portal (Goodstein *et al*. 2012). Once the upstream sequences were obtained, the PlantCARE web-based tool (Lescot *et al*. 2002) was employed to predict and identify potential CREs within these regions.

2.6. Transcriptome Analysis of *CO* Transcription Factors in Quinoa

To access the expression profiles of the *CO* genes in quinoa, we utilized three publicly available microarray datasets from the NCBI GEO (Barrett *et al*. 2013), including GSE128155 (Tovar & Quillatupa 2020), GSE139174 (Liu *et al*. 2021), and GSE156523. The raw sequencing data were downloaded and subjected to quality control using FastQC to assess sequencing quality and identify technical issues. Gene expression levels were quantified by using the fold-change and FPKM values. Visualization tools such as heatmaps and expression profile plots were generated using packages like ggplot2 in R.

3. Results

3.1. Comprehensive Identification of The *CO* TF Family in Quinoa

To identify *CO* TFs within the quinoa genome, we employed a homology-based approach anchored on well-characterized *CO* genes from *Arabidopsis*. After screening by using the Pfam database, a total of 20 members of the *CO* TF family in quinoa have been reported. The locus name of each member of the *CO* TF family in quinoa has been provided in Table 1. At the same time, their coding DNA sequences, genomic DNA sequences, and full-length protein sequences were obtained for further analyses. To systematically represent the *CO* TF family in quinoa, we have designated these genes as "*CqCOL*." The prefix "Cq" denotes *Chenopodium quinoa*, directly

Table 1. Information on the *CO* TF family in quinoa

Gene name	Locus name	Length (aa)	mW (KDa)	pI	Gravy	AI
CqCOL01	AUR62035217	305	32.81	5.59	-0.42	65.80
CqCOL02	AUR62035221	442	50.44	5.66	-0.77	70.18
CqCOL03	AUR62041089	332	37.23	4.69	-0.73	61.39
CqCOL04	AUR62030805	467	52.80	5.07	-0.93	56.19
CqCOL05	AUR62040293	389	42.94	5.74	-0.62	62.96
CqCOL06	AUR62037849	348	39.34	4.61	-0.78	61.64
CqCOL07	AUR62009440	367	40.73	5.31	-0.63	56.13
CqCOL08	AUR62023118	367	40.88	5.24	-0.69	54.01
CqCOL09	AUR62037140	411	44.67	5.24	-0.52	63.87
CqCOL10	AUR62034638	398	44.63	5.91	-0.73	61.48
CqCOL11	AUR62020307	305	32.84	5.59	-0.43	66.13
CqCOL12	AUR62002867	403	46.00	5.82	-0.83	64.37
CqCOL13	AUR62028867	565	62.97	5.06	-0.82	61.47
CqCOL14	AUR62025058	411	44.77	5.11	-0.55	63.14
CqCOL15	AUR62039984	389	42.69	5.38	-0.54	66.22
CqCOL16	AUR62001856	253	28.45	6.29	-0.80	61.46
CqCOL17	AUR62008578	398	44.53	5.87	-0.68	64.17
CqCOL18	AUR62030486	430	48.66	5.32	-0.95	54.42
CqCOL19	AUR62005692	575	64.00	5.53	-0.84	61.91
CqCOL20	AUR62009650	253	28.29	6.59	-0.77	62.85

linking the gene nomenclature to the studied species. The suffix "*COL*" stands for "*CONSTANS-LIKE*", a standard convention used to identify genes homologous to the original *CONSTANS* gene characterized in *Arabidopsis*.

3.2. Estimation of The General Parameters of The *CO* TF Family in Quinoa

In this study, we assessed the parameters of the *CO* TFs in quinoa, including mW, pI, AI, and GRAVY. By evaluating these properties, a comprehensive profile of the structural and stability characteristics of the *CqCOL* proteins was established, providing essential information for further biochemical and structural studies. As a result, the general characteristics of the *CqCOL* proteins in quinoa are provided in Table 1.

The physicochemical features of the *CqCOL* TFs in quinoa were comprehensively analyzed using ProtParam, revealing significant diversity among the family members. The mW of the *CqCOL* proteins ranged from approximately 28.29 (*CqCOL20*) to 64.00 kilodaltons (*CqCOL19*), indicating variability in their amino acid lengths and compositions. The theoretical pI varied widely, spanning from pH 4.61 (*CqCOL06*) to pH 6.59 (*CqCOL20*), which suggests differences in their net electrical charges under physiological conditions. The AI values differed among the *CqCOL* proteins, between 54.01 (*CqCOL08*) and 70.18 (*CqCOL02*), providing insights into their relative thermostability, with higher values suggesting greater stability at elevated temperatures. The GRAVY values of all *CqCOL* proteins were negative, ranging from -0.95 (*CqCOL18*) to -0.42 (*CqCOL01*). These physicochemical parameters establish a foundational profile of the *CqCOL* protein family in quinoa, facilitating further biochemical and structural investigations.

3.3. Phylogenetic Categorization of The *CO* TF Family in Quinoa

To investigate the evolutionary relationships among *CO* TFs in quinoa, we constructed an unrooted phylogenetic tree. This analysis included 20 *CO* protein sequences from quinoa, along with well-characterized *CO* proteins from *Arabidopsis* and sugar beet serving as reference sequences. Utilizing the Maximum Likelihood method, we generated a phylogenetic tree presented in Figure 1. This figure illustrated the relationships among all members of the *CO* TF families in the quinoa and related species, providing valuable insights into their evolutionary connections. Based on the phylogenetic analysis of the *CqCOL* TFs in quinoa, these members can be classified into three distinct groups. The first group, group I, consists of 6 (out of 20) *CqCOL* proteins. Group II includes 6 (out of 20) *CqCOL* proteins, while group III contains 8 remaining *CqCOL* proteins.

3.4. Investigation of The Gene Organization of The *CO* TF Family in Quinoa

This study analyzed the gene structure of *CqCOL* genes in quinoa to investigate exon-intron arrangements and structural diversity within the gene family. Using the GSDS tool, full-length genomic sequences were

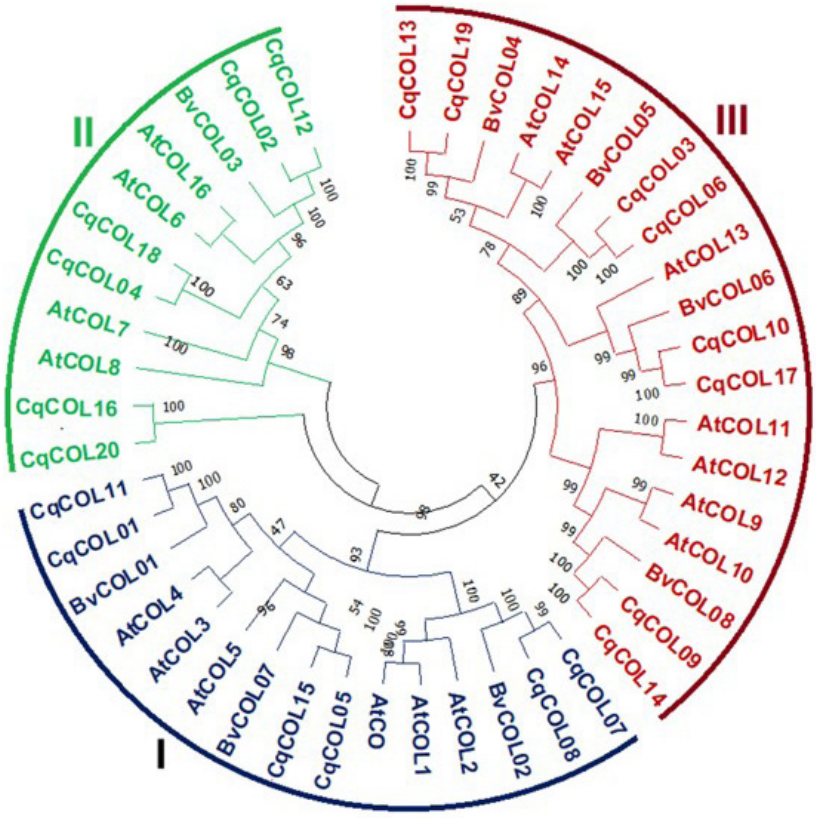

Figure 1. Phylogenetic categorization of the *CO* TF families in quinoa, arabidopsis, and sugar beet

aligned with their corresponding coding sequences to identify exon/intron structure. The analysis revealed notable variation in the exon-intron structures among the 20 *CqCOL* genes (Figure 2).

The results indicated that the number of exons varied from 2 to 4, reflecting structural diversity in the gene family. Specifically, 8 out of the 20 *CqCOL* genes exhibited a simple structure with only 2 exons separated by a single intron. Another 8 genes displayed more complex architectures, containing 4 exons, while the remaining 4 genes had 3 exons. These variations in exon-intron organization suggest evolutionary diversification within the *CqCOL* gene family, potentially influencing the functional roles of individual genes in photoperiodic flowering regulation and other developmental processes.

3.5. Analysis of Stress-Responsive and Phytohormone-Responsive CREs in The Promoter Sequences of The *CO* TF Family in Quinoa

To analyze CREs in the promoter sequences of *CqCOL* genes in quinoa, we extracted the 2 kb upstream of the transcription start site for each *CO* gene. Using the PlantCARE tool, several putative CREs related to phytohormone and stress responsiveness were explored. As a result, Tables 2 and 3 described the foundation of the stress-responsive and phytohormone-inducible CREs of all members of the *CqCOL* genes in quinoa, respectively.

The analysis of the promoter regions of *CqCOL* genes in quinoa revealed the presence of 4 key stress-responsive CREs. These include the LTRE (CRE associated with low-temperature responsiveness), MYBRS (MYB recognize site involved in drought responsiveness), MBS (MYB binding site involved in drought-inducibility), and TC-rich repeats (element linked to defense and stress responsiveness). We realized that two drought-responsive CREs were localized in 8 *CO* genes, including *CqCOL03, CqCOL06, CqCOL07, CqCOL12, CqCOL15, CqCOL16, CqCOL17,* and *CqCOL18*. The presence of LTRE was also recorded in the promoter sequences of 4 *CO* genes, including *CqCOL02, CqCOL06, CqCOL07,* and *CqCOL08*, while TC-rich repeats have been found in the promoter sequences of 8 *CO* genes, such as *CqCOL04, CqCOL07, CqCOL08, CqCOL10, CqCOL15, CqCOL16,* and *CqCOL19*. Our predictions suggested that these *CO* genes may be putative in stress responsiveness.

Next, we investigated the occurrences of phytohormone-responsive CREs in the promoter sequences of the *CqCOL* genes. Briefly, a total of 10 CREs related to

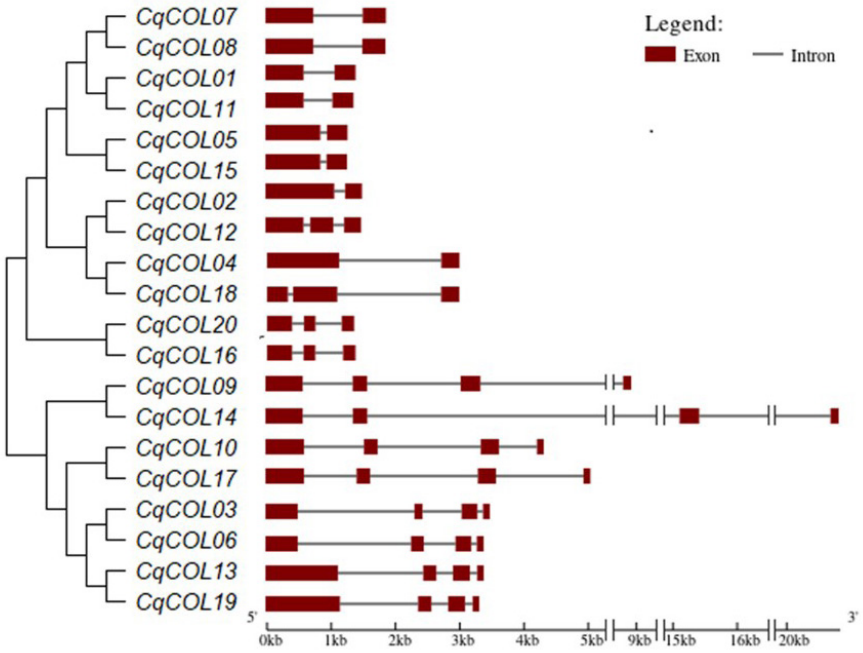

Figure 2. Structural analysis of genes encoding the *CO* TF family in quinoa

Table 2. A list of stress-responsive *cis*-regulatory elements in the promoter regions of genes encoding the *CO* TF family in quinoa

	Stress-related CREs			
Gene name	MYBRS	LTRE	MBS	TC-rich repeats
CqCOL01				
CqCOL02		Y		
CqCOL03	Y			
CqCOL04				Y
CqCOL05				
CqCOL06	Y	Y		
CqCOL07		Y		Y
CqCOL08		Y	Y	Y
CqCOL09				
CqCOL10				Y
CqCOL11				
CqCOL12	Y			
CqCOL13				
CqCOL14				
CqCOL15	Y			Y
CqCOL16	Y			Y
CqCOL17	Y			
CqCOL18	Y			
CqCOL19			Y	Y
CqCOL20				

MYBRS: MYB recognize site involved in drought responsiveness, LTRE: CRE associated with low-temperature responsiveness, MBS: MYB binding site involved in drought-inducibility, TC-rich repeats: element linked to defense and stress responsiveness

phytohormone responsiveness, including ABRE (abscisic acid-responsive element), TATC-box (CRE related to gibberellin-responsiveness), TGACG-motif (CRE involved in the jasmonic acid-responsiveness), ERE (ethylene-responsive element), TCA-element (CRE involved in salicylic acid responsiveness), GARE-motif (Gibberellin-responsive element), P-box (Gibberellin responsive element), TGA-element (auxin-responsive element), TGACG-motif (CRE related to the jasmonic acid-responsiveness), CGTCA-motif (CRE related to the jasmonic acid-responsiveness), has been predicted. As expected, a foundation of the hormone-responsive elements was enriched in the promoter sequences of the *CO* genes in quinoa. Among them, a large number (11 out of 20) *CO* genes contained ABRE, while 12 (out of 20) *CO* genes had 3 CREs related to gibberellin-responsiveness (GARE-motif, TATC-box, and P-box). Next, 3 CREs involved in jasmonic acid-responsiveness, including TGACG-motif, CGTCA-motif, and TGACG-motif, were identified in the promoter sequences of 11 *CO* genes. Additionally, 1, 4, and 7 *CO* genes contained CREs related to ethylene, auxin, and salicylic acid-responsiveness, respectively. This finding provided important insights into the potential regulatory pathways involved in the stress adaptation of quinoa.

Table 3. A list of phytohormone-induced *cis*-regulatory elements in the promoter regions of genes encoding the *CO* TF family in quinoa

Gene name	ABRE	TATC	TGACG	ERE	GARE	P	TGA	TCA	TGACG	CGTCA
CqCOL01	Y			Y			Y			
CqCOL02						Y	Y			Y
CqCOL03	Y									
CqCOL04			Y		Y					
CqCOL05	Y				Y	Y				Y
CqCOL06		Y	Y							Y
CqCOL07	Y					Y		Y		
CqCOL08	Y					Y				
CqCOL09	Y		Y		Y	Y		Y		Y
CqCOL10	Y	Y								
CqCOL11										
CqCOL12	Y		Y				Y			
CqCOL13							Y	Y		
CqCOL14	Y				Y	Y		Y	Y	
CqCOL15	Y					Y		Y		Y
CqCOL16								Y		Y
CqCOL17	Y	Y	Y							
CqCOL18					Y					
CqCOL19			Y							
CqCOL20								Y		

ABRE: abscisic acid-responsive element, TATC: TATC-box (CRE involved in gibberellin-responsiveness), TGACG: TGACG-motif (CRE involved in the jasmonic acid-responsiveness), ERE: ethylene-responsive element, GARE: GARE-motif (Gibberellin-responsive element), P: P-box (Gibberellin responsive element), TGA: TGA-element (auxin-responsive element), TCA-element (CRE involved in salicylic acid responsiveness), TGACG: TGACG-motif (CRE involved in the jasmonic acid-responsiveness), CGTCA: CGTCA-motif (CRE involved in the jasmonic acid-responsiveness)

3.6. Transcriptomic Analysis of The *CO* Genes in Quinoa Under Adverse Environmental Conditions

To access the expression profiles of the *CO* genes in quinoa, we performed a comprehensive transcriptomic analysis under various treatments by exploring recent RNA-Seq datasets. As a result, Figures 3 and 4 were constructed to describe the expression levels of the *CqCOL* genes under high-temperature stress and normal conditions, respectively.

Firstly, most (18 out of 20) *CqCOL* genes exhibited non-different expression (|fold-change| < 1.5) in heated roots and/or shoots. Meanwhile, only two *CqCOL* genes, *CqCOL02* and *CqCOL12*, were up-regulated in treated roots by approximately 1.85 and 1.91-fold, respectively. Interestingly, the promoter regions of *CqCOL02* and *CqCOL12* contained CREs involved in the gibberellin, auxin, jasmonic acid, and abscisic acid responsiveness. These findings suggested that these *CO* genes may act as positive regulators in heat stress response via the crosstalk between gibberellin, auxin, jasmonic acid, and abscisic acid.

Additionally, the *CO* genes exhibited variable expression profiles in major organs in quinoa plants under normal conditions. Among them, the expression levels of 3 *CO* genes, *CqCOL01, CqCOL11,* and *CqCOL18*, were high in flower tissues of both white quinoa and/or yellow quinoa plants (Figure 4A). Under normal conditions, it has been realized that 2 *CO* genes, particularly *CqCOL01* and *CqCOL11*, were specific in stem and leaf tissues, while *CqCOL18* was exclusively expressed in leaf tissues of yellow quinoa plants (Figure 4B). We also found that no *CqCOL* genes were expressed in examined root samples. Our findings strongly suggested a hypothesis of the potential function of expressed *CqCOL* genes in specific organs related to regulating photoperiodic flowering and other developmental processes in quinoa plants.

4. Discussion

4.1. Phylogenetic and Structural Diversity of *CO* Transcription Factors Across Plant Species

A comparative analysis of the *CO* TF families across various plant species reveals notable differences in the number of *CO* genes, highlighting the evolutionary complexity and functional multiplicity within this TF family. We identified twenty *CO* family genes in quinoa, which is relatively higher than several other species. For example, in the model plant *Arabidopsis*, 16 *CO* TFs have been found and thoroughly annotated (Griffiths *et al*. 2003). Rice also possesses 16 *CO* genes, as previously reported (Griffiths *et al*. 2003). In the Chinese white pear, 15 *CO* genes were detected

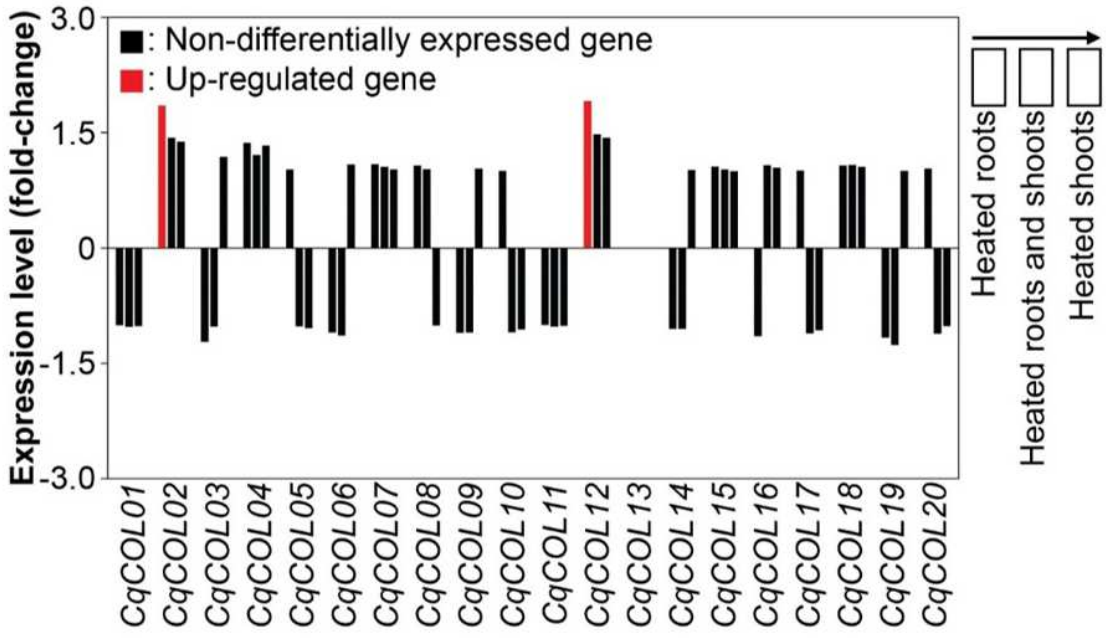

Figure 3. Expression levels of genes encoding the *CO* TF family in quinoa under heat stress condition

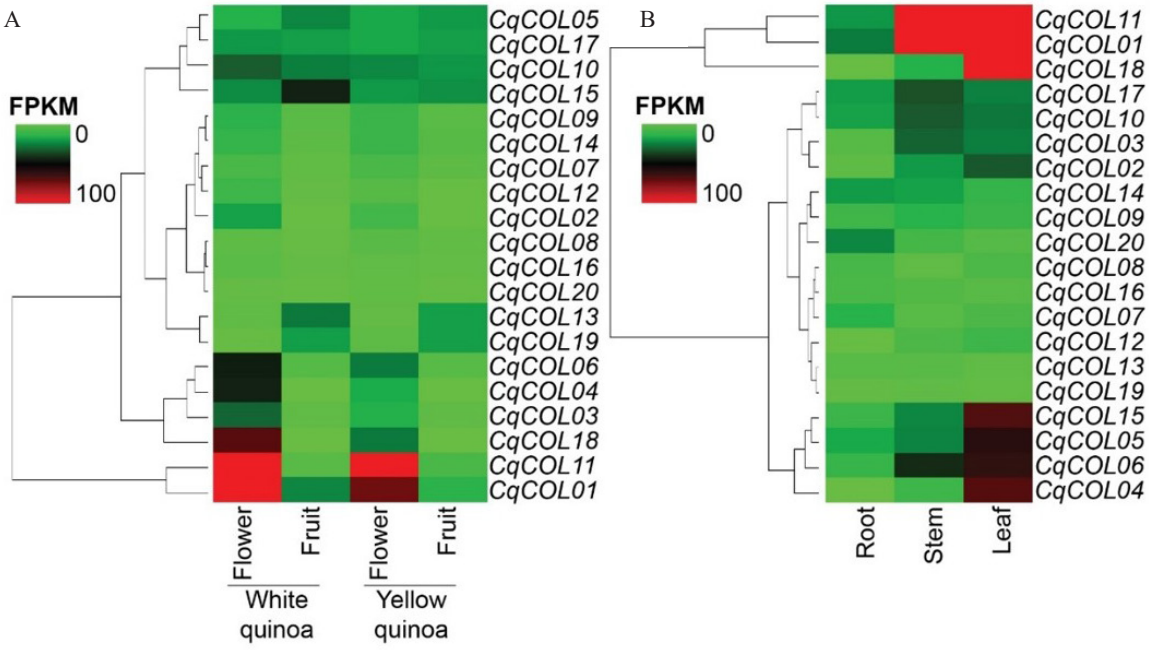

Figure 4. Expression levels of genes encoding the *CO* TF family in (A) flower and fruit tissues in white quinoa and yellow quinoa plants and (B) root, stem, and leaf in yellow quinoa plants

in the latest assembly (Wang *et al*. 2017). Conversely, chrysanth has at least 7 *CO* genes reported in its genome (Fu *et al*. 2015). Additionally, 19 members of the *CO* family have been identified in ginkgo (Yan *et al*. 2017), and mango has been found to contain 21 such genes (Liu *et al*. 2022). This variation in *CO* gene numbers among different species underscores the intricate evolutionary dynamics. It suggested that gene expansion or contraction may be crucial in regulating flowering time and facilitating adaptation to diverse environmental conditions.

Previous studies have extensively investigated the characteristics of *CO* TF families in various higher plant species, revealing significant diversity in their genetic sequences and physicochemical properties. For instance, in mango, the *CO* TF family has been reported to possess gene sequences ranging from 492 to 1,536 base pairs in length (Liu *et al*. 2022). These sequences encode proteins consisting of 163 to 511 aa residues, resulting in mW values between 18.11 and 56.63 kilodaltons and predicted pI spanning from 3.93 to 8.85 (Liu *et al*. 2022). This wide range indicates considerable protein size variation and net charges under physiological pH conditions (Liu *et al*. 2022). Similarly, in grapevine, the predicted open reading frames of *CO* genes exhibit substantial variation, ranging from approximately 1,044 base pairs for *VviCOL4* to about 1,425 base pairs for *VviCOL14b* (Wang *et al*. 2019). The corresponding proteins vary in length from 347 to 474 aa residues, leading to calculated molecular masses between roughly 38.00 and 51.43 kilodaltons (Wang *et al*. 2019). In the case of Chrysanthemum, the sizes of *CO* TF proteins have been reported to range from 337 to 434 aa residues (Fu *et al*. 2015). Additionally, in Chinese white pear, the *CO* TF family includes proteins varying between 340 and 488 aa residues in length (Wang *et al*. 2017). These observations underscore the complexity and evolutionary dynamics of the *CO* TF family across different plant species. The physicochemical properties, such as mW, pI, AI, and GRAVY values, provide valuable insights into these proteins' structural stability and functional potential. By comprehensively analyzing these physicochemical parameters, we establish a crucial groundwork for the *CO* TF family's future structural and functional studies. This detailed characterization is instrumental in enhancing our understanding of the molecular mechanisms by which *CO* proteins regulate flowering time in response to photoperiodic signals. We gain deeper insights into the complex regulatory networks governing plant growth and reproduction by elucidating how these proteins interact with environmental cues to influence developmental processes. This advanced knowledge is vital for progress in plant developmental biology. Moreover, it has significant practical implications, as it can guide breeding programs to enhance crop adaptation to various environmental conditions. By integrating this molecular-level understanding into agricultural practices, we can construct crop lines with improved resilience and productivity, contributing to food security and sustainable agriculture.

Previous research has documented the classification of *CO* TF families in various higher plant species. In grapevine, 20 *CO* genes were found and categorized into three groups via phylogenetic analysis (Wang *et al*. 2019). Specifically, three genes, *VviCOL2, VviCOL4,* and *VviCOL5*, were assigned to group I; *VviCOL16a* and *VviCOL16b* were placed in group II; and the remaining CO genes, including *VviCOL9a, VviCOL9b, VviCOL11a, VviCOL11b, VviCOL13, VviCOL14a,* and *VviCOL14b*, were grouped into group III (Wang *et al*. 2019). Furthermore, to explore the relationships among *CO* members, a phylogenetic tree was constructed using 60 *CO* members from *A. thaliana*, grapevine, and mango (Liu *et al*. 2022). This analysis revealed that these *CO* proteins clustered into three separate clades, corresponding closely with their structural differences (Liu *et al*. 2022). Similar classification models were reported in *CO* TF families of other crop species (Griffiths *et al*. 2003; Fu *et al*. 2015;Wang *et al*. 2017; Yan *et al*. 2017). Taken together, the classification of *CqCOL* TFs in quinoa, based on phylogenetic analysis, reveals evolutionary divergence and functional diversity similar to that observed in other plant species (Griffiths *et al*. 2003; Fu *et al*. 2015; Wang *et al*. 2017; Yan *et al*. 2017; Wang *et al*. 2019; Liu *et al*. 2022). Grouping the *CqCOL* proteins into three distinct clades emphasizes the conservation and diversification within this gene family. This categorization highlights the critical function of *CO* TFs across various plant species.

Comparative analyses highlight similar structural variability in *CO* TF families across several crop species, including mango (Liu *et al*. 2022) and grapevine (Wang *et al*. 2019). For instance, in mango, the analysis of 30 *CO* genes revealed a range of intron numbers, from intronless genes (*MiCOL11*) to genes with up to 4 introns. Notably, 10 genes had only 1 intron, while 8 and 9 genes possessed 3 and 2 introns, respectively (Liu *et al*. 2022). This diversity in exon-

intron organization suggests functional differentiation within the gene family, even within a single species.

In grapevine, a similar phenomenon of structural variability was recorded. For example, *CO* genes like *VviCOL9a, VviCOL11a,* and *VviCOL14a* each contained 3 introns, whereas others, such as *VviCOL2, VviCOL4,* and *VviCOL16a*, had only 1 intron (Wang *et al*. 2019). These observations underscore the evolutionary dynamics of *CO* genes across plant species and suggest that differences in gene structure may have functional implications for their roles in developmental processes, including flowering regulation. This comparative insight enhances our investigation of the *CqCOL* gene family's diversity and its potential adaptive significance.

4.1. Regulatory Mechanisms of *CO* Genes in Quinoa: Insights from *cis*-regulatory Elements and Expression Profiles

The transcriptional regulation of genes is largely influenced by CREs located in their promoter regions, which serve as binding sites for TFs and modulate gene expression in response to environmental and hormonal cues. In this study, a comprehensive analysis of the promoter sequences of *CqCOL* genes in quinoa revealed the presence of multiple CREs associated with stress responses and phytohormone signaling pathways. Among these, stress-responsive elements such as LTRE and MBS are involved in drought-inducibility) and TC-rich repeats (associated with defense and stress responses) were identified in several *CqCOL* genes. Notably, the presence of MBS elements in promoter regions of the *CqCOL* genes suggests their potential role in drought stress adaptation. MYB TFs are known to regulate drought-responsive pathways in plants. Similarly, the presence of LTRE indicated that these *CqCOL* genes may contribute to cold tolerance, aligning with previous studies that have linked LTRE-containing genes to low-temperature resilience.

Additionally, the analysis of hormone-responsive elements revealed an abundance of CREs associated with abscisic acid (ABRE), gibberellins (GARE-motif, TATC-box, and P-box), jasmonic acid (TGACG-motif and CGTCA-motif), and salicylic acid (TCA-element). The enrichment of ABRE motifs in several *CqCOL* genes suggested a regulatory role in abscisic acid-mediated stress responses, particularly in drought and osmotic stress conditions. Furthermore, the presence of gibberellin-responsive elements supported the hypothesis that these *CqCOL* genes may be involved in growth and developmental processes, particularly in the regulation of flowering time. Gibberellins are known to interact with *CO* genes to influence the transition from vegetative to reproductive phases in plants. The identification of multiple hormone-responsive elements in the promoter regions of *CqCOL* genes provides insights into their potential regulatory mechanisms. It suggests that these transcription factors play an integral role in coordinating stress adaptation and developmental processes in quinoa.

To elucidate the functional significance of *CqCOL* genes, we conducted a transcriptomic analysis under both normal and stress conditions. The results indicated that while the majority of *CqCOL* genes exhibited stable expression patterns under heat stress, two genes, *CqCOL02* and *CqCOL12*, were significantly upregulated in root tissues. This suggests that these genes may serve as key regulators in the heat stress response, potentially mediating signaling pathways that enhance root resilience under high-temperature conditions. The presence of gibberellin-, auxin-, and jasmonic acid-responsive elements in the promoter regions of these genes further supports their role in hormonal crosstalk during stress adaptation. These findings align with previous studies demonstrating that *CO* genes can function in stress adaptation by integrating environmental signals with hormonal responses.

Overall, the integration of CRE analysis and expression profiling provides valuable insights into the regulatory mechanisms governing the *CqCOL* gene family in quinoa. The presence of stress- and hormone-responsive elements in their promoter regions, coupled with their dynamic expression under environmental and developmental conditions, underscores the complexity of their functional roles. These findings highlight the importance of *CqCOL* genes in stress adaptation and flowering regulation and provide a molecular framework for future functional studies aimed at improving quinoa's resilience and reproductive success.

In conclusion, To sum up, this study presents a comprehensive analysis of the *CO* TFs in quinoa, identifying 20 members designated as *CqCOL*. Physicochemical characterization revealed significant diversity among these proteins regarding mW, pI, AI, and GRAVY values. Phylogenetic analysis classified the *CqCOL* proteins into three distinct groups, indicating evolutionary divergence and potential functional differentiation. Structural analysis showed diversity in exon-intron organization, with exon

numbers ranging from two to four, suggesting possible influences on gene expression and function. Promoter analysis uncovered several stress-responsive and phytohormone-responsive CREs, implying that certain *CqCOL* genes may participate in stress responses and hormonal signaling pathways. Transcriptomic analyses revealed that while most *CqCOL* genes did not significantly change expression under heat stress, *CqCOL02* and *CqCOL12* were upregulated, suggesting roles in heat stress adaptation. Additionally, tissue-specific expression patterns under normal conditions indicate potential involvement in photoperiodic flowering regulation and developmental processes. These findings enhance our understanding of the *CO* TFs in quinoa and supply an establishment for future functional studies and breeding strategies to improve stress tolerance and optimize flowering time.

Acknowledgements

This work was funded by Hung Vuong University.

References

Angeli, V., Silva, P.M., 2020. Quinoa (*Chenopodium quinoa* Willd.): An overview of the potentials of the "Golden grain" and socio-economic and environmental aspects of its cultivation and marketization. *Foods*. 9, 216. https://doi.org/10.3390/foods9020216

Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., Yefanov, A., Lee H., Zhang, N., Robertson, C.L., Serova, N., Davis, S., Soboleva, A., 2013. NCBI GEO: archive for functional genomics data sets - update. *Nucleic Acids Res*. 41, D991-995. https://doi.org/10.1093/nar/gks1193

Bazile, D., Jacobsen, S.E., Verniau, A., 2016. The global expansion of Quinoa: Trends and limits. *Front Plant Sci*. 7, 622. https://doi.org/10.3389/fpls.2016.00622

Cai, X., Ballif, J., Endo, S., Davis, E., Liang, M., Chen, D., DeWald, D., Kreps, J., Zhu, T., Wu, Y., 2007. A putative CCAAT-binding transcription factor is a regulator of flowering timing in *Arabidopsis*. *Plant Physiol.* 145, 98-105. https://doi.org/10.1104/pp.107.102079

Chia, T.Y., Müller, A., Jung, C., Mutasa-Göttgens, E.S., 2008. Sugar beet contains a large CONSTANS-LIKE gene family, including a *CO* homologue that is independent of the early-bolting (B) gene locus. *J Exp Bot*. 59, 2735-2748. https://doi.org/10.1093/jxb/ern129

Chu, H.D., Nguyen, K.H., Watanabe, Y., Le, D.T., Pham, T.L.T., Mochida, K., Tran, L.P., 2018. Identification, structural characterization and gene expression analysis of members of the Nuclear Factor-Y family in chickpea (*Cicer arietinum* L.) under dehydration and abscisic acid treatments. *Int J Mol Sci*. 19, 3290. https://doi.org/10.3390/ijms19113290

Chu, H.D., Yen, T.T.H., Pham, C.T., Quynh, L.T.N., Huyen, T.T.T., Trung, N.Q., Dong, H.G., Tran, V.T., Cao, P.B., 2024. Identification of two enzymes for trehalose synthesis and their potential function in growth and development in peanut (*Arachis hypogaea*). *Journal of Tropical Life Science*. 14, 83-94. https://doi.org/10.11594/jtls.14.01.10

Fu, J., Yang, L., Dai, S., 2015. Identification and characterization of the CONSTANS-like gene family in the short-day plant *Chrysanthemum lavandulifolium*. *Mol Genet Genomics*. 290, 1039-1054. https://doi.org/10.1007/s00438-014-0977-3

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., Bairoch, A., 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*. 31, 3784-3788. https://doi.org/10.1093/nar/gkg563

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein identification and analysis tools on the ExPASy server, in: Walker, J.M., *The Proteomics Protocols Handbook*. Springer. Totowa, pp 571-607.

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 Res*. 40, D1178-1186. https://doi.org/10.1093/nar/gkr944

Griffiths, S., Dunford, R.P., Coupland, G., Laurie, D.A., 2003. The evolution of *CONSTANS-like* gene families in barley, rice, and *Arabidopsis*. *Plant Physiol*. 131, 1855-1867. https://doi.org/10.1104/pp.102.016188

Guo, Y., Wang, Q., Zhang, H., Huang, T., Zhang, X., Xie, H., Liu, J., Zhang, P., Li, L., Qin, P., 2022. Responses to waterlogging stress in quinoa seedlings based on metabolomic and transcriptomic analysis. *Environ Exp Bot*. 203, 105044. https://doi.org/10.1016/j.envexpbot.2022.105044

Hinojosa, L., González, J.A., Barrios-Masias, F.H., Fuentes, F., Murphy, K.M., 2018. Quinoa abiotic stress responses: A review. *Plants*. 7, 106. https://doi.org/10.3390/plants7040106

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. https://doi.org/10.1093/bioinformatics/btu817

Jarvis, D.E., Ho, Y.S., Lightfoot, D.J., Schmöckel, S.M., Li, B., Borm, T.J.A., Ohyanagi, H., Mineta, K., Michell, C.T., Saber, N., Kharbatia, N.M., Rupper, R.R., Sharp, A.R., Dally, N., Boughton, B.A., Woo, Y.H., Gao, G., Schijlen, E.G., Guo, X., Momin, A.A., Negrão, S., Al-Babili, S., Gehring, C., Roessner, U., Jung, C., Murphy, K., Arold, S.T., Gojobori, T., Linden, C.G., van Loo, E.N., Jellen, E.N., Maughan, P.J., Tester, M., 2017. The genome of *Chenopodium quinoa*. *Nature*. 542, 307-312. https://doi.org/10.1038/nature21370

Kim, S.Y., Yu, X., Michaels, S.D., 2008. Regulation of CONSTANS and FLOWERING LOCUS T expression in response to changing light quality. *Plant Physiol*. 148, 269-279. https://doi.org/10.1104/pp.108.122606

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 35, 1547-1549. https://doi.org/10.1093/molbev/msy096

La, H.V., Chu, H.D., Tran, C.D., Nguyen, K.H., Le, Q.T.N., Hoang, C.M., Cao, B.P., Pham, A.T.C., Nguyen, B.D., Nguyen, T.Q., Nguyen, V.L., Ha, C.V., Le, H.T., Le, H.H., Le, T.D., Tran, L.P., 2022. Insights into the gene and protein structures of the CaSWEET family members in chickpea (*Cicer arietinum*), and their gene expression patterns in different organs under various stress and abscisic acid treatments. *Gene*. 819, 146210. https://doi.org/10.1016/j.gene.2022.146210

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouzé, P., 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 Res*. 30, 325-327. https://doi.org/10.1093/nar/30.1.325

Liu, M., Sun, W., Li, C., Yu, G., Li, J., Wang, Y., Wang, X., 2021. A multilayered cross-species analysis of GRAS transcription factors uncovered their functional networks in plant adaptation to the environment. *J Adv Res*. 29, 191-205. https://doi.org/10.1016/j.jare.2020.10.004

Liu, Y., Luo, C., Liang, R., Lan, M., Yu, H., Guo, Y., Chen, S., Lu, T., Mo, X., He, X., 2022. Genome-wide identification of the mango CONSTANS (CO) family and functional analysis of two *MiCOL9* genes in transgenic *Arabidopsis*. *Front Plant Sci*. 13, 1028987. https://doi.org/10.3389/fpls.2022.1028987

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., 2021. Pfam: The protein families database in 2021. *Nucleic Acids Res*. 49, D412-D419. https://doi.org/10.1093/nar/gkaa913

Nguyen, L.V., Bertero, D., Hoang, D.T., Long, N.V., 2022. Variation in quinoa roots growth responses to drought stresses. *Journal of Agronomy and Crop Science*. 208, 830-840. https://doi.org/10.1111/jac.12528

Niu, L., Chu, H.D., Tran, C.D., Nguyen, K.H., Pham, H.X., Le, D.T., Li, W., Wang, W., Le, T.D., Tran, L.-S.P., 2020. The GATA gene family in chickpea: Structure analysis and transcriptional responses to abscisic acid and dehydration treatments revealed potential genes involved in drought adaptation. *J Plant Growth Regul*. 39, 1647-1660. https://doi.org/10.1007/s00344-020-10201-5

Oustani, M., Mehda, S., Halilat, M.T., Chenchouni, H., 2023. Yield, growth development and grain characteristics of seven Quinoa (*Chenopodium quinoa* Willd.) genotypes grown in open-field production systems under hot-arid climatic conditions. *Sci Rep*. 13, 1991. https://doi.org/10.1038/s41598-023-29039-4

Pathan, S., Siddiqui, R.A., 2022. Nutritional composition and bioactive components in quinoa (*Chenopodium quinoa* Willd.) greens: A review. *Nutrients*. 14, 558. dhttps://doi.org/10.3390/nu14030558

Patiranage, D.S.R., Asare, E., Maldonado-Taipe, N., Rey, E., Emrani, N., 2021. Haplotype variations of major flowering time genes in quinoa unveil their role in the adaptation to different environmental conditions. *Plant Cell Environ*. 44, 2565-2579. https://doi.org/10.1111/pce.14071

Pereira, E., Encina-Zelada, C., Barros, L., Gonzales-Barron, U., Cadavez, V., Ferreira, I.C.F.R., 2019. Chemical and nutritional characterization of *Chenopodium quinoa* Willd (quinoa) grains: A good alternative to nutritious food. *Food Chem*. 280, 110-114. https://doi.org/10.1016/j.foodchem.2018.12.068

Potter, S.C., Luciani, A., Eddy, S.R., Park, Y., Lopez, R., Finn, R.D., 2018. HMMER web server: 2018 update. *Nucleic Acids Res*. 46, W200-W204. https://doi.org/10.1093/nar/gky448

Robson, F., Costa, M.M., Hepworth, S.R., Vizir, I., Piñeiro, M., Reeves, P.H., Putterill, J., Coupland, G., 2001. Functional importance of conserved domains in the flowering-time gene CONSTANS demonstrated by analysis of mutant alleles and transgenic plants. *Plant J*. 28, 619-631. https://doi.org/10.1046/j.1365-313x.2001.01163.x

Suarez-Lopez, P., Wheatley, K., Robson, F., Onouchi, H., Valverde, F., Coupland, G., 2001. CONSTANS mediates between the circadian clock and the control of flowering in *Arabidopsis*. *Nature*. 410, 1116-1120. https://doi.org/10.1038/35074138

Tovar, J.C., Quillatupa, C., 2020. Heating quinoa shoots results in yield loss by inhibiting fruit production and delaying maturity. *Plant Journal*. 102, 1058-1073. https://doi.org/10.1111/tpj.14699

Turcios, A.E., Weichgrebe, D., Papenbrock, J., 2016. Potential use of the facultative halophyte *Chenopodium quinoa* Willd. as substrate for biogas production cultivated with different concentrations of sodium chloride under hydroponic conditions. *Bioresour Technol*. 203, 272-279. https://doi.org/10.1016/j.biortech.2015.12.061

Wang, P., Liu, Z., Cao, P., Liu, X., Wu, X., Qi, K., Zhang, S., Wu, J., 2017. PbCOL8 is a clock-regulated flowering time repressor in pear. *Tree Genetics & Genomes*. 13, 107. dhttps://doi.org/10.1007/s11295-017-1188-y

Wang, L., Xue, J., Dai, W., Tang, Y., Gong, P., Wang, Y., Zhang, C., 2019. Genome-wide identification, phylogenetic analysis, and expression profiling of *CONSTANS-like (COL)* genes in *Vitis vinifera*. *Journal of Plant Growth Regulation*. 38, 631-643. https://doi.org/10.1007/s00344-018-9878-8

Wenkel, S., Turck, F., Singer, K., Gissot, L., Le Gourrierec, J., Samach, A., Coupland, G., 2006. CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of *Arabidopsis*. *Plant Cell*. 18, 2971-2984. https://doi.org/10.1105/tpc.106.043299

Yan, J., Mao, D., Liu, X., Wang, L., Xu, F., Wang, G., Zhang, W., Liao, Y., 2017. Isolation and functional characterization of a circadian-regulated CONSTANS homolog (GbCO) from *Ginkgo biloba*. *Plant Cell Rep*. 36, 1387-1399. https://doi.org/10.1007/s00299-017-2162-8