

Genome-wide Identification of the *BHLH* Family in *Chenopodium Quinoa*

Yuyang Tao*, Yulin Huang, Ningrui Liu

Sichuan Agricultural University, Ya'an, Sichuan, China.

How to cite this paper: Yuyang Tao, Yulin Huang, Ningrui Liu. (2023). Genome-wide Identification of the *BHLH* Family in *Chenopodium Quinoa*. *Advance in Biological Research*, 4(2), 52-59.

DOI: 10.26855/abr.2023.12.004

Received: November 29, 2023

Accepted: December 27, 2023

Published: January 23, 2024

Corresponding author: Yuyang Tao, Sichuan Agricultural University, Ya'an, Sichuan, China.

Abstract

Chenopodium quinoa is a plant rich in a variety of nutrients, with strong resistance to cold, drought, salt, and barren conditions. *BHLH* transcription factors are one of the largest transcription factor families in plants, and they have a basic/helix-loop-helix structure. *BHLH* transcription factors are involved in the regulation of plant growth and development and play a role in stress resistance. Besides, the *bHLH* family also participates in various signal transduction and anabolic processes, such as light signal transduction and hormone synthesis. In view of the important roles of the *bHLH* family in plant growth, development, anabolism, signaling, and resistance to stress, the *Chenopodium quinoa bHLH* (*CqbHLH*) family was identified in the genome by using bioinformatics methods. Moreover, the structure, chromosomal assignment, replication, and collinearity of all *CqbHLH* genes were analyzed. The results of this study have laid the foundation for identifying *CqbHLH* genes that play a key role in the regulation of plant growth and development and in stress resistance.

Keywords

Chenopodium quinoa, *bHLH* transcription factor, plant resistance, gene identification

1. Introduction

Chenopodium quinoa is an annual herbaceous plant, belonging to the *Chenopodium* genus of the *Chenopodiaceae* family. The ear part of the plant can turn red, purple, or yellow, and the plant resembles *Chenopodium album* in appearance. The ears of the plant in a mature state are like those of the sorghum. The plant height is influenced mainly by the environment and hereditary factors, ranging between 0.3m and 3m. The stem is hard in texture and can be branched or unbranched. The duck palm-shaped leaves are arranged in an alternate fashion, and the leaf margin can be serrated or smooth. The flowers are bisexual and arranged like a corymb, a spike, or a cone. The plant seeds are shaped as small rounded pills and have a small size, with a diameter of 1.5~2mm and a weight of 1.4~3g/1000 grains [1]. *Chenopodium quinoa* contains abundant nutrients and is certified as the only food that can provide nutrients essential to the human body. It is described as the perfect nutrient-dense food for human beings [2]. *Chenopodium quinoa* has a high content of proteins, fibers, and vitamins, and a low content of fat and sugar. It also contains a large number of total polyphenols, saponins, flavonoids, and polysaccharides, playing an important role in the research and development of functional foods, cosmetics, pharmaceuticals, and biopesticides [2]. The protein content in *Chenopodium quinoa* is similar to that in beef and the protein quality is comparable to that in meat and milk. There are diverse amino acids in *Chenopodium quinoa*, including 9 essential amino acids and many other nonessential amino acids, especially lysine, which is not present in most crops. Moreover, *Chenopodium quinoa* is rich in a variety of mineral elements and contains many vitamins necessary for the normal metabolism of the human body [3]. The sugar content, fat content, and calories are low in *Chenopodium quinoa*. All of the above features make *Chenopodium quinoa* an ultra-high-nutrition plant. In addition, *Chenopodium quinoa* is highly resistant to abiotic stress, such as cold [4], drought [5] and salt [6].

Transcription factors play a crucial role in the regulation of many biological processes of plants, such as growth and development, stress resistance, and protection against damage. *BHLH* transcription factors are known for having highly conserved basic/helix-loop-helix regions [7]. There are 50 to 60 amino acids in the *BHLH* structural domain, which consists of two regions with totally different functions, namely, a basic amino acid region and a helix-loop-helix region. The basic region at the N end contains 15 amino acids, 6 of which are basic. This region regulates the transcription of genes mainly by binding with the DNA cis-element E-box (5'-CANNTG-3') [8, 9]. The helix-loop-helix region at the C end comprises 40 to 50 amino acids, which are connected by two hydrophilic and lipophilic α -helices and a hydrophobic amino acid ring structure with unfixed length to form homologous or heterologous dimers for interactions with other *bHLH* proteins [10]. *BHLH* genes are greatly different from each other, except that they all have a highly conserved *bHLH* structural domain [11]. Previous studies have proven that *bHLH* transcription factors play an important role in stress resistance in many types of plants. For instance, the rapid expression of the *GhbHLH1* gene in the leaves of upland cotton can be induced by ABA and PEG, but not by high salt and low temperature. It suggests that *GhbHLH1* possibly participates in responses to osmotic stress [12]. The expression of the *rd22BP1* gene in *Arabidopsis thaliana* is mainly induced by drought and ABA stresses [9]. The *AtMYC2* gene in the *bHLH* family and the *AtMYB2* gene in the MYB family specifically combine with the MYC and MYB sites in the promoter sequence of the *rd22BP1* gene, respectively, thus regulating the ABA-induced expression of genes and enhancing the stress resistance of transgenic plants [9]. *OsbHLH1* is a gene separated from *Oryza sativa* and it is similar to ICE1. Transgenic plants overexpressing *OsbHLH1* show significant improvement in the resistance to low temperatures [13]. In *Populus koreana*, *PkbHLH2* induces the expression of its own target gene after being activated by the transcription factor of its upstream gene *PkbZIP2*, thus enabling the plant to resist drought [14]. Wang et al. [15] isolated *CsICE1* and *CsCBF1* genes from camellia trees. These two genes are homologous to *AtICE1* in *Arabidopsis thaliana* and have been demonstrated to participate in the ICE1-CBF cold stress regulatory pathway.

The *bHLH* family in plants has not been thoroughly studied, and existing studies only focus on model plants *Arabidopsis thaliana* and *Oryza sativa*. Further research is required to study the specific functions of *bHLH* transcription factors in other species. In addition, *bHLH* transcription factors have aroused wide concern among scholars as key regulatory proteins in the stress-resisting signaling network of plants. Since the discovery of the first *bHLH* transcription factor in corn [16, 17], multiple biological techniques such as transcriptomics, proteomics, bioinformatics, and genetics approaches have been used to explore the functions of *bHLH* proteins in plants. It has been found that *bHLH* proteins are involved in plant growth, development, and metabolism processes, but more importantly, in stress responses. To date, the *bHLH* genes in wheat, mulberry trees, *Clerodendrum japonicum*, and many other plant species [18-20] have been systematically identified and analyzed, but there is no report about the analysis of *bHLH* genes in *Chenopodium quinoa*.

Considering the role of *bHLH* transcription factors in plant resistance to stress and the lack of studies on the *bHLH* genes in *Chenopodium quinoa*, the present study identified 246 *Chenopodium quinoa bHLH* (*CqbHLH*) genes in the genome using bioinformatics approaches and systematically analyzed their structure, location on chromosomes, replication, and evolutionary relationships. The study results provide a basis for the screening of stress-resisting genes in *Chenopodium quinoa* and for the in-depth study of the functions of *CqbHLH* genes.

2. Methods

2.1 Genomic data acquisition

Genomic data of *Chenopodium quinoa*, *Oryza sativa*, *Beta vulgaris*, and *Glycine max* plants were extracted from the Phytozome database (<https://phytozome.jgi.doe.gov/>). Genomic data of *Arabidopsis thaliana* and *Fagopyrum tataricum* were obtained from the Arabidopsis Information Resource (TAIR) database (<https://www.arabidopsis.org/>) and the Molecular Breeding Knowledgebase (<http://www.mbkbase.org/Pinku1>), respectively.

2.2 Identification of *CqbHLH* genes

The genomic sequence of *Chenopodium quinoa* was downloaded from plant genome databases. The amino acid sequences of all *Arabidopsis thaliana bHLH* (*AtbHLH*) proteins were derived from the TAIR database. By taking the *AtbHLH* genes as targets, we used TB tools to identify all possible *bHLH* genes in the genome of *Chenopodium quinoa* [21]. After comparing the amino acid sequences of *CqbHLH* proteins using the BLASTP program of NCBI, non-target sequences were deleted. In addition, the Web CD tool of NCBI was used to predict the conserved region of *CqbHLH* genes, and then the wrong genes were removed.

2.3 Phylogenetic analysis of *CqbHLH* genes

Multiple sequence alignments of *CqbHLH* proteins were made using MEGA7. Then, the phylogenetic tree of *CqbHLH* and *AtbHLH* proteins was constructed using the maximum likelihood (ML) method, with the Bootstrap value set at 1000. The *CqbHLH* genes on adjacent branches of the evolutionary tree were divided into subgroups according to the classification of *AtbHLH* proteins. Then, an evolutionary tree for *CqbHLH* proteins was constructed using the same method for subsequent gene structure analysis.

2.4 Chromosomal assignment, gene structure, and protein motif analysis of *CqbHLH* genes

The locations of genes were extracted from the genome annotation Gff file. The structure of *CqbHLH* genes was visualized with the help of TB tools. Then, *CqbHLH* proteins were analyzed using the online MEME analysis tool (<http://meme-suite.org/>), with the number of motifs set at 10, totaling 5–200 amino acids in length. Moreover, the structure annotation of each motif for replacement was obtained from the Pfam database (<https://pfam.xfam.org/>) in order to further confirm the core structural domain of *bHLH* genes.

2.5 Collinearity analysis

Collinearity genes in the genome of *Chenopodium quinoa* and between *Chenopodium quinoa* and other species were identified using the multiple collinearity scan toolkit MCScanX. Based on the information of locations given in the genome annotation files of *Chenopodium quinoa* and other species, the locations of all collinearity gene pairs were visualized on their respective chromosomes.

3. Results and discussion

3.1 Genome-wide identification and phylogenetic analysis of *CqbHLH* genes

A total of 246 *CqbHLH* genes were identified from the genome of *Chenopodium quinoa* through bidirectional BLASTP alignments. After multiple sequence alignments of 158 *AtbHLHs* and 246 *CqbHLHs*, the phylogenetic tree of the *CqbHLH* and *AtbHLH* families was constructed to show the evolutionary relationships of *CqbHLH* proteins (Figure 1).

According to the classification method of the *AtbHLH* family, all *CqbHLH* and *AtbHLH* genes were divided into 23 subgroups, each of which contained one or several *CqbHLH* genes. Among all the subgroups, subgroup No.6 has the least number of *CqbHLH* genes (only 1 gene), while subgroup No.23 contains the most *CqbHLH* genes (23 genes) (Figure 1).

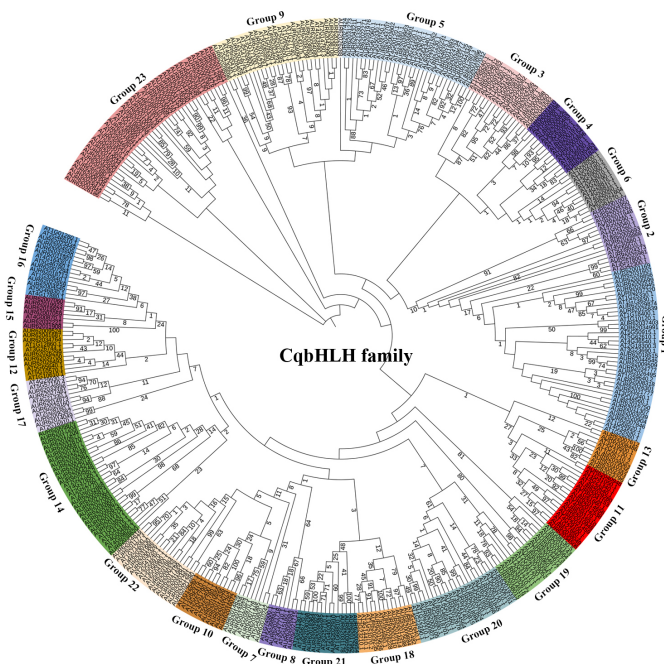
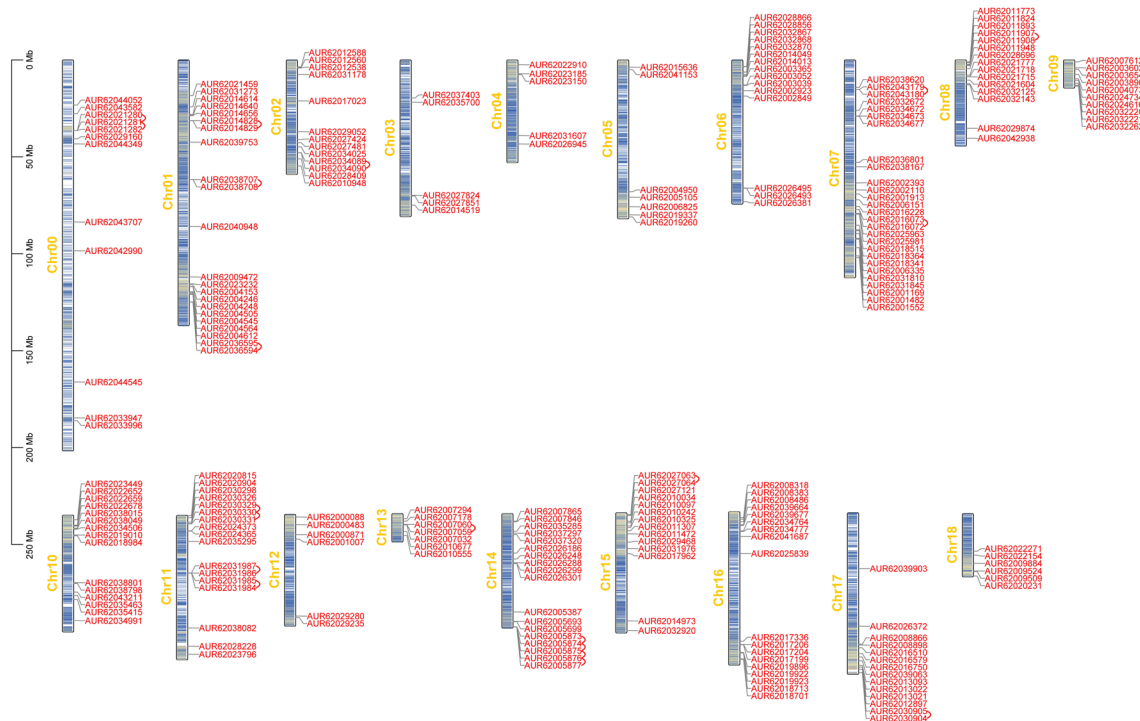


Figure 1. Phylogenetic tree of 246 *CqbHLH* and 158 *AtbHLH* genes.

3.2 Chromosomal assignment of *CqbHLH* genes

The information on the locations of all *CqbHLH* genes was acquired from the genomic structure annotation files of *Chenopodium quinoa*, and these genes were separately assigned to 18 chromosomes of *Chenopodium quinoa* (Figure 2). The location results showed that among 246 *CqbHLH*s, 234 genes were successfully located on chromosomes No.1 to No.18, and 12 genes were located on chromosome No.0 due to match failure possibly caused by low genome quality. The 234 successfully located *CqbHLH*s were distributed unevenly on the 18 chromosomes. Chromosome No.7 contains the most *CqbHLH* genes (27 genes), while there are fewer *CqbHLH* genes on chromosomes No.3, No.4, No.5, No.12, No.13, and No.18 (Figure 2). Twenty-one tandem repetitive genes were detected on 10 chromosomes of *Chenopodium quinoa*, and they may arise from the tandem duplication of genes (Figure 2).



Red lines represent tandem repeats.

Figure 2. *CqbHLH*s' locations on chromosomes and tandem repeats.

3.3 Structure and protein motif analysis of *CqbHLH* genes

Based on the chromosomal assignment results, the structure and motifs of the *CqbHLH* family were further analyzed. The protein motif analysis showed that almost all *CqbHLH* proteins have motifs No.1 and No.2, indicating that these two motifs possibly constitute the core structure domain of the *bHLH* family and play a decisive role in certain functions of *bHLH* proteins. The motif type varies among different subgroups of *bHLH* proteins and is associated with certain functions of the proteins. Some proteins contain a few types of motifs. For example, there is only motif No.1 in *AUR6208866* and *AUR62031845* and only motif No.2 in *AUR62009524*. In addition, most genes from the same branch have identical or similar motifs, which may explain their similar functions. A closer kinship is associated with a more similar composition and arrangement of conserved motifs. It suggests that these genes possibly have identical or similar functions in the growth and development of *Chenopodium quinoa* (Figure 3A). Gene structure analysis revealed that the number of introns ranges from 0 to 4 in 246 *CqbHLH*s on average. However, there are 8 introns in *AUR62029235*, and the introns in *AUR62024734* and *AUR62011893* have long base sequences. The introns may play a potential role in the regulation of gene expression. The number of exons ranges from 2 to 8, and the genes from the same branch have a similar structure, indicating that *CqbHLH* proteins maintain basic functions in the process of evolution. The structural domain varies among different subgroups, implying that *bHLH* genes have developed different structures during the long-term evolution and thus participate in different biological processes of plants (Figure 3B).



Advance in Biological Research

(A) Conserved motifs of proteins, with different colors representing different types of motifs; (B) Gene structure, with different colors representing different structures

3.4 Intraspecific and interspecific collinearity analysis of *CqbHLH* genes

Intraspecific collinearity analysis of *CqbHLH* genes was made, and the results showed that there are 14 pairs of segmental duplicated genes. It demonstrates the high conservativeness of *bHLH* genes in the evolution process and the important role of conserved sequences in performing particular functions (Figure 4).

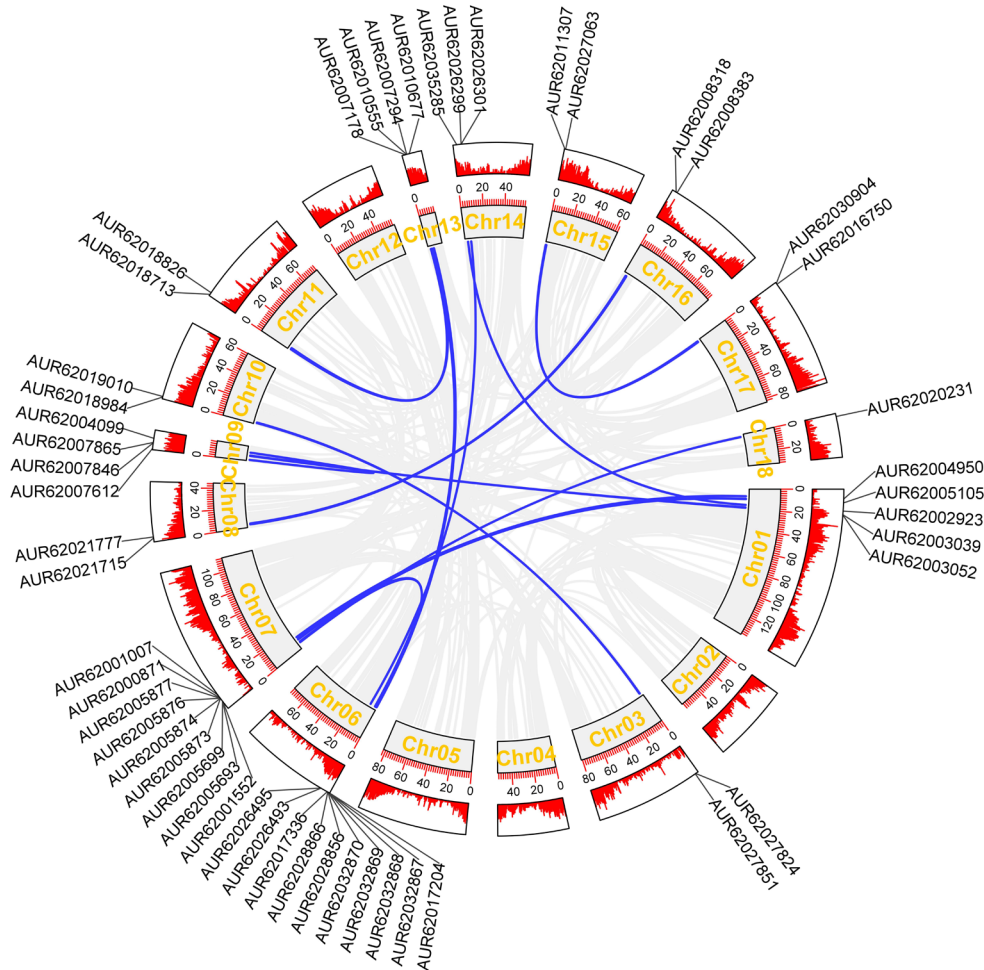


Figure 4. Intraspecific collinearity analysis of *CqbHLH* genes.

Interspecific collinearity analysis of genes is also an approach to investigating the evolution of species. The collinearity between *CqbHLH* genes and *bHLH* genes in typical dicotyledon and monocotyledon plants was examined. The results showed that only 32 pairs of collinear genes were found between *CqbHLH* genes and *AtbHLH* genes, and the number is relatively small. It indicates that these two types of plants differentiate early in the evolution process. There are 107, 60, and 64 pairs of collinear genes between *CqbHLH* genes and *Beta vulgaris* genes, *Fagopyrum tataricum* genes, and *Oryza sativa* genes, respectively, and the numbers are moderate. The largest number of collinear gene pairs (173 pairs) were observed between *CqbHLH* genes and soybean genes. This finding suggests that *bHLH* genes in *Chenopodium quinoa* and *Glycine max* are highly homologous, and these two plants differentiate late (Figure 5). According to the intraspecific and interspecific collinearity analysis, gene duplication and collinearity genes between species are the driving force and cause of the rapid evolution and expansion of the *bHLH* family.

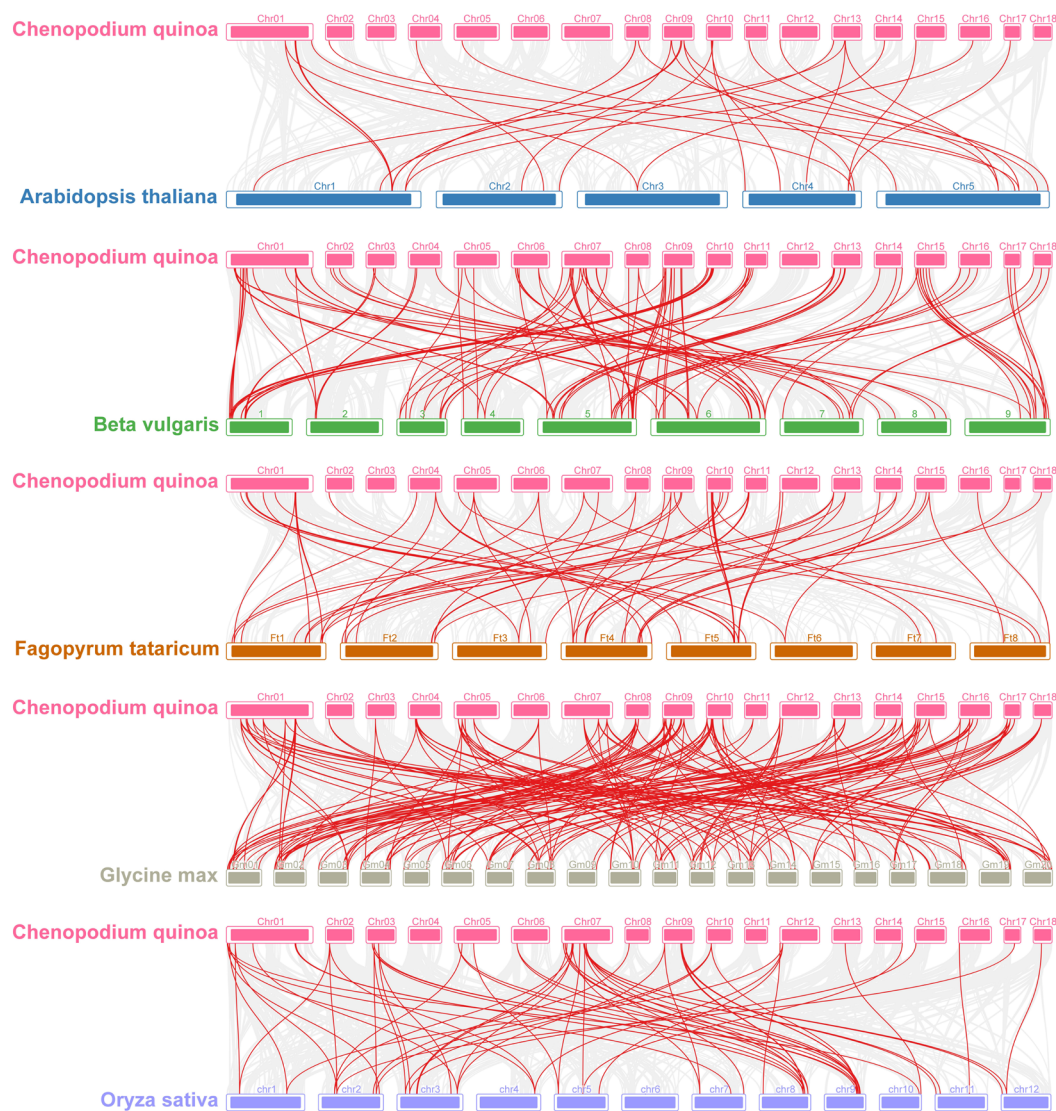


Figure 5. Interspecific collinearity analysis between *CqbHLH* genes and *AtbHLH* genes, *Beta vulgaris* genes, *Fagopyrum tataricum* genes, *Glycine max* genes, and *Oryza sativa* genes.

4. Conclusion

In this study, 246 *bHLH* genes were identified in the genome of *Chenopodium quinoa*, and the number is large, compared with 162 *bHLH* genes in *Oryza sativa* [22], 159 *bHLH* genes in *Solanum tuberosum* [23], and 152 *bHLH* genes in *Dioscorea esculenta* [24]. For this reason, *Chenopodium quinoa* has good stress resistance. According to the gene replication analysis, tandem duplication and segmental duplication play an important role in the amplification of the *CqbHLH* family. Gene structure and protein motif analysis indicate that motifs No.1 and No.2 of *CqbHLH* proteins are highly conserved in subgroups, which may be related to the biological functions of *CqbHLH* proteins. Different subgroups differentiate into different motifs to perform different functions. Intraspecific and interspecific collinearity analysis reveals that *CqbHLH* genes expand through intraspecific gene duplication and interspecific gene differentiation. Taken above, the number of *bHLH* genes in *Chenopodium quinoa* is more than that in other plants. Intraspecific tandem replication, fragment duplication, and interspecific differentiation promote the expansion of the *CqbHLH* family. The systematical identification of *CqbHLH* genes in this study helps with the selection of stress-resistant *bHLH* genes. The study results have important theoretical and practical significance for cultivating stress-resistant *Chenopodium quinoa* varieties and lay a foundation

for the verification of the functions of these varieties.

References

- [1] Yuxing Sun & Wenjuan Chi. Analysis of the promotion prospects of *Chenopodium quinoa* [J]. *Journal of Green Science and Technology*, 2017(7): 2.
- [2] Yunshuang Ding, Yawen Zeng, Kang Min, et al. A comprehensive study and utilization of functional components in *Chenopodium quinoa* [J]. *Current Biotechnology*, 2015, 000(005): 340-346.
- [3] Liming Wang, Ning Ma, Song Li, et al. The nutritional value and application prospects of *Chenopodium quinoa* [J]. *Science and Technology of Food Industry*, 2014, 35(1): 5.
- [4] Ming Cheng, Shenju Suo, & Qinan Yang. Analysis of cold resistance in *Chenopodium quinoa* seedlings [J]. *Modern Agricultural Science and Technology*, 2020(19): 3.
- [5] Liyuan Hou, Yanhui Dong, Yali Li, et al. Research progress and prospect of drought resistance in *Chenopodium quinoa* [J]. *Jiangsu Agricultural Sciences*, 2021, 49(11): 7.
- [6] Lili Li, Qiyang Jiang, Fengjuan Niu, et al. Advances in research on the salt resistance mechanism of quinoa *Chenopodium quinoa* [J]. *Journal of Agricultural Science and Technology*, 2016(2): 10.
- [7] Friedrichsen D M, Nemhauser J, Muramitsu T, et al. Three Redundant Brassinosteroid early response genes encode putative bhlh transcription factors required for normal growth [J]. *Genetics*, 2002.
- [8] Chinnusamy V, Ohta M, Kanrar S, et al. ICE1: a regulator of cold-induced transcriptome and freezing tolerance in *Arabidopsis* [J]. *Genes & Development*, 2003, 17(8):1043-1054.
- [9] Abe H, Yamaguchi-Shinozaki K, Urao T, et al. Role of *Arabidopsis* MYC and MYB Homologs in Drought and Abscissic Acid-Regulated Gene Expression [J]. *The Plant Cell*, 1997, 9(10):1859-1868.
- [10] White P L, Alvistur E, Dias C, et al. Nutritive values of crops, nutrient content and protein quality of quinoa and caihua, edible seed products of the andes mountains [J]. *Journal of Agricultural & Food Chemistry*, 1955, 3(6): 531-534.
- [11] Albert N W, Lewis D H, Huaibi Z, et al. Light-induced vegetative anthocyanin pigmentation in *Petunia* [J]. *Journal of Experimental Botany*, 2009, 60(7): 2191-2202.
- [12] Meng C M, Zhang T Z, & Guo W Z. Molecular cloning and characterization of a Novel *Gossypium hirsutum* L. bHLH Gene in Response to ABA and drought stresses [J]. *Plant Molecular Biology Reporter*, 2009, 27(3): 381-387.
- [13] Wang Y J, Zhang Z G, He X J, et al. A rice transcription factor OsbHLH1 is involved in cold stress response [J]. *Theoretical & Applied Genetics*, 2003, 107(8):1402-1409.
- [14] Yingjie Wu. Analysis of the functions of bHLH genes in *populus Korea* [D]. Northeast Forestry University, 2011.
- [15] Wang Y, Jiang C J, Li Y Y, et al. CsICE1 and CsCBF1: two transcription factors involved in cold responses in *Camellia sinensis* [J]. *Plant Cell Reports*, 2012, 31(1):27-34.
- [16] Zhang T T, Lv W, Zhang H S, et al. Genome-wide analysis of the basic Helix-Loop-Helix (bHLH) transcription factor family in maize [J]. *BMC plant biology*, 2018, 18(1).
- [17] Ludwig S R, Habera L F, Dellaporta S L, et al. Lc, a member of the maize R gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcriptional activators and contains the myc-homology region [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1989, 86(18):7092-7096.
- [18] Guihua Liao, Yu Duan, Congcong Wang, et al. Identification and bioinformatics analysis of the bHLH transcription factor family in *clerodendrum japonicum* [J]. *Molecular Plant Breeding*, 2023:1-17.
- [19] Tian Hui, Bingqi Shen, Lianchun Wang, et al. Genome-wide identification and analysis of the bHLH transcription factor family in *Morus alba* [J]. *Molecular Plant Breeding*, 2019.
- [20] Silu Fu. Identification and expression analysis of bHLH transcription factors in *Triticum aestivum* [D]. Chinese Academy of Agricultural Sciences, 2014.
- [21] Chen C, Chen H, Zhang Y, et al. TBtools: An integrative toolkit developed for interactive analyses of big biological data [J]. *Molecular Plant*, 2020, 13(8).
- [22] Li X, Duan X, Jiang H, et al. Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and *arabidopsis* [J]. *Plant Physiology*, 2006, 141(4): 1167.
- [23] Sun H, Fan H J, & Ling H Q. Genome-wide identification and characterization of the bHLH gene family in tomato [J]. *BMC Genomics*, 2015.
- [24] Lingquan Pei. Identification and preliminary analysis of the bHLH family in *dioscorea esculenta* [D]. Jiangsu Normal University, 2017.