Genome-wide identification of trihelix transcription factors in the apple genome in silico

Polina Kuzmitskaya (1401polina@gmail.com)
Institute of Genetics and Cytology NAS of Belarus: GNU Institut genetiki i citologii Nacional'noj akademii nauk Belarusi https://orcid.org/0000-0002-1929-4315

Ekaterina Koroleva
Institute of Genetics and Cytology NAS of Belarus: GNU Institut genetiki i citologii Nacional'noj akademii nauk Belarusi

Oksana Urbanovich
Institute of Genetics and Cytology NAS of Belarus: GNU Institut genetiki i citologii Nacional'noj akademii nauk Belarusi

Research Article

Keywords: Malus, trihelix transcription factors, expression analysis, stress

Posted Date: March 17th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2573217/v1

Version of Record: A version of this preprint was published at Journal of Applied Genetics on July 16th, 2023. See the published version at https://doi.org/10.1007/s13353-023-00770-y.

Abstract

Trihelix transcription factors are involved in the growth and development of plants, as well as various stress responses. In this work, we have identified 37 genes of the trihelix family in the apple genome (MdTH). They were located on 13 chromosomes. Phylogenetic analysis showed that apple trihelix transcription factors belong to one of six subfamilies (GT-1, GT-2, SH4, SIP1, GTγ and GTδ). Genes from different groups have differences in the gene structure and conserved functional domains. Analysis of the promoter regions of apple trihelix transcription factors showed that their expression may be regulated by light, phytohormones, stress factors such as anaerobic stress, drought, low temperature, a pathogen attack, an injury, etc. In our work, we have demonstrated that drought, salinity, as well as high and low temperatures, affect the expression of genes of the apple trihelix family. Among the genes under study, the greatest increase in the expression level under soil salinity was observed in MdTH4 and MdTH24 genes. The MdTH4 gene also showed the greatest increase in the relative expression level during drought. With decreasing temperature, the greatest change in the expression level was observed in MdTH30. The same gene was among the three genes whose expression level was significantly decreasing when exposed to drought. With an increase in temperature, the expression level of three genes demonstrated the greatest increase: MdTH8, MdTH20, and MdTH36.

Introduction

Trihelix transcription factors are characteristic of plants only. The genes of this family were first discovered in peas in 1990s. It was demonstrated that they are able to bind to the core sequence 5’-G-Pu-(T/A)-A-A-(T/A)-3’ of the promoter region of the *rbcS-3A* gene involved in the regulation of light-dependent expression (Green et al. 1987). The family of these factors was originally named GT for the ability to bind to the light-dependent homonymous regulatory element. The DNA-binding domain of these transcription factors has a tandem three-helix structure (helix-loop-helix-loop-helix), which led to the subsequent renaming of the family. Subsequent studies showed that the three-helix structure of GT factors is similar to the structure of the Myb/SANT-LIKE-DNA-binding domain (Nagano 2000). In the plant genome, genes encoding GT factors have evolved from the genes encoding Myb/SANT-LIKE transcription factors. Breaks between the helices in trihelix transcription factors form a DNA-binding structure, which is different from the proteins containing the Myb/SANT-LIKE domain (Nagano 2000; Qin et al. 2014). According to the data provided in protein domain databases such as Pfam, the Myb/SANT-LIKE domain is a conserved domain of trihelix transcription factors.

In recent years, trihelix transcription factors have been actively explored. Systematic work was carried out on such plants as *Arabidopsis*, tomato, chrysanthemum, rice and other (Kaplan-Levy et al. 2012; Yu et al. 2015; Song et al. 2016; Li et al. 2019). Thus, 30 members of the family were found in *Arabidopsis*, which belonged to several subfamilies (GT-1, GT-2, GTγ, SH4, and SIP1) (Kaplan-Levy et al. 2012; Gao et al. 2009). Thirty six genes belonging to the trihelix family were identified in tomato. They were classified into six subfamilies (GT-1, GT-2, SH4, SIP1, GT-γ and GT-δ) (Yu et al. 2015). The last family was not found in *Arabidopsis*. In the rice genome, 41 genes encoding the transcription factors of the trihelix family were

identified, which were represented by the members of five subfamilies (SIP1, GTγ, GT, SH4, and GTδ) (Li et al. 2019).

Study of the expression patterns of genes encoding proteins with the trihelix domain showed their participation in a variety of processes occurring during the life of a plant. In rice, trihelix transcription factors are expressed in four tissues at six developmental stages, but their expression patterns differ. The expression level is affected by abiotic stress (drought, salinity), as well as signaling molecules, such as abscisic acid and hydrogen peroxide (Li et al. 2019). In *Arabidopsis*, genes belonging to the GT-1 subfamily may be involved in a response to salinity and the attack of pathogens (Murata et al. 2002). In soybean seedlings, the transcription factors GmGT-2A and GmGT-2B are induced by abscisic acid, drought, high salt levels, and cold (Xie et al. 2009). *Arabidopsis* mutants with the damaged GTL1 gene had fewer stomata than wild-type plants, and therefore, they were more resistant to drought due to less water loss (Yoo et al. 2010). In rice, the expression of the *OsGTγ-1* gene increased by 2.5–10 times in response to salt stress; it also increased when exposed to abscisic acid (Fang et al. 2010). However, in chrysanthemum, the expression of some genes belonging to the trihelix family decreased after exposure to abscisic acid (Song et al. 2016). It was also demonstrated that genes belonging to the trihelix family are involved in plant morphogenesis. In *Arabidopsis*, for example, the transcription factor PETAL LOSS (PTL) determines the number of petals per flower and sepal fusion (Brewer et al. 2004). The SH4 subfamily gene in rice promotes the development and functioning of the deciduous layer in mature seed peduncles (Li et al. 2006). Thus, genes encoding trihelix transcription factors play an important role in the life of plants, participating in many processes (Kaplan-Levy et al. 2012).

An apple tree is an important fruit crop. Its yield depends on many factors, including light, water regime, soil composition and many other factors. The identification and study of genes that influence both normal developmental processes and a stress response may contribute to a better understanding of molecular mechanisms that enable the life of this plant. In the work presented, we carried out the *in silico* identification of genes encoding Trihelix transcription factors in the domestic apple genome of the Golden Delicious variety, analysis of the structural organization of hypothetical proteins encoded by them, evaluation of their promoter regions, study of their phylogenetic relationships with the homologues of other species, as well as their response to abiotic stress (drought, high and low temperatures).

Materials And Methods

Identification and analysis of trihelix family members in the apple tree

Nucleotide and hypothetical protein sequences of apple genes (*Malus x domestica*, Golden Delicious variety) were obtained from the Genome Database for Rosaceae (https://www.rosaceae.org/) and chromosome sequences from the NCBI GenBank. The protein HMM model of the Myb/SANT-LIKE (PF13837) domain was downloaded from the PFAM database (El-Gebali et al. 2018). It was used to search for matches among hypothetical protein sequences of the apple tree using HMMER3 (http://hmmer.org/). The identified potential matches were tested by scanning with SMART

(http://smart.embl-heidelberg.de/) (Letunic et al. 2020). The sequences containing the Myb/SANT-LIKE domain were considered as candidates for the role of transcription factors belonging to the trihelix family. Alignment of gene sequences was performed using the BLAST algorithm (Altschul et al. 1997). The online service Circoletto (Darzentas 2010) based on Circos (Krzywinski et al. 2009) was used to visualize sequence similarity. The exon-intron structure of genes was generated using the Gene Structure Display Server v.2.0 (GSDS v.2.0) (Hu et al. 2014).

Analysis of hypothetical MdTH proteins

The molecular mass, the isoelectric point of the MdTH were calculated using the Expasy Bioinformatics Resource Portal: https://www.expasy.org/ (Gasteiger et al. 2003). The intracellular localization was determined using the Localizer 1.0.4 (https://localizer.csiro.au/cgi-bin/script.py) (Sperschneider et al. 2017).

Promoter analysis of MdTH genes

Aiming to study the regulatory elements of genes, an analysis of a DNA sequence of 2000 bp above the first codon was carried out. Identification of *cis*-regulatory elements was performed using the PlantCARE database (Lescot et al. 2002).

Calculation of potential miRNA targets and MAPK-specific phosphorylation sites

Potential miRNA targets were calculated using the psRNATarget online server (https://www.zhaolab.org/psRNATarget/) (Dai et al. 2018) with the following parameters: the maximum likelihood – 2, UPE (maximum energy to unpair the target site) – less than 25. At the time of analysis, 207 miRNAs of the apple tree were introduced in the database. MAPK-specific phosphorylation sites were predicted by the MusiteDeep tool (http://musite.net) (Wang et al. 2020), and the default confidence threshold was used for calculations.

Multiple sequence alignment and phylogenetic analysis

With a view to performing a phylogenetic analysis of the amino acid sequences of the transcription factors belonging to the trihelix family, *Arabidopsis* sequences were loaded from the GenBank database, according to the list presented in (Kaplan-Levy et al. 2012). The sequences were aligned using the ClustalW algorithm. The phylogenetic tree was constructed using the maximum likelihood method and MEGA11 (Tamura et al. 2021).

Plant materials and qRT-PCR

In order to study the expression of MdTH, the clonal rootstocks of the MM-106 apple variety were used. They are characterized by medium frost and low drought resistance. The rootstocks were grown under long daylight conditions (16 hours per day/8 hours per night) at 22°C. The plants were divided into five groups, and one of the groups, the control, was under the conditions described above. The second group

was exposed to drought by removing the plants from the earthen lump and then placing them on the filter paper; the third group was exposed to elevated temperatures (40°C) and the fourth to low (4°C). The plants of the fifth group were watered with a salt solution at a concentration of 0.2 mol/L. Each group was represented by three trees.

The selection of leaves was carried out at 0, 2nd, 4th, and 24th hour. The selection at 0 hour was carried out immediately after the exposure of plants to stress factors. The selected leaves were immediately frozen in liquid nitrogen.

RNA extraction from frozen leaves was performed using the CTAB method (Jaakola et al. 2001). The quality of extracted RNA was assessed using agarose gel electrophoresis. The concentration of the RNA obtained was measured using the NanoDrop device (ND-8000 Spectrophotometer, Thermo Scientific). RNA was purified from DNA using the DNase I (RNase-free) reagent (Thermo Scientific, EU), according to the protocol. Construction of the cDNA minus strand was performed using the RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo Scientific, EU), according to the manufacturer's protocol.

Primer sequences for setting qPCR are presented in Supplementary Table 1. Primer efficiency was assessed by the qPCR reaction on the cDNA matrix of the MM-106 apple variety, and the results obtained demonstrated that the efficiency of each pair of primers is within the range of acceptable values of 90-110% (Bustin et al. 2009). The elongation factor *Ef-1a* was chosen as the internal control (Zhao et al. 2012). As the negative control, instead of cDNA, the equal amount of deionized water was used. To analyze the expression of genes encoding MdTH in response to abiotic stress factors, three biological repeats were used in each experimental condition. The calculation of a relative expression level of the genes under study was carried out using the method 2^(−ΔΔCT) (Rao et al. 2013).

Results

Identification of genes encoding trihelix transcription factors in the domestic apple genome

A search for genes encoding the transcription factors of the trihelix family was carried out using the HMMER3 software and the HMM profile downloaded from the PFAM database. Confirmation of candidate genes belonging to the trihelix family was performed using SMART. In total, 37 genes belonging to the trihelix family (marked as MdTH) with conserved domains characteristic of the transcription factors of this family were found in the apple genome of the Golden Delicious variety.

The length of the amino acid sequence encoded by them varies from 278 to 917 amino acid residues (514.5 on average). The molecular weight of MdTH proteins ranges from 31.9 to 101.3 kDa (57.7 kDa on average). Estimated isoelectric points range from 5.23 to 9.78 (7.4 on average). The results of the cellular localization assessment *in silico* demonstrate that 28 hypothetical transcription factors of the MdTH are located in the nucleus, 2 — in chloroplasts (Table 1).

Evaluation of the chromosomal distribution of MdTH showed that 13 out of 17 chromosomes (except for 1, 8, 9, and 15) carry at least one of such genes. The largest number of genes (six) is located on 5th and 6th chromosomes, while chromosomes 3, 7, 13, 16 and 17 carry only one gene each. There is no clear relationship between the chromosome length and the number of genes located on it (Fig. 1). Genes located on different chromosomes may have more similarities than those on the same chromosome evidencing that polyploidization processes played a role in the formation of the apple genome. The domestic apple genome was formed as a result of duplication and subsequent reorganization of nine chromosomes of a precursor species (Velasco et al. 2010). A model for the origin of the *M. × domestica* genome designed on the basis of sequencing results suggests that chromosomes 5 and 10 originated from chromosome I of an ancestral species. This explains a high degree of similarity in the structural organization of MdTH8 and MdTH21, MdTH9 and MdTH22, MdTH10 and MdTH23 genes (Fig. 2). Chromosomes 2 and 7, 4 and 12, 6 and 14, 12 and 14 or their individual regions are also combined by common origin (Velasco et al. 2010). As a result, we can observe that MdTH genes located on chromosomes 2 and 7, 4 and 12, 6 and 14, 12 and 14, as well as on chromosomes 5 and 10, are characterized by a higher degree of homology relative to each other compared with other genes of this family (Fig. 2).

Table 1 Trihelix family members identified in the apple tree

Gene name	Gene ID	Protein length (a.a.)	MW (kDa)	PI	Localization of cells	Chromosome of the gene localization
MdTH1	MD02G1219900	366	40481.75	9.16	Nucleus	2
MdTH2	MD02G1244500	491	53176.73	8.57		2
MdTH3	MD02G1247700	569	65668,97	6.06	Nucleus	2
MdTH4	MD02G1318400	742	81180.12	9.75	Nucleus	2
MdTH5	MD03G1089900	674	74397.21	9.06	Nucleus	3
MdTH6	MD04G1154600	287	34785.25	7.82		4
MdTH7	MD04G1221900	658	74968.15	5.84	Nucleus	4
MdTH8	MD05G1024300	448	51480.28	6.12	Nucleus	5
MdTH9	MD05G1024400	665	74379.05	5.66	Nucleus	5
MdTH10	MD05G1024500	572	64620.71	8.07	Nucleus	5
MdTH11	MD05G1174500	291	32988.47	9.33	Nucleus	5
MdTH12	MD05G1322900	474	53022.84	9.78	Nucleus	5
MdTH13	MD05G1361500	412	46153.60	5.61	Nucleus	5
MdTH14	MD06G1021000	288	33181.41	8.88		6
MdTH15	MD06G1046900	353	39091.11	9.32	Nucleus	6
MdTH16	MD06G1127800	764	83305.38	5.75	Nucleus	6
MdTH17	MD06G1143600	917	100966.22	7.62	Chloroplast	6
MdTH18	MD06G1172900	485	55103.10	6.50		6
MdTH19	MD06G1196500	338	37845.55	7.74	Nucleus	6
MdTH20	MD07G1068800	566	64781.83	5,69	Nucleus	7
MdTH21	MD10G1024800	447	51549,42	6.11	Nucleus	10
MdTH22	MD10G1025100	664	74452.43	6.47	Nucleus	10
MdTH23	MD10G1025200	563	63544.23	8.56	Nucleus	10
MdTH24	MD10G1163600	278	31889.18	7.72	Nucleus	10
MdTH25	MD10G1338800	130	15328.55	9.79	Nucleus	10
MdTH26	MD11G1017600	593	65658.60	9.21	Nucleus	11
MdTH27	MD11G1079000	497	54971.24	6.35	Nucleus	11

MdTH28	MD12G1018900	538	60775.83	6.36	Nucleus	12
MdTH29	MD12G1168300	296	35422.79	8.35		12
MdTH30	MD12G1238000	656	73982.06	5.70	Nucleus	12
MdTH31	MD13G1109800	365	42097.23	6.21	Nucleus	13
MdTH32	MD14G1016900	532	60326.45	6.48	Nucleus	14
MdTH33	MD14G1058900	674	74986.94	9.10		14
MdTH34	MD14G1143600	727	79127.51	5.23	Nucleus	14
MdTH35	MD14G1158900	917	101294.49	8.36	Chloroplast	14
MdTH36	MD16G1109700	382	43571.81	6.05		16
MdTH37	MD17G1017400	417	46616.93	5.47	Nucleus	17

Phylogenetic analysis of trihelix proteins

In order to assess phylogenetic relationships between the apple trihelix transcription factors and the model plant *Arabidopsis thaliana*, an unrooted phylogenetic tree, containing 33 hypothetical proteins of trihelix *Arabidopsis thaliana* and 37 of the apple tree, was constructed (Fig. 3). Apple trihelix transcription factors belong to one of six subfamilies (GT-1, GT-2, SH4, GTγ, GTδ, and SIP1) depending on the conserved amino acids of the GT domain, the number of DNA-binding motifs, and the classification of homologues.

The largest is the cluster that combines GT-2 subfamily genes. It includes 14 apple tree genes and 6 *Arabidopsis* genes. The SH4 cluster is represented by 13 genes and eight out of them are of the apple tree. Fifteen genes are included in the SIP1 cluster; five out of them are apple tree genes. The GT-1 cluster combines 10 genes and six out of them are of the apple tree. The GTγ cluster is represented by five genes, including two of the apple tree. The GTδ cluster is represented by two apple tree genes. The genes of this subfamily were not originally isolated from *Arabidopsis* (Gao et al. 2009). They are described in tomato, sunflower, and rice (Song et al. 2021; Yu et al. 2015; Li et al. 2019).

Analysis of the gene structure and motif of the MdTH

MdTHs contain the consensus sequences of conserved amino acid motifs (Fig. 4A). Sequences, containing similar motifs, are grouped into clusters the location of which correlates with phylogenetic analysis results (Fig. 3, Fig. 4A). Analysis of conservative motifs within the subfamilies showed that motifs 2, 3, 6, and 8 are found only in GT-2 subfamily members. Representatives of the GT-2 subfamily contain the largest number of motifs. Four of them have all ten motifs, two have nine, and another two contain eight. Motifs 5 and 10 contain GT-2 and GT-1 subfamily sequences.

Motif 1 turned out to be the most widespread and may be found in the composition of all sequences. Motif 7 was discovered in 36 out of 37 sequences, and it could be found no more than once in each. Motif 4 is the next most frequently occurring and was found in 32 proteins. Motif 6 was represented only in 6 sequences belonging to the TG2 subfamily. Motif 10 may be found only in the composition of six sequences belonging to two subclasses — GT-2 and GT-1. The SH4 subfamily was characterized by the smallest number of conserved motifs from 1 to 3.

The coding sequences of most MdTH genes are separated by introns and their number varies from 1 to 16 (Fig. 4c). The number of introns in GT-1, GT-2, and SH4 subfamily members ranges from 1 to 8. The intron length and location in the gene differ both in the representatives of different subfamilies and within the subfamilies themselves. The greatest differences from other subfamilies in the exon-intron organization were found in MdTH17 and MdTH35 genes belonging to the GTδ subfamily. Both representatives of the GTδ subfamily carry 16 introns in their composition.

Seven genes have no introns. These are MdTH8 and MdTH21 genes belonging to the GTγ subfamily. Introns are absent in four out of five representatives of the SIP1 subfamily, while the remaining representative of MdTH2 carries six introns in the gene sequence. The coding MdTH25 GT-2 subfamily sequences are not separated by introns.

Identification of hypothetical *cis*-elements in the promoter regions of MdTH

Analysis of the nucleotide structure of regions located before the MdTH genes suggests that they are involved in a complex network of interactions between regulatory proteins controlling the vital activity of plant cells.

One of the key factors influencing the expression of MdTH genes is light. The elements involved in light-dependent signaling pathways were discovered in the promoter regions of all family members with no exceptions found (Figure 5). Among them, G-box (detected in the promoter regions of 33 out of 37 genes, which is 89.2%), Box 4 (78.4%), GT-1 (64.9%), TCT-motif (56.8%), and AE-box (51.3%) are the most frequently occurring. The 3-AF1 binding site, AAAC-motif, A-box, ACA-motif, ACE, AT1-motif, ATC-motif, ATCT-motif, Box II, CAAT-box, CAG-motif, chs-CMA1a/ 2a, chs-Unit 1 m1, GA-motif, Gap-box, GATA-motif, G-Box, GT1-motif, GTGGC-motif, I-box, LAMP-element, L-box, MRE, Sp1, TCC- motif, and TCCC-motif elements were also found.

The expression of MdTH genes under stress conditions is evidenced by the detected STRE elements (91.9% of the promoter regions contain them) and TC-reach repeats (29.8%). ARE elements involved in the response to anaerobic stress were found in the regulatory regions of 81.1% genes; CG involved in the response to anoxia in 21.6%; MBS, DRE1 and DRE regulating the expression of genes under the conditions of drought in 51.4%, 5.4% and 13.5% respectively; LTR involved in the response to low temperatures in 48.6%. The possible expression of MdTH genes in response to wound stress and pathogen attack may be judged by the presence of WUN, Box S, WRE3 and WRE motifs found in the

regulatory regions of 35.1%, 18.9%, 43.2% and 5.4% genes respectively. At least one of these elements contains the regulatory regions of 67.6% MdTH genes (Figure 6).

In the promoter regions of each of the MdTH genes, the binding sites of regulatory proteins with the MYB domain were found. They are involved in the regulation of a response to biotic and abiotic stress, including epigenetic control, hormonal signaling pathways, regulation of cell differentiation and shape, and phenylpropanoid biosynthesis. As found, 97.3% of the MdTH genes contain characteristic DNA sequences for binding transcription factors from the MYC family. They are part of the hormonal regulation system for jasmonic and salicylic acids participating in coordination of plant growth and development in response to various types of stress and other biological processes. With regard to analyzed regulatory regions, 54.1% contains W-box — the DNA sequence to which WRKY transcription factors bind. They perform many functions, such as the formation of resistance to diseases, stress, ontogeny, and other, including hormonal regulation (Figure 6).

Induction of the expression of trixelix transcription factors may be indirectly triggered by plant hormones: abscisic acid as evidenced by the presence of at least one of the following elements: ABRE, ABRE2, ABRE3a, ABRE4, and AT-ABRE in the regulatory regions of 91.9% of genes; gibberellin (at least one of p-box, TACT-box, and GARE elements was found in 75.7%), salicylic acid (TCA and/or *as-1* elements have regulatory regions of 91.9% of genes), ethylene (ERE elements were found in 48.6% of cases). CGTCA and TGACG sequences involved in the methyl jasmonate-induced signaling pathway were found in the promoter regions of 75.7% and 83.8% of genes respectively. Certain elements associated with signaling cascades regulated by hormones were found in the promoter regions of each gene (Figure 6).

The expression of MdTH genes may be tissue-specific and depend on the plant development stage. This is evidenced by AC1, AC2, telo-box, and F-box elements (the latter is also involved in the response to biotic and abiotic stress); dOCT, which is involved in the regulation of motif I (expression in the root) growth and development. The expression of MdTH genes may be indicated by the following elements: AP-1, which regulates flowering; O2 — prolamin metabolism; MBS-I — flavonoid biosynthesis; GCN-4 associated with the expression in the endosperm; CAT — in meristems; and HD-ZIP1 responsible for palisade mesophyll differentiation. Apparently, a list of functions of trixelix transcription factors also includes the regulation of proliferation and a cell cycle, as evidenced by the detected regulatory elements: re2f-1, NON, MSA-like, and e2fb.

Thus, analysis of the nucleotide structure of regions located directly before the genes encoding the transcription factors of the trixelix family suggests that they are involved in a complex system of interaction of regulatory processes essential for plant life. Analysis of detected regulatory elements allows assuming that the expression of MdTH genes occurs continuously; however, the conditions of plant existence may lead to a change in its level. It may be influenced by ontogeny stages, the specifics of tissues and organs, various environmental factors, including changes in the level of light, as well as various types of biotic and abiotic stress effects. The results obtained are consistent with the data derived during the study of the expression of trihelix proteins on other objects demonstrating that most of

them are expressed under normal conditions, but their expression level may change significantly when exposed to stress.

Searching for phosphorylation sites specific for mitogen-activated protein kinase (MAPK) and predicting miRNA targets

The paper (Li et al., 2015) demonstrates that in *Arabidopsis* MAP KINASE4 triggered the rapid phosphorylation of the ASR3 gene, belonging to the trihelix family, after treatment with MAMP (microbe-associated molecular patterns) (Li et al. 2015). This suggests that the members of this family may be post-translationally regulated by phosphorylation. Analysis of hypothetical MdTH proteins showed that each of them has at least one putative phosphorylation site. At that, serine residues predominate among them. Thus, among GT-2 family members, the number of supposed phosphorylated serine residues is 1-19 (more than 10 in most representatives) and of threonine is from 0 to 5; in SH4 family members, from 2 to 49 of serine residues and from 0 to 7 of threonine residues; in SIP1 representatives, the number of putative phosphorylation sites by serine ranges from 7 to 22 and threonine 0-4. In GT-1 family members, the number of putative phosphorylated serine residues ranges from 6 to 11 and threonine – from 1 to 3. No threonine phosphorylation sites were found in GTγ family members with 13-15 putative serine phosphorylation sites. Representatives of GTδ have three putative phosphorylated threonine and 31-40 serine residues each.

miRNAs are a class of non-coding regulatory RNAs that are involved in the regulation of gene expression by inhibiting translation or cleavage of the target mRNA (Unver et al. 2009; Eldem et al. 2013; Zhang 2015). Analysis of mRNAs encoding MdTH showed that 21 out of them contain at least one of the specific binding sites for one or more of 75 apple miRNAs able to cleave mRNA or inhibit its translation. The largest number of binding sites for various miRNAs (32) was found in the mRNA sequence MdTH34. MdTH12 and MdTH17 contain 19 binding sites each; MdTH36 and MdTH8 — six each; MdTH5 and 15 — five each; MdTH35 has four; MdTH7, MdTH29 and MdTH30 have three each. MdTH4, MdTH6 and MdTH31 contain two binding sites each. MdTH1, MdTH9, MdTH14, MdTH20, MdTH28, MdTH32 and MdTH37 have one. At that, MdTH34 contains two copies of binding sites mdm-miR535b and mdm-miR535c, while all other genes contain only one copy of each binding site.

MdTH gene expression patterns under conditions where MM-106 apple rootstocks are exposed to stress

Various adverse environmental factors, such as drought, high and low temperatures, and soil salinity, may affect the growth, development and productivity of plants. In order to adapt to the conditions of stress, plants change the expression of genes associated with stress. In the work presented here, we tested a response of 14 MdTH genes to the effect of certain unfavorable abiotic factors, since their representatives are known to be involved in adaptation processes in other plants (Fang et al. 2010; Li et al. 2019; Murata et al. 2002; Xie et al. 2009).

Evaluation of the expression of 14 MdTH genes in response to drought (Fig. 7) showed that in 10 out of them a relative expression level had increased and most strongly in MdTH4 at 24th hour of exposure. The

expression level of three genes (MdTH21, MdTH23, and MdTH30) had, on the contrary, a tendency to decrease when exposed to drought.

A decrease in temperature to 4°C caused a significant decrease in the MdTH4 expression at the 2nd and later hours of exposure (Figure 7). The greatest increase in the expression level was observed in the MdTH30 gene at the 4th hour. Most genes (MdTH8, MdTH9, MdTH20, MdTH21, MdTH31, MdTH35, and MdTH36) demonstrated an increase in their expression level at the 2nd and/or 4th hour of exposure. A day later, the relative level of their expression turned out, as a rule, to be lower than the starting point of sampling, but not always (in the MdTH35 gene, for example, the relative level of expression at the 24th hour of exposure was higher than at the zero point of sampling).

In order to study the effect of elevated temperatures, the plants were placed in the climate chamber with a temperature of 40°C. Evaluation of the expression levels of MdTH genes in MM-106 rootstocks (Figure 7) showed that in most of the genes under study the expression changes slightly. In some (e.g. MdTH4, MdTH11, MdTH22 etc.), it decreases. In three genes, it increases significantly, e.g. MdTH8 at the 24th hour of exposure; MdTH20 and MdTH36 at the 4th hour with the latter having the strongest changes in response to elevated temperatures.

Among all of the stress factors investigated, exposure to a saline solution caused the greater homogenous reaction of the genes under study: we observed an increase in the expression level of most of them (MdTH4, MdTH8, MdTH9, MdTH11, MdTH20, MdTH21, MdTH22, MdTH24, MdTH30, MdTH31, and MdTH36) at the 2nd hour of expose. At that, the magnitude of a response of different genes was not the same. The greatest increase in the expression level was observed in MdTH4 and MdTH24 genes. At later sampling points, MdTH gene expression levels decreased (Figure 7).

Discussion

Characterization of MdTH genes

In this work, 37 apple genes were identified as encoding the transcription factors of the trihelix family using the HMM-based search followed by the confirmation using the SMART algorithm for the presence of a characteristic Myb/SANT-LIKE domain. The trihelix family of tomato contains the same number of genes, while other plants may have fewer or more (Yu et al. 2015). The wheat genome, for example, encodes 94 of such genes, the soybean genome encodes 71, *Brassica rapa* – 52, rice – 41, Tatar buckwheat – 31, and chrysanthemum – 20 (Xiao et al. 2019; Liu et al. 2020; Wang et al. 2017; Li et al. 2019; Ma et al. 2019; Song et al. 2016). The number of genes in the trihelix family could be influenced by various polyploidization processes that some species had gone through in the course of evolution. Thus, the hexaploid wheat genome was formed, for example, as a result of interspecies hybridization of three diploid species as a result of which trihelix genes have homologous copies in the A, B, and D genome of *T. aestivum*, while diploid *T. uratru* possesses 22 genes of this family (Xiao et al. 2019. Duplication of the ancestral genome plays a great role in the formation of the domestic apple genome (Velasco et al. 2010). In this regard, gene pairs located on the chromosomes of common origin are characterized by high

homology in structure (Fig. 2), as well as exon-intron structure and the distribution of conserved amino acid sequence motifs (Fig. 4).

MdTH belong to one of six subfamilies (GT-1, GT-2, SH4, GTδ, GTγ and SIP1). Five of these subfamilies are commonly found in all species for which trihelix family genes have been identified. The subfamily GTδ is not always distinguished. Individual sunflower, tomato, and rice genes have been assigned to this subfamily (Song et al. 2021; Yu et al. 2015; Li et al. 2019). The genes of the GTδ subfamily have differences in exon-intron structure and the distribution of conserved amino acid sequence motifs relative to the genes of other families. They have more introns, and characteristic conserved motifs are located at the C-terminus of the amino acid sequence (Fig. 4) (Song et al. 2021).

Other subfamily members also demonstrate similarities in the arrangement of conserved amino acid sequence motifs (Figure 4). The distribution of conservative motifs in the amino acid sequence of genes correlates with phylogenetic analysis results. This tendency is characteristic of the transcription factors of the trihelix family found in other species (Xiao et al. 2019; Liu et al. 2020; Wang et al. 2017; Li et al. 2019; Ma et al. 2019; Song et al. 2016; Song et al. 2021). It is possible that the genes of a similar structure perform similar functions.

Trihelix family genes have a different exon-intron structure. In the study presented here, both the genes that do not contain introns and the genes with the number of introns from 1 to 16 were found in the apple tree. Members of the GTγ subfamily have no introns. They were not found in the genes of this family and in the genome of wheat, pineapple, *Brachypodium distachyon*, and tomato (Xiao et al. 2019; Wang et al. 2022; Wang et al. 2019; Yu et al. 2015). The maximum number of introns is found in the representatives of the GTδ subfamily in the apple tree, sunflower, tomato, and rice (Yu et al. 2015; Li et al. 2019; Song et al. 2021). The similarity of the exon-intron organization of genes encoding the transcription factors of the trihelix family combined in one group was also noted when studying these genes in other plants, e.g. the buckwheat *Fagopyrum tataricum* (Ma et al. 2019). Similar to the apple tree, the number of introns in this plant ranged from 1 to 16, and most intronless genes were the representatives of GTγ and SIP1.

Analysis of the promoter regions of MdTH genes showed that their expression may be regulated by light. The regulatory region of each gene contains from 2 to 10 types of elements involved in light-induced signaling cascades. Induction of the expression of MdTH genes may be triggered by various phytohormones: abscisic acid, gibberellin, salicylic acid, ethylene, and methyl jasmonate. Certain elements associated with signaling cascades regulated by hormones were found in the promoter regions of each gene. Plant hormones are involved in the regulation of many processes, e.g. abscisic acid is involved, through transcriptional and post-transcriptional mechanisms, in the regulation of plant growth and development, its response to stress, leaf aging and bud germination, seed germination, stomatal closure, and many other processes (Chen et al. 2020). Gibberellic acid stimulates various trigger transitions in plant ontogenesis, e.g. the transition of seeds from dormancy to germination, meristems to shoot growth, the transition of leaves from their juvenile to the mature stage, from the vegetative stage to flowering, and it also participates in signal cascades triggered by changes in light, temperature, water

regimes, etc. (Gupta and Chakrabarty 2013). Salicylic acid is an activator of protective reactions in plants and their growth regulator. It is involved in the regulation of their immunity, resistance to biotic and abiotic stresses, and the interaction with the soil microbiome (Koo et al. 2020). Methyl jasmonate, jasmonic acid, and its amino acid conjugates, collectively referred to as jasmonates, are important cellular regulators mediating various developmental processes, including root growth, pollen production, and plant resistance to insects and pathogens (Creelman and Mullet 1997; Kessler and Baldwin 2002). For example, when *Nicotiana attuata* is eaten by herbivores, methyl jasmonates are involved in the regulation of the synthesis of volatile substances that attract natural enemies of herbivores (Kessler and Baldwin 2001; Mattiacci et al. 1995) and secondary metabolites that act as direct means of defense, e.g. nicotine neurotoxin (Baldwin 1999). Ethylene influences various processes of plant development, e.g. seed germination, ripening, aging, and fruit drop, as well as a response to various types of stress (flooding, salinization, and soil compaction) (Iqbal et al. 2017; Binder 2020). The presence of regulatory elements for hormonal regulation in the promoter regions of MdTH genes may indicate that they may also be involved in such processes.

In the promoter region of each of the MdTH genes, at least one element was found indicating their expression when exposed to stress, e.g. anaerobic stress, drought, low temperature, pathogen attack, and injury.

Thus, it may be expected that the expression of MdTH genes will differ in different organs of a plant and may vary depending on the stage of ontogenesis, as well as changes in external environmental factors: light, temperature regime, oxygen and salt levels in the soil, pathogen or herbivorous animal attacks.

Being immobile organisms, plants cannot escape the impact of abiotic stress. In the process of plant adaptation to stress, genes play an important role enhancing their resistance. Previously, a response of genes encoding trihelix proteins to stress factors, such as overwetting, salinity, and other adverse effects, as well as treatment with hormones involved in a response to stress, e.g. abscisic acid, was shown (Fang et al. 2010; Li et al. 2019; Murata et al. 2002; Xie et al. 2009; Liu et al. 2020). In our work, we demonstrated that drought, salinity, high and low temperatures affect the expression of genes of the MdTH. Among the genes studied, the greatest increase in the expression level under soil salinity was observed in MdTH4 and MdTH24 genes belonging to the SH4 and SIP1 cluster respectively. The MdTH4 gene also demonstrated the greatest increase in the relative expression level in drought. With decreasing temperature, the greatest change in the expression level was observed in MdTH30 belonging to the GT-2 cluster. Notably is that the same gene was among the three whose expression level significantly decreased when exposed to drought. With increasing temperature, the expression level of three genes demonstrated the highest jump: MdTH8, MdTH20, and MdTH36 belonging to the clusters GTγ, GT-2 and GT-1 respectively.

Thus, the representatives of different subfamilies of MdTH may respond to stress conditions. An increase in the relative expression level of MdTH4, MdTH24, MdTH30, MdTH8, MdTH20, and MdTH36 genes

under stress conditions may suggest that they are involved in the formation of a response to corresponding types of abiotic stress.

Declarations

This work was supported by Grant 2.1.5 of the NASB Research Program "Biotechnology-2"

Author Contributions

All authors contributed to the study conception and design. Material preparation and data collection were performed by Polina Kuzmitskaya and Ekaterina Koroleva. Data analysis was performed by Polina Kuzmitskaya and Oksana Urbanovich. The first draft of the manuscript was written by Polina Kuzmitskaya and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Conflict of interest

The authors declare no competing interests

References

1. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25 (17):3389-3402. https://doi.org/10.1093/nar/25.17.3389
2. Baldwin IT (1999) Inducible nicotine production in native *Nicotiana* as an example of adaptive phenotypic plasticity. Journal of Chemical Ecology 25 (1):3-30. https://doi.org/10.1023/A:1020880931488
3. Binder BM (2020) Ethylene signaling in plants. The Journal of biological chemistry 295 (22):7710-7725. https://doi.org/10.1074/jbc.REV120.010854
4. Brewer PB, Howles PA, Dorian K, Griffith ME, Ishida T, Kaplan-Levy RN, Kilinc A, Smyth DR (2004) PETAL LOSS, a trihelix transcription factor gene, regulates perianth architecture in the Arabidopsis flower. Development 131 (16):4035-4045. https://doi.org/10.1242/dev.01279
5. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem. 2009 (4):611-622. https://doi.org/10.1373/clinchem.2008.112797
6. Chen K, Li GJ, Bressan RA, Song CP, Zhu JK, Zhao Y (2020) Abscisic acid dynamics, signaling, and functions in plants. Journal of integrative plant biology 62 (1):25-54. https://doi.org/10.1111/jipb.12899
7. Creelman RA, Mullet JE (1997) biosynthesis and action of jasmonates in plants. Annual review of plant physiology and plant molecular biology 48:355-381.

https://doi.org/10.1146/annurev.arplant.48.1.355

8. Dai X, Zhuang Z, Zhao PX (2018) psRNATarget: a plant small RNA target analysis server (2017 release). Nucleic Acids Res 46 (W1):W49-W54. https://doi.org/10.1093/nar/gky316
9. Darzentas N (2010) Circoletto: visualizing sequence similarity with Circos. Bioinformatics (Oxford, England) 26 (20):2620-2621. https://doi.org/10.1093/bioinformatics/btq484
10. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer EL L, Hirsh L, Paladin L, Piovesan D, Tosatto SC E, Finn RD (2018) The Pfam protein families database in 2019. Nucleic Acids Res 47 (D1):D427-D432. https://doi.org/10.1093/nar/gky995
11. Eldem V, Okay S, Ünver T (2013) Plant microRNAs: new players in functional genomics. Turkish Journal of Agriculture and Forestry 37 (1):1-21. https://doi.org/10.3906/tar-1206-50
12. Fang Y, Xie K, Hou X, Hu H, Xiong L (2010) Systematic analysis of GT factor family of rice reveals a novel subfamily involved in stress responses. Molecular genetics and genomics : MGG 283 (2):157-169. https://doi.org/10.1007/s00438-009-0507-x
13. Gao MJ, Lydiate DJ, Li X, Lui H, Gjetvaj B, Hegedus DD, Rozwadowski K (2009) Repression of seed maturation genes by a trihelix transcriptional repressor in Arabidopsis seedlings. The Plant cell 21 (1):54-71. https://doi.org/10.1105/tpc.108.061309
14. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res 31 (13):3784-3788. https://doi.org/10.1093/nar/gkg563
15. Green PJ, Kay SA, Chua NH (1987) Sequence-specific interactions of a pea nuclear factor with light-responsive elements upstream of the rbcS-3A gene. The EMBO journal 6 (9):2543-2549. https://doi.org/10.1002/j.1460-2075.1987.tb02542.x
16. Gupta R, Chakrabarty SK (2013) Gibberellic acid in plant: still a mystery unresolved. Plant Signal Behav 8 (9):e25504. https://doi.org/10.4161/psb.25504
17. Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G (2014) GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics 31 (8):1296-1297. https://doi.org/10.1093/bioinformatics/btu817
18. Iqbal N, Khan NA, Ferrante A, Trivellini A, Francini A, Khan MIR (2017) Ethylene role in plant growth, development and senescence: interaction with other phytohormones. Frontiers in Plant Science 8. https://doi.org/10.3389/fpls.2017.00475
19. Jaakola L, Pirttilä AM, Halonen M, Hohtola A (2001) Isolation of high quality RNA from bilberry (*Vaccinium myrtillus* L.) fruit. Molecular biotechnology 19 (2):201-203. https://doi.org/10.1385/MB:19:2:201. PMID: 11725489.
20. Kaplan-Levy RN, Brewer PB, Quon T, Smyth DR (2012) The trihelix family of transcription factors–light, stress and development. Trends in plant science 17 (3):163-171. https://doi.org/10.1016/j.tplants.2011.12.002
21. Kessler A, Baldwin IT (2001) Defensive function of herbivore-induced plant volatile emissions in nature. Science (New York, NY) 291 (5511):2141-2144.

https://doi.org/10.1126/science.291.5511.2141

22. Kessler A, Baldwin IT (2002) Plant responses to insect herbivory: the emerging molecular analysis. Annual review of plant biology 53:299-328. https://doi.org/10.1146/annurev.arplant.53.100301.135207
23. Koo YM, Heo AY, Choi HW (2020) Salicylic Acid as a Safe Plant Protector and Growth Regulator. Plant Pathol J 36 (1):1-10. https://doi.org/10.5423/PPJ.RW.12.2019.0295
24. Krzywinski MI, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: An information aesthetic for comparative genomics. Genome Research. https://doi.org/10.1101/gr.092759.109
25. 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 (1):325-327. https://doi.org/10.1093/nar/30.1.325
26. Letunic I, Khedkar S, Bork P (2020) SMART: recent updates, new developments and status in 2020. Nucleic Acids Res 49 (D1):D458-D460. https://doi.org/10.1093/nar/gkaa937
27. Li B, Jiang S, Yu X, Cheng C, Chen S, Cheng Y, Yuan JS, Jiang D, He P, Shan L (2015) Phosphorylation of trihelix transcriptional repressor ASR3 by MAP KINASE4 negatively regulates Arabidopsis immunity. The Plant cell 27 (3):839-856. https://doi.org/10.1105/tpc.114.134809
28. Li C, Zhou A, Sang T (2006) Rice domestication by reducing shattering. Science (New York, NY) 311 (5769):1936-1939. https://doi.org/10.1126/science.1123604
29. Li J, Zhang M, Sun J, Mao X, Wang J, Wang J, Liu H, Zheng H, Zhen Z, Zhao H, Zou D (2019) Genome-wide characterization and identification of trihelix transcription factor and expression profiling in response to abiotic stresses in rice (*Oryza sativa* L.). Int J Mol Sci 20 (2):251. https://doi.org/10.3390/ijms20020251
30. Liu W, Zhang Y, Li W, Lin Y, Wang C, Xu R, Zhang L (2020) Genome-wide characterization and expression analysis of soybean trihelix gene family. PeerJ 8:e8753-e8753. https://doi.org/10.7717/peerj.8753
31. Ma Z, Liu M, Sun W, Huang L, Wu Q, Bu T, Li C, Chen H (2019) Genome-wide identification and expression analysis of the trihelix transcription factor family in tartary buckwheat (*Fagopyrum tataricum*). BMC Plant Biol 19 (1):344-344. https://doi.org/10.1186/s12870-019-1957-x
32. Mattiacci L, Dicke M, Posthumus MA (1995) beta-Glucosidase: an elicitor of herbivore-induced plant odor that attracts host-searching parasitic wasps. Proceedings of the National Academy of Sciences of the United States of America 92 (6):2036-2040. https://doi.org/10.1073/pnas.92.6.2036
33. Murata J, Takase H, Hiratsuka K (2002) Characterization of a novel GT-box binding protein from *Arabidopsis*. Plant Biotechnology 19 (2):103-112. https://doi.org/10.5511/plantbiotechnology.19.103
34. Nagano Y (2000) Several features of the GT-factor trihelix domain resemble those of the Myb DNA-binding domain. Plant Physiol 124 (2):491-494. https://doi.org/10.1104/pp.124.2.491

35. Qin Y, Ma X, Yu G, Wang Q, Wang L, Kong L, Kim W, Wang HW (2014) Evolutionary history of trihelix family and their functional diversification. DNA Research 21 (5):499-510. https://doi.org/10.1093/dnares/dsu016
36. Rao X, Huang X, Zhou Z, Lin X (2013) An improvement of the 2ˆ (−delta delta CT) method for quantitative real-time polymerase chain reaction data analysis. Biostatistics, bioinformatics and biomathematics 3 (3):71
37. Song A, Wu D, Fan Q, Tian C, Chen S, Guan Z, Xin J, Zhao K, Chen F (2016) Transcriptome-wide identification and expression profiling analysis of chrysanthemum trihelix transcription factors. Int J Mol Sci 17 (2):198. https://doi.org/10.3390/ijms17020198
38. Song J, Shen WY, Shaheen S, Li YY, Liu ZR, Wang Z, Pang HB, Ahmed Z (2021) Genome-wide identification and analysis of the trihelix transcription factors in sunflower. Biologia plantarum 65 (1):80-87. https://doi.org/10.32615/bp.2021.006
39. Sperschneider J, Catanzariti A-M, DeBoer K, Petre B, Gardiner DM, Singh KB, Dodds PN, Taylor JM (2017) LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. Sci Rep 7 (1):44598. https://doi.org/10.1038/srep44598
40. Tamura K, Stecher G, Kumar S (2021) MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Molecular Biology and Evolution 38 (7):3022-3027. https://doi.org/10.1093/molbev/msab120
41. Unver T, Namuth-Covert DM, Budak H (2009) Review of current methodological approaches for characterizing microRNAs in plants. Int J Plant Genomics 2009:262463-262463. https://doi.org/10.1155/2009/262463
42. Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, Salvi S, Pindo M, Baldi P, Castelletti S, Cavaiuolo M, Coppola G, Costa F, Cova V, Dal Ri A, Goremykin V, Komjanc M, Longhi S, Magnago P, Malacarne G, Malnoy M, Micheletti D, Moretto M, Perazzolli M, Si-Ammour A, Vezzulli S, Zini E, Eldredge G, Fitzgerald LM, Gutin N, Lanchbury J, Macalma T, Mitchell JT, Reid J, Wardell B, Kodira C, Chen Z, Desany B, Niazi F, Palmer M, Koepke T, Jiwan D, Schaeffer S, Krishnan V, Wu C, Chu VT, King ST, Vick J, Tao Q, Mraz A, Stormo A, Stormo K, Bogden R, Ederle D, Stella A, Vecchietti A, Kater MM, Masiero S, Lasserre P, Lespinasse Y, Allan AC, Bus V, Chagné D, Crowhurst RN, Gleave AP, Lavezzo E, Fawcett JA, Proost S, Rouzé P, Sterck L, Toppo S, Lazzari B, Hellens RP, Durel C-E, Gutin A, Bumgarner RE, Gardiner SE, Skolnick M, Egholm M, Van de Peer Y, Salamini F, Viola R (2010) The genome of the domesticated apple (*Malus × domestica* Borkh.). Nature Genetics 42 (10):833-839. https://doi.org/10.1038/ng.654
43. Wang C, Wang Y, Pan Q, Chen S, Feng C, Hai J, Li H (2019) Comparison of Trihelix transcription factors between wheat and Brachypodium distachyon at genome-wide. BMC genomics 20 (1):142. https://doi.org/10.1186/s12864-019-5494-7
44. Wang D, Liu D, Yuchi J, He F, Jiang Y, Cai S, Li J, Xu D (2020) MusiteDeep: a deep-learning based webserver for protein post-translational modification site prediction and visualization. Nucleic Acids Res 48 (W1):W140-W146. https://doi.org/10.1093/nar/gkaa275

45. Wang J, Ouyang Y, Wei Y, Kou J, Zhang X, Zhang H (2022) Identification and characterization of trihelix transcription factors and expression changes during flower development in pineapple. Horticulturae 8 (10). https://doi.org/10.3390/horticulturae8100894
46. Wang W, Wu P, Liu T, Ren H, Li Y, Hou X (2017) Genome-wide analysis and expression divergence of the trihelix family in *Brassica rapa*: insight into the evolutionary patterns in plants. Sci Rep 7 (1):6463-6463. https://doi.org/10.1038/s41598-017-06935-0
47. Xiao J, Hu R, Gu T, Han J, Qiu D, Su P, Feng J, Chang J, Yang G, He G (2019) Genome-wide identification and expression profiling of trihelix gene family under abiotic stresses in wheat. BMC genomics 20 (1):287. https://doi.org/10.1186/s12864-019-5632-2
48. Xie Z-M, Zou H-F, Lei G, Wei W, Zhou Q-Y, Niu C-F, Liao Y, Tian A-G, Ma B, Zhang W-K, Zhang J-S, Chen S-Y (2009) Soybean trihelix transcription factors GmGT-2A and GmGT-2B improve plant tolerance to abiotic stresses in transgenic *Arabidopsis*. PLOS ONE 4 (9):e6898. https://doi.org/10.1371/journal.pone.0006898
49. Yoo CY, Pence HE, Jin JB, Miura K, Gosney MJ, Hasegawa PM, Mickelbart MV (2010) The Arabidopsis GTL1 transcription factor regulates water use efficiency and drought tolerance by modulating stomatal density via transrepression of SDD1. The Plant cell 22 (12):4128-4141. https://doi.org/10.1105/tpc.110.078691
50. Yu C, Cai X, Ye Z, Li H (2015) Genome-wide identification and expression profiling analysis of trihelix gene family in tomato. Biochemical and biophysical research communications 468 (4):653-659. https://doi.org/10.1016/j.bbrc.2015.11.010
51. Zhang B (2015) MicroRNA: a new target for improving plant tolerance to abiotic stress. J Exp Bot 66 (7):1749-1761. https://doi.org/10.1093/jxb/erv013
52. Zhao T, Liang D, Wang P, Liu J, Ma F (2012) Genome-wide analysis and expression profiling of the DREB transcription factor gene family in Malus under abiotic stress. Molecular genetics and genomics 287 (5):423-436. https://doi.org/ 10.1007/s00438-012-0687-7

Figures

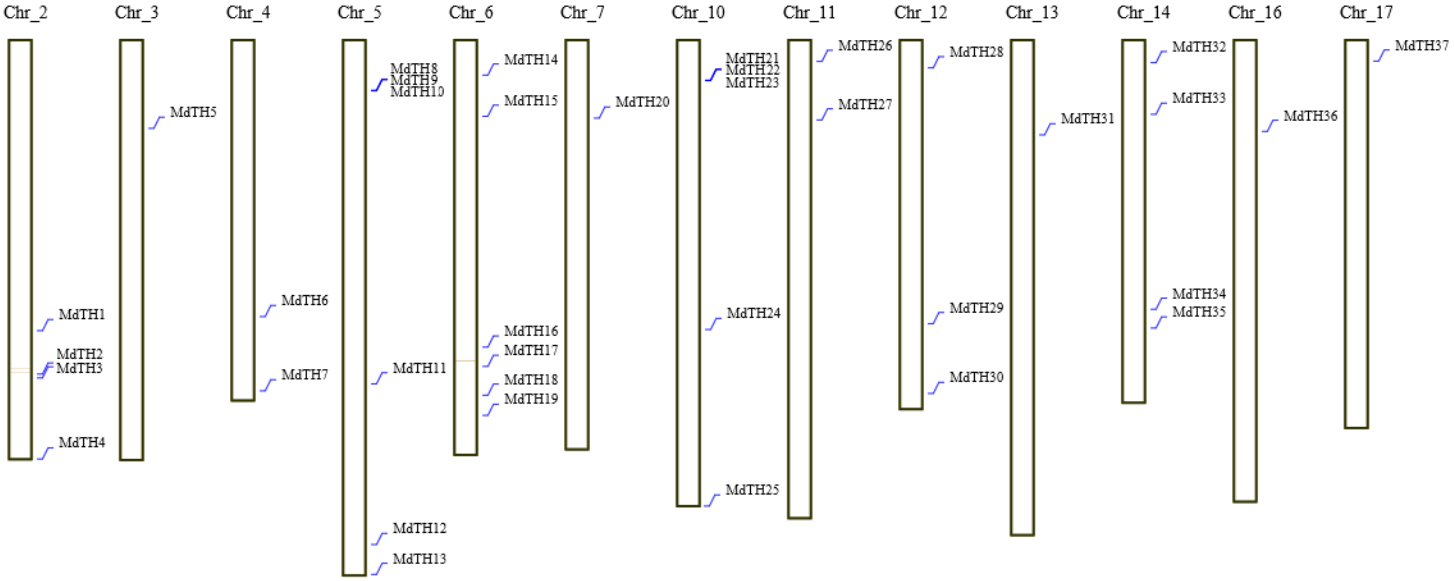

Figure 1

Distribution of trihelix genes in apple tree chromosomes

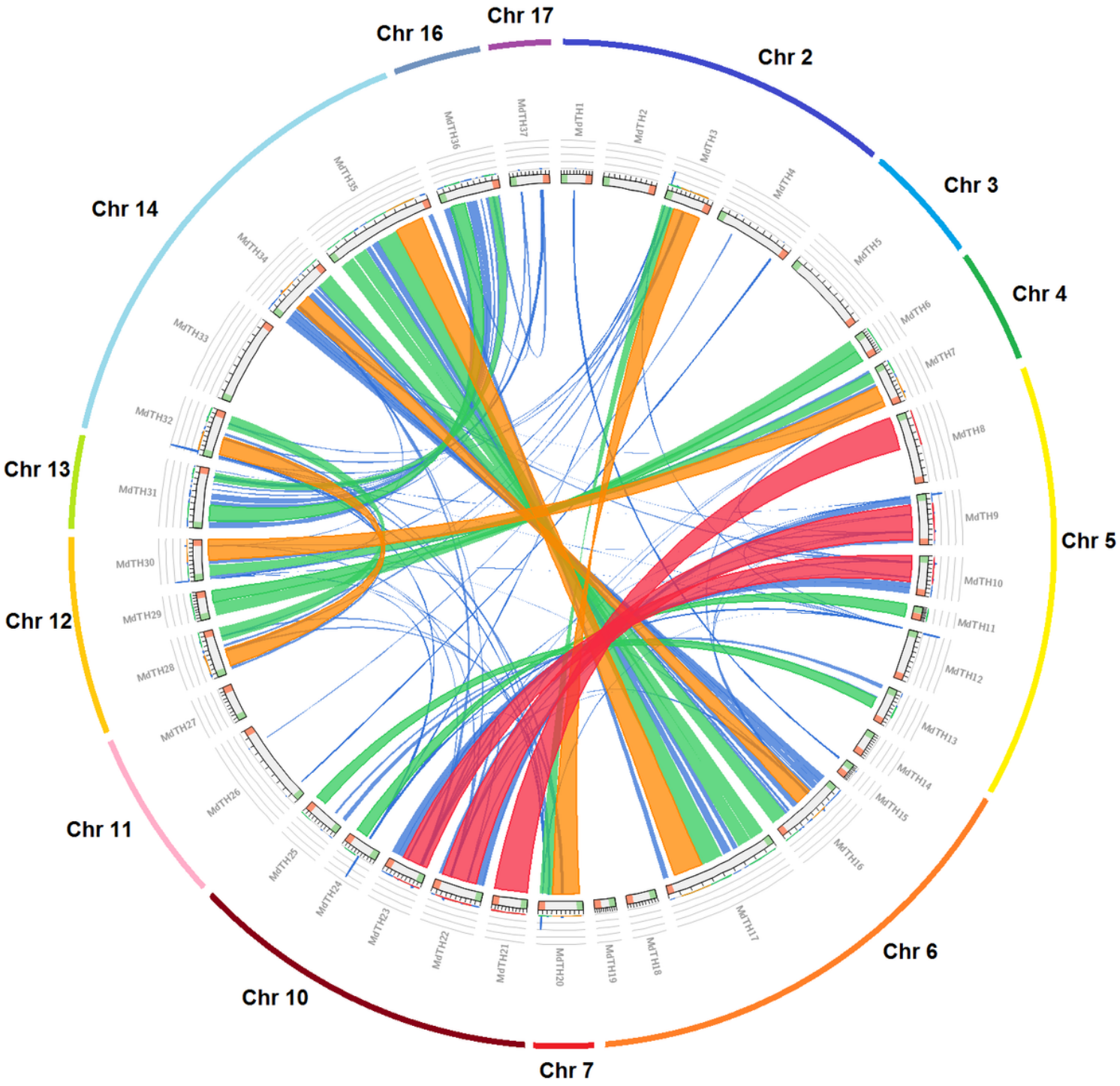

Figure 2

Schematic representation of segmental duplications of apple trihelix genes. Colored according to the score (min → max: blue → green → orange → red)

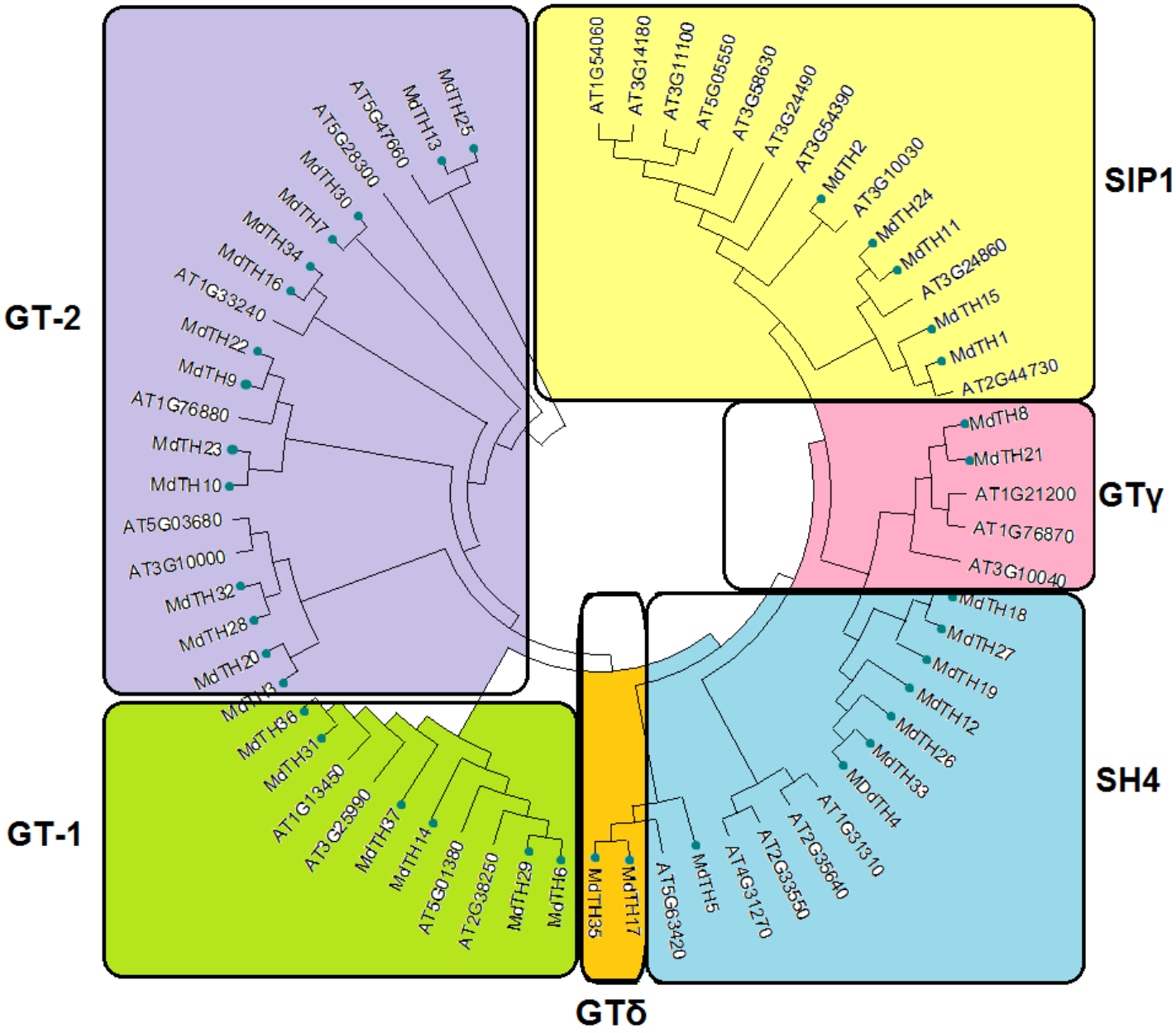

Figure 3

Phylogenetic tree of apple (Md) and *Arabidopsis thaliana* (AT) trihelix proteins

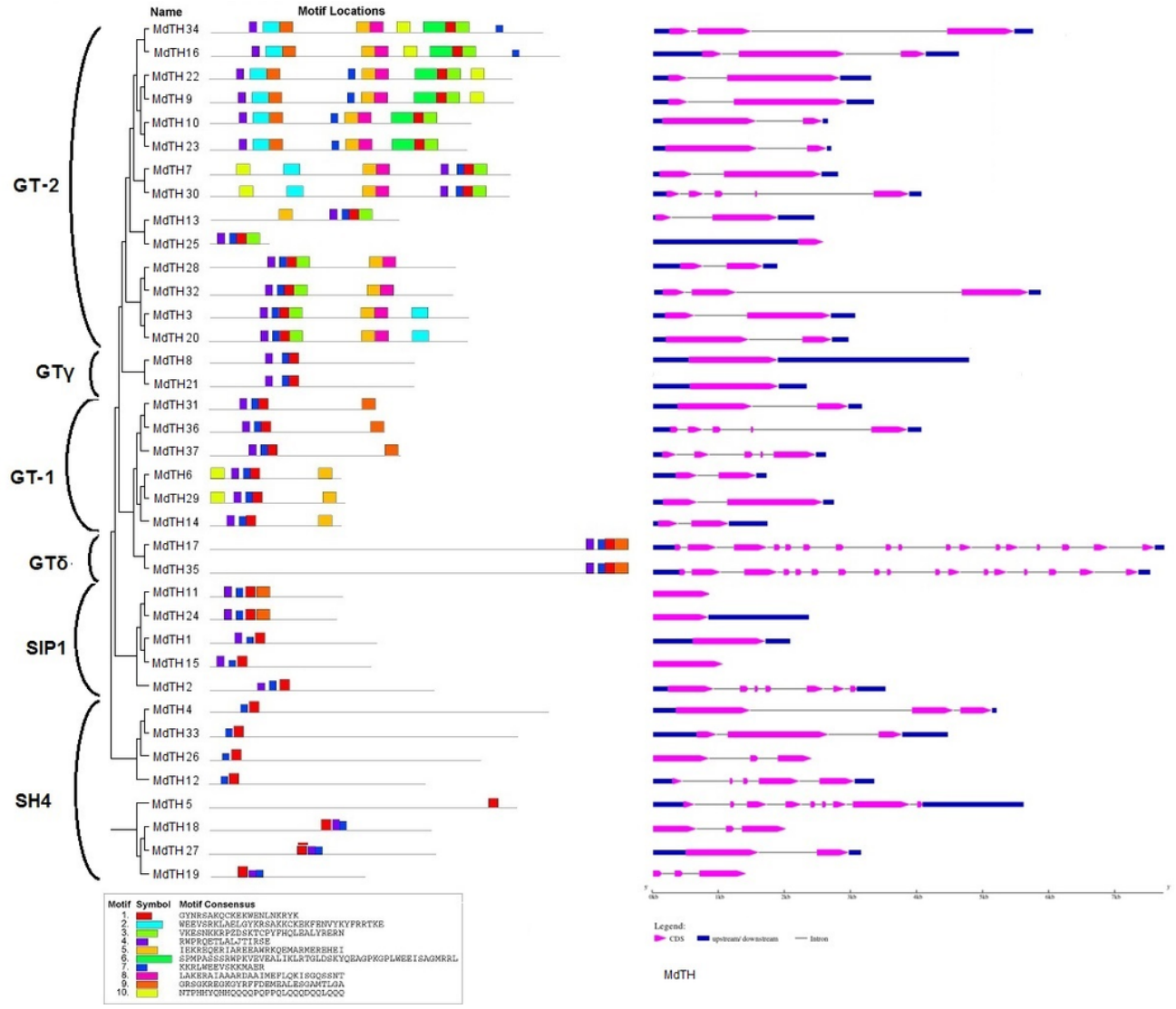

Figure 4

Apple trihelix gene structure and motifs.

A Phylogenetic relationship (left) and conserved motifs (right) of 37 trihelix proteins. MEME was used to predict motifs. Motifs were represented by the boxes of different color.

B Gene structures. Intron, exon and upstream/downstream sequences were represented by black lines, red boxes and blue boxes respectively.

Figure 5

Binding sites for the transcription factors involved in light-dependent signaling pathways found in the promoter regions of MdTH genes.

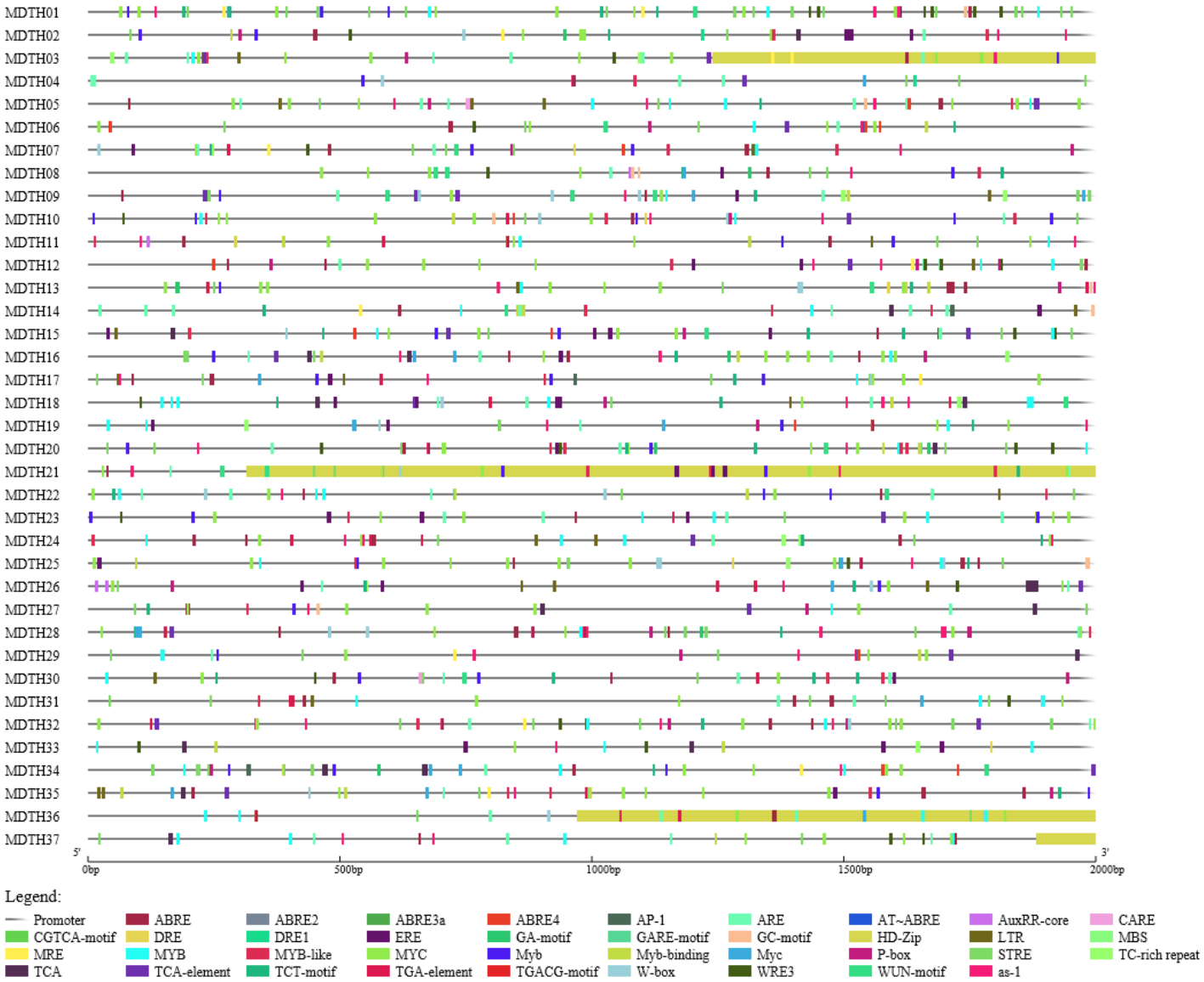

Figure 6

Binding sites of the transcription factors involved in stress- and hormone-induced signaling pathways found in the promoter regions of MdTH genes.

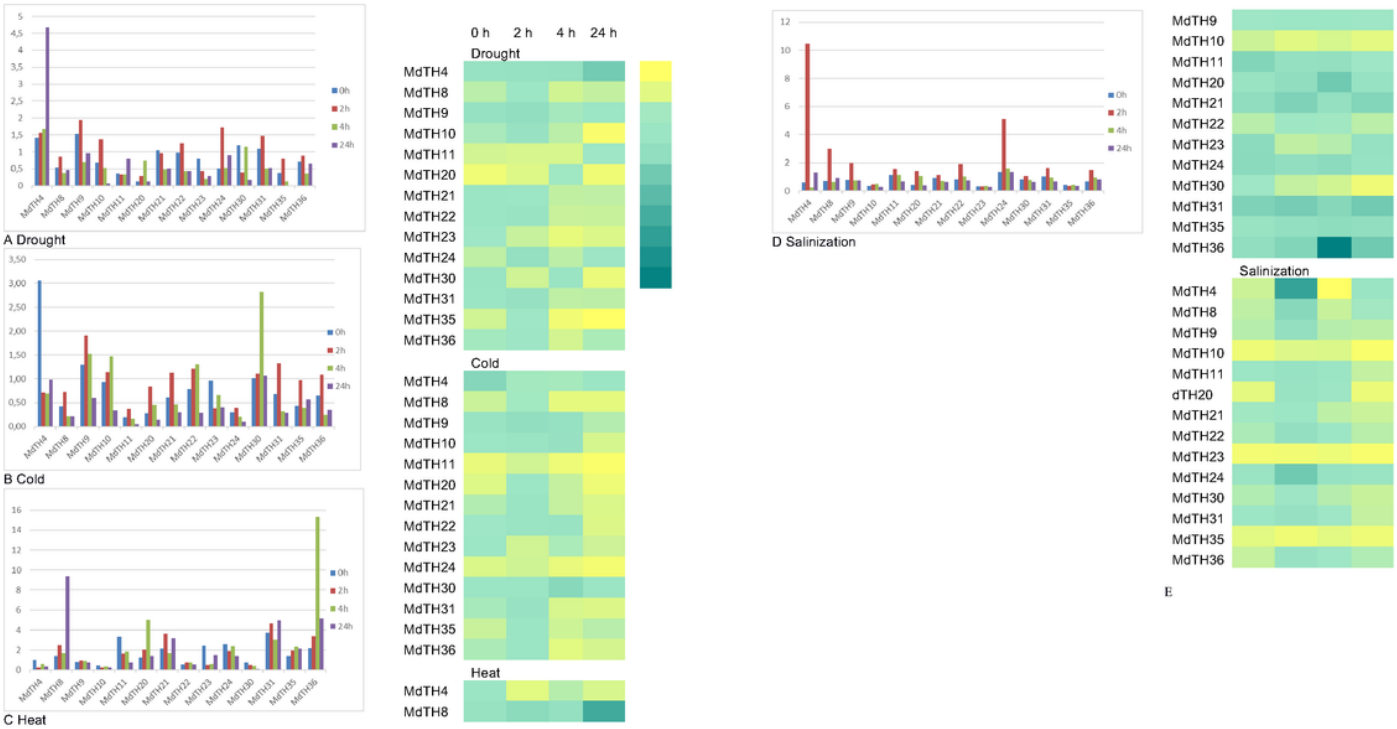

Figure 7

MdTH gene expression profiles under the condition of drought (A), low temperatures (B), high temperatures (C), salinity (D) on the MM-106 apple variety with measurement points at 0, 2nd, 4th, and 24th hour. The data were normalised in relation to the apple housekeeping gene *Ef1-α*. The vertical stripes show the standard error of the mean. The significance value is a=0.05

E Expression profiles of MdTH genes. The color scale indicates the log 2 values of transcript per million. Green and yellow colors demonstrate lower and higher expression levels respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementary3h2.docx