



Full Length Article

Expression Analysis of Trihelix Gene Family in Maize (*Zea mays*) using Bioinformatics

Long Jiang, Xiaoming Yu*, Dianyuan Chen*, Huan Wang, Duo Cao, Jingzhe Zhang and Jianming Li

College of Agronomy, Jilin Agricultural Science and Technology University, Jilin, Jilin 132101, Jilin Province, P.R. China

*For correspondence: yuxm446@nenu.edu.cn; jllykjxyjl@163.com

Received 29 November 2019; Accepted 31 December 2019; Published 03 March 2020

Abstract

Trihelix transcription factors of plants are strongly induced by diseases, chilling, drought and high salt and widely involved in the response to biotic and abiotic stresses. This manuscript investigates the basics of Trihelix, such as its structural domain, evolutionary relationship, chromosome location and subcellular location by means of bioinformatics. The results showed that the Trihelix gene family of maize can be divided into five subfamilies, each subfamily domain has different distribution; chromosome mapping showed that only one Trihelix gene on chromosome 7 and two or more Trihelix genes on other chromosomes; most members of the Trihelix gene family of maize are highly conserved through multiple developmental stages. Trihelix expression analysis showed an important role of this gene family in the growth and development of maize. © 2020 Friends Science Publishers

Keywords: Maize; Trihelix transcription factor; Bioinformatics analysis

Introduction

Grain production is the foundation of agricultural development in a country. As one of the main grain crops, maize plays an important role in agriculture (Zhao *et al.* 2016) and is rich in various nutrients, such as fat, vitamins, protein, cellulose and trace elements (Zhao *et al.* 2018). However, there are many disadvantages in the process of traditional breeding due to the variety of maize and the complexity of heredity. Using molecular biology to identify gene function and assist traditional breeding can speed up breeding.

As a small family in plants, Trihelix transcription factor family was only limited to the study of light response until the end of the 20th century. With the progress of biotechnology and the rapid development of bioinformatics, more and more members of Trihelix transcription factor family have been identified (Zhou *et al.* 2015). For example, 30 Trihelix genes have been identified in *Arabidopsis*, 31 in rice and other species (Zhou *et al.* 2015). More studies have found that Trihelix family genes play an important role in the growth and development of different parts of plants, such as flowers, epidermal hairs, stomata, seeds and embryos, as well as in the response of biological and abiotic stresses such as diseases, salt stress, drought stress and cold stress (Zhang *et al.* 2017). However, the study of identifying and analyzing the gene family based on the whole genome level of maize has not been reported. In this study, the Trihelix transcription factor family in maize was analyzed

and the basic information statistics, conservative domain analysis, gene structure prediction, gene location on chromosome, amino acid physical and chemical properties and expression analysis of Trihelix transcription factor gene family in maize were carried out by using bioinformatics method. This study will provide useful information for further analysis of the evolution and biological function of trihelix transcription factor gene family in maize.

Materials and Methods

Plant material

From the plant transcription factor database PlantTFDB (<http://plantfdb.cbi.pku.edu.cn/>), the nucleic acid and protein sequences of maize Trihelix transcription factor family gene and *Arabidopsis* Trihelix transcription factor family gene for correlation analysis of this study was downloaded (Liu *et al.* 2015).

Test method

Acquisition of trihelix transcription factor sequence: The Trihelix sequences of *Arabidopsis* and maize were downloaded from the plant transcription factor database PlantTFDB (<http://plantfdb.cbi.pku.edu.cn/>), 34 and 57 amino acid sequences of transcription factor family proteins were obtained respectively. The Trihelix protein sequences of maize and *Arabidopsis* obtained in the above database

include the proteins translated by multiple transcriptions isoforms with different Trihelix genes. In this study, the longest transcripts and their encoded proteins selected by Trihelix gene with multiple transcripts were analyzed subsequently, and 28 Trihelix gene coding regions and their proteins of *Arabidopsis* were obtained. Sequence and 44 Trihelix gene coding regions and their protein sequences of maize.

Construction of phylogenetic tree of transcription factor trihelix gene

Clustal X software was used to carry out multiple matching analysis on the protein sequences of maize and *Arabidopsis*. Based on the comparison results, using MEGA5.05 to build neighbor joining NJ, set Bootstrap as 1000 repetitions and other as default parameters (Zhao *et al.* 2015).

Analysis of conserved domains of amino acid sequences of trihelix transcription factor family proteins

Using the Pfam Domain pattern redrawer function of TBtools software, the conservative domain of amino acid sequence of Trihelix protein in maize was predicted.

Location of trihelix gene on chromosome

According to the published information of maize genome B73 RefGen_v3, the distribution of 39 Trihelix transcription factor family genes on 10 maize chromosomes was determined. The location of Trihelix gene on chromosome was found by BLSAT comparison in Maize GDB database (<https://www.maizegdb.org/>).

Analysis of physico-chemical properties of amino acids

The number of amino acids, molecular weight, theoretical isoelectric point, number of aliphatic amino acids and hydrophobicity of proteins were analyzed by online ProtParam tool provided by ExPASy (<https://web.expasy.org/protparam/>) (Xie *et al.* 2014).

Sub-cellular localization of trihelix

Using Plant-mPLoc (www.csbio.sjtu.edu.cn/), the location of 44 Trihelix transcription factor family proteins in cells was predicted (Zhu *et al.* 2012).

Expression analysis of trihelix family gene

Based on the published transcriptome data of maize tissues at different developmental stages, the expression patterns of 39 genes of Trihelix transcription factor family in different maize tissues and developmental stages were analyzed. The heat map through the Amazing Heatmap function in TBtools was drawn.

Results

Identification of trihelix transcription factors

According to the Trihelix protein identified in the plant transcription factor database plantTFDB, for the proteins translated from different transcripts with the same Trihelix gene, the protein with the longest amino acid sequence was selected for the following analysis. After screening, 28 Trihelix proteins of *Arabidopsis* and 44 Trihelix proteins of maize were obtained, which were AC209784.3, GRMZM2G002978, GRMZM2G016637, GRMZM2G016649, GRMZM2G021831, GRMZM2G023119, GRMZM2G031493, GRMZM2G037128, GRMZM2G037493, GRMZM2G037823, GRMZM2G047370, GRMZM2G063203, GRMZM2G080583, GRMZM2G081445, GRMZM2G084684, GRMZM2G111760, GRMZM2G126148, GRMZM2G134439, GRMZM2G149590, GRMZM2G153575, GRMZM2G156348, GRMZM2G157219, GRMZM2G162840, GRMZM2G163157, GRMZM2G169580, GRMZM2G301122, GRMZM2G305362, GRMZM2G314660, GRMZM2G320827, GRMZM2G326783, GRMZM2G334722, GRMZM2G339957, GRMZM2G375307, GRMZM2G379179, GRMZM2G380094, GRMZM2G392168, GRMZM2G414159, GRMZM2G415229, GRMZM2G427087, GRMZM2G428470, GRMZM2G469873, GRMZM2G481163, GRMZM5G818655 and GRMZM5G850092.

Phylogenetic analysis of trihelix gene in maize and *Arabidopsis thaliana*

The 28 Trihelix genes from *Arabidopsis* and 44 Trihelix genes from maize were sequenced by ClustalW and the phylogenetic tree was established. Results showed that Trihelix transcription factors of maize and *Arabidopsis* can be divided into five subfamilies, named GT-1 subfamily, GT-2 subfamily, SIP1 subfamily, GT γ subfamily and SH4 subfamily respectively (Fig. 1). Among them, GT-1 subfamily has 6 members, GT-2 subfamily; 11 members, SIP1 subfamily; 9 members, GT γ subfamily; 9 members and SH4 subfamily also has 9 members. It was found that AT1g54060 and AT3g14180 in *Arabidopsis* regulated seed development (Zhou *et al.* 2015). These two genes are located in SIP1 subfamily, hence, it is speculated that the genes in SIP1 subfamily may have the function of regulating seed development. In GT-2 subfamily, *Arabidopsis* AT1g33240 (*GTL1*) is related to the development of epidermal hair, and has the function of water conservation under drought stress. Therefore, the gene of GT-2 subfamily may be related to the formation of maize stem, leaf and organ. It can reduce transpiration and keep water under drought condition, but the specific function needs to be further verified.

Analysis of conserved domain of trihelix transcription factor family protein sequence

Conservative domain is a kind of highly conserved domain in the process of biological evolution. The analysis of the

Table 2: Analysis of physical and chemical properties and subcellular localization of Trihelix transcription factor family proteins

Gene	Locus	Amino acid number	Molecular weight	Theoretical isoelectric point	Fat coefficient	Average coefficient of hydrophobicity	Predicted location
Trihelix1	AC209784.3	682	72410.16	9.61	83.65	-0.152	Cell membrane. Nucleus.
Trihelix2	GRMZM2G002978	519	55331.10	7.19	69.79	-0.438	Chloroplast.
Trihelix3	GRMZM2G016637	335	36762.77	5.95	61.58	-0.805	Nucleus.
Trihelix4	GRMZM2G016649	774	83274.20	5.90	50.58	-0.882	Nucleus.
Trihelix5	GRMZM2G021831	439	46481.48	8.87	53.08	-0.803	Nucleus.
Trihelix6	GRMZM2G023119	210	23061.93	10.02	56.24	-0.796	Nucleus.
Trihelix7	GRMZM2G031493	277	30181.88	8.94	67.15	-0.570	Nucleus.
Trihelix8	GRMZM2G037128	669	72502.03	5.91	81.08	-0.311	Nucleus.
Trihelix9	GRMZM2G037493	714	76269.47	5.93	52.59	-0.862	Nucleus.
Trihelix10	GRMZM2G037823	214	24343.26	8.74	61.67	-0.742	Nucleus.
Trihelix11	GRMZM2G047370	405	46187.92	5.91	68.91	-0.896	Nucleus.
Trihelix12	GRMZM2G063203	379	41935.71	6.28	63.11	-0.727	Nucleus.
Trihelix13	GRMZM2G080583	668	71709.60	5.78	53.8	-0.758	Nucleus.
Trihelix14	GRMZM2G081445	318	34724.07	9.59	63.08	-0.750	Nucleus.
Trihelix15	GRMZM2G084684	381	40870.93	9.38	69.87	-0.640	Nucleus.
Trihelix16	GRMZM2G111760	366	38952.08	5.01	71.53	-0.589	Nucleus.
Trihelix17	GRMZM2G126148	664	70652.29	5.61	54.04	-0.796	Chloroplast.
Trihelix18	GRMZM2G134439	436	49782.63	6.47	59.79	-1.059	Nucleus.
Trihelix19	GRMZM2G149590	510	56488.55	8.81	78.22	-0.454	Nucleus.
Trihelix20	GRMZM2G153575	334	36440.24	5.89	66.95	-0.794	Nucleus.
Trihelix28	GRMZM2G314660	533	57191.20	6.09	58.91	-0.727	Nucleus.
Trihelix29	GRMZM2G320827	725	76401.90	6.73	51.71	-0.732	Chloroplast.
Trihelix30	GRMZM2G326783	208	22949.16	11.09	66.35	-0.682	Nucleus.
Trihelix31	GRMZM2G334722	673	76473.17	9.57	75.23	-0.447	Nucleus.
Trihelix32	GRMZM2G339957	387	40507.55	4.55	70.26	-0.522	Nucleus.
Trihelix33	GRMZM2G375307	350	38799.78	9.7	61.71	-0.847	Nucleus.
Trihelix34	GRMZM2G379179	337	36717.50	6.94	69.08	-0.631	Nucleus.
Trihelix35	GRMZM2G380094	319	35873.99	5.6	60.75	-0.966	Nucleus.
Trihelix36	GRMZM2G392168	402	42967.31	7.15	58.41	-0.805	Nucleus.
Trihelix37	GRMZM2G414159	392	42106.22	6.61	54.16	-0.846	Nucleus.
Trihelix38	GRMZM2G415229	776	82452.12	6.11	48.45	-0.830	Nucleus.
Trihelix39	GRMZM2G427087	271	31896.87	8.94	58.38	-1.204	Chloroplast.
Trihelix40	GRMZM2G428470	206	22628.80	11.19	68.88	-0.657	Nucleus.
Trihelix41	GRMZM2G469873	317	33468.76	9.17	69.15	-0.466	Nucleus.
Trihelix42	GRMZM2G481163	398	45440.35	5.97	73.54	-0.814	Nucleus.
Trihelix43	GRMZM5G818655	321	34106.44	9.41	65.26	-0.567	Nucleus.
Trihelix44	GRMZM5G850092	528	56690.91	7.38	61.31	-0.722	Nucleus.

proteins are different, and the amino acid number, molecular weight, theoretical isoelectric point, fat coefficient, and hydrophilic average coefficient are quite different in different Trihelix transcription factors. As shown in Table 2, the maximum number of amino acids is Trihelix 21, up to 1021; the minimum is Trihelix 40, up to 206; the maximum number of fatty amino acids is Trihelix 1, up to 83.65; the minimum number of fatty amino acids is Trihelix 4, only 50.58. The proteins of Trihelix transcription factor family in maize contain both acid amino acids and basic amino acids. The molecular weight of amino acids is between 24343.26 and 110243.1, most of which are neutral and basic. The average coefficient of hydrophilicity is only negative but not positive, which indicates that all the transcription factor family proteins are hydrophilic rather than hydrophobic. Fat coefficient can be used as an index of protein stability. Generally, the higher fat coefficient is, the higher protein stability. It was found that the fat coefficient of Trihelix transcription factor family was between 48-83, hence was speculated that the stability of Trihelix transcription factor family was poor. This specific situation needs further study.

Subcellular localization of trihelix transcription factor

We use the online tool Plant-mPLoc (<http://www.csbio.sjtu.edu.cn/bioinfo/plant-multi/>) to predict subcellular localization (Zhu *et al.* 2012). Results showed that the Trihelix family of transcription factors are basically located in the nucleus, in which Trihelix 1 is present in both the nucleus and the cell membrane; Trihelix 2, Trihelix 17, Trihelix 29 and Trihelix 39 are present in the chloroplast (Table 2). In conclusion, the Trihelix family of transcription factors plays a major role in the nucleus, which may have transcriptional regulation function.

Tissue-specific expression of trihelix in maize

We analyzed the expression of 39 Trihelix genes in different stages of maize development by using the transcriptome sequencing data released by Stelpflug *et al.* (2016) and drew a heatmap based on the FPKM value of each gene in each stage of maize development. The tissues analyzed include germinated seeds, different regions of roots, seedlings, stems at different positions, apical

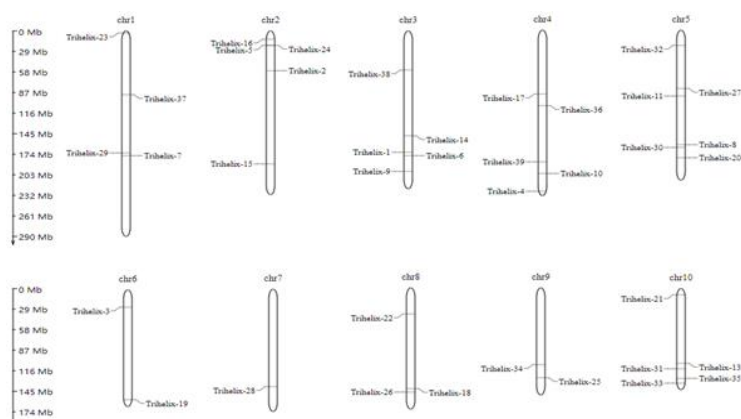


Fig. 3: Location of Trihelix transcription factor family genes on different chromosomes of maize

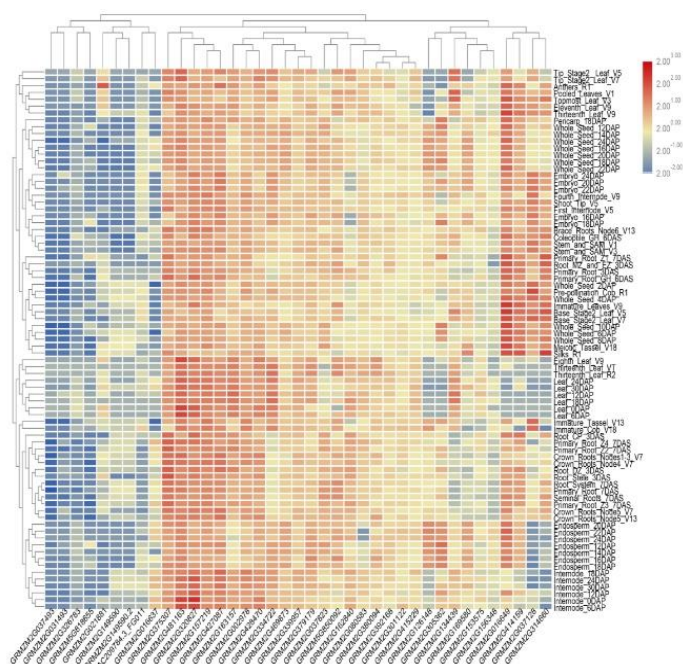


Fig. 4: Expression map of Trihelix gene in different tissues of maize

meristem of stems, leaves, internodes, spikes, anthers, and filaments (maize whiskers) at different stages of development. The expression pattern of Trihelix transcription factor family gene is different in different maize tissues and development stages (Fig. 4). Most of the genes were highly expressed in different tissues and periods, except GRMZM2G037493, GRMZM2G031493, GRMZM2G326783 and GRMZM5G818655. At the stage of Anthers R1, GRMZM2G021881 expression level was increased. The overall expression of GRMZM2G149590, GRMZM2G149590.2, AC207984.3_FG011 and GRMZM2G016637 was also relatively low. Compared with the overall expression level, expression of GRMZM2G375307, GRMZM2G481163 and GRMZM2G320827 were highest. Compared with the

overall expression level, the five genes, GRMZM2G157219, GRMZM2G427087, GRMZM2G002978, GRMZM2G428470 and GRMZM2G334722, were also expressed higher. The overall expression of GRMZM2G469873, GRMZM2G339957, GRMZM2G379179, GRMZM2G037823, GRMZM2G850092, GRMZM2G162840, GRMZM2G080583, GRMZM2G380094, GRMZM2G392168, GRMZM2G301122 and GRMZM2G414229 were also high. The expression levels of GRMZM2G01649, GRMZM2G414159, GRMZM2G037128 and GRMZM2G314660 were significantly increased in meiosis and root formation.

Discussion

Transcription factors play an important role in plant growth, development and response to changes in the external environment, and are the key to regulate various physiological activities (Zhu *et al.* 2019). In recent years, many transcription factors related to drought, high salt, low temperature, hormone, pathogen response and development have been isolated from plants. It has been found that the over-expression of some transcription factors can enhance the resistance and adaptability of plants to stress by transgenic means (Zhuang *et al.* 2009; Liu *et al.* 2010).

Trihelix transcription factor family plays an important role in plant growth and development and response to stress (Zhou *et al.* 2017). It is involved in plant growth and development, including light response gene regulation, flower organ morphogenesis and response to stress, including abiotic stress such as cold damage, drought, high salt and biological stress such as pathogen stress (Liang *et al.* 2017). In this study, 44 Trihelix sequences were screened from the Trihelix transcription factor family of maize by bioinformatics analysis, which is more than 28 Trihelix sequences of *Arabidopsis*. It is suggested that the expansion of Trihelix gene may be to make maize better adapt to some environments and then to develop evolutionary characteristics. By comparing the physical and chemical properties of Trihelix transcription factor protein, it was found that the number of acid amino acids in 44 Trihelix transcription factor proteins was more than basic amino acids. There are only hydrophilic proteins in this gene family, and the molecular weight difference between each gene is large, indicating that Trihelix transcription factor family protein is relatively rich.

In this study, the distribution relationship of Trihelix transcription factors in maize was discovered by evolutionary tree. It was found that the members of Trihelix gene family involved in seed development were located in SIP1 subfamily, and the members involved in abiotic stress were mostly located in GT-2 subfamily, indicating that the function and classification of Trihelix transcription factors in maize were not significantly related (Li *et al.* 2015). In the conservative domain analysis, it was found that each member of the Trihelix transcription factor family has the same domain, but some members also have different conservative domains, suggesting the diversity of gene function. According to the expression analysis, GRMZM2G016649, GRMZM2G414159, GRMZM2G037128 and GRMZM2G314660 were expressed in a large number in meiosis and root formation, suggesting that they play an important role in the seed development of plants.

Conclusion

The Trihelix transcription factor family in maize was identified by bioinformatics and will provide some basic data for further utilization of Trihelix transcription factors in

maize breeding.

Acknowledgements

The authors acknowledge the Scientific Research Project fund of Jilin Provincial Education Department (JJKH20190979KJ), Science and Technology Innovation Development Project of Jilin City (#201831781 to X.Y.), Natural Science Foundation of Jilin Province of China (#20180101233JC to Z.-Y.X.) and Doctor Talent Research Foundation of Jilin Agricultural Science and Technology University (#20185002 to L. J.).

References

- Li Y, XD Liu, YM Dong, ZM Xie, SY Chen (2015). Cloning and functional analysis of the cotton Trihelix transcription factor GhGT29. *Genetics* 37:1218–1227
- Liang SM, DL Luo, YJ Cheng, LA Yu, YJ Yang, JY Chen, WJ Lu, JF Kuang (2017). Identification and characterization of a Trihelix transcription factor MaGTL1a from banana. *J Fruit Sci* 34:935–945
- Liu C, XX Li, YL Su, YF Guo (2015). Genome-wide identification, phylogenetic analysis and expression profiling of the SBP transcription factor family in *Nicotiana glauca*. *Chin Tob Sci* 36:1–11
- Liu CF, J Zou, XB Chen (2010). Advances in DREB transcription factors and plant abiotic stress tolerance. *Biol Bull* 10:26–30
- Niu YL, XM Jiang, XY Xu (2016). Research advances on transcription factor MYB gene family in plant. *Mol Plant Breed* 14:2050–2059
- Stelpflug SC, RS Sekhon, B Vaillancourt (2016). An expanded maize gene expression atlas based on RNA sequencing and its use to explore root development. *Plant Genome* 9:85–94
- Xie T, S Wang, L Huang, X Wang, LP Kang, LP Guo (2014). Transcriptome-based bioinformatics analysis of *Arnebia euchroma* ERF transcription factor family. *Chin J Trad Chin Med* 39:4732–4739
- Zhang CL, Q Yuan, TT Zhao, XY Xu (2017). Bioinformatics analysis of Trihelix transcription factors from Chilli Pepper. *Chin Agric Bull* 33:32–38
- Zhao JL, WJ Yao, SJ Wang, YB Jiang, BR Zhou (2015). AP2/ERF gene family in *Populus trichocarpa* by bioinformatics. *J Northeast For Univ* 43:21–29
- Zhao JR, S Wang, M Li, HY Lv, DW Wang, YQ Ge, X Wei, WC Yang (2018). Current status and perspective of maize breeding. *J Plant Genet Res* 19:435–446
- Zhao JR, RH Wang, XX Liu (2016). The current situation of maize industry and the development trend of biological breeding in China. *Biol Bus* 3:45–52
- Zhou H, J Qian, RF Li, DD Chen, RJ Fang, WG Zhao (2017). Research of the *Arabidopsis* Trihelix transcription factor gene family. *J Jiangsu Sci Technol Univ* 31:231–236
- Zhou YJ, H Wang, JH Ji (2015). Classification, structure and function of Trihelix transcription factor family in plants. *J Hubei Agric Sci* 54:5501–5503
- Zhu J, Z Sun, D Shi, S Song, L Lian, L Shi, A Ren, H Yu, M Zhao, (2019). Dual functions of AreA, a GATA transcription factor, on influencing ganoderic acid biosynthesis in *Ganoderma lucidum*. *Environ Microbiol* 21:4166–4179
- Zhu S, L Zeng, PZ Wu, YP Chen, HW Jiang, GJ Wu, MR Li (2012). Cloning and sequence analysis of the cDNA of plastid-located glycerol-3-phosphate acyltransferase (JcGPAT2) gene from *Jatropha curcas* L. *J Guangdong Agric Sci* 39:1–5
- Zhuang J, RH Peng, F Gao, XY Fu, B Zhu, XF Jin, J Zhang, AS Xiong, QH Yao (2009). Cloning and analysis of AP2/ERF-B1 Subfamily transcription factors from *Brassica napus* L. *Huyou* 15. *J Nucl Agric Sci* 23:435–441