Journal of Genetics (2019) 98:13
https://doi.org/10.1007/s12041-018-1049-y

RESEARCH ARTICLE

Genomewide analysis of homeobox gene family in apple (*Malus domestica* Borkh.) and their response to abiotic stress

RONG LI[1], HONGJUAN GE[2], YAQING DAI[1], LI YUAN[1], XIN LIU[1], QINGHUA SUN[1]* and XIAOYUN WANG[1]*

[1] *College of Life Science, State Key Laboratory of Crop Biology, Shandong Agricultural University, Taian 271018, Shandong, People's Republic of China*
[2] *Qingdao Academy of Agricultural Science, Qingdao 266100, Shandong, People's Republic of China*
*For correspondence. E-mail: Qinghua Sun, qhsun@sdau.edu.cn; Xiaoyun Wang, xyunwang@sdau.edu.cn.

Received 7 April 2018; revised 9 August 2018; accepted 3 September 2018; published online 21 February 2019

Abstract. Homeobox proteins (HOXs) comprise a large family in eukaryotes and share a highly conserved DNA-binding motif, the homeodomain (HD). HOXs play an important role in the regulation of plant growth, development and stress response. However, systematic analysis and expression profiling of these genes have not been reported in *Malus domestica*. In this study, a total of 207 HOXs of *M. domestica* (*Md*HOXs) were identified and classified into 11 distinct subfamilies, and an unclassified group according to their functional domains. The *MdHOX*s were localized in all 17 chromosomes with various densities and a majority of them tended to form gene clusters. Analysis of the K_a/K_s ratios suggested that the duplicated genes of *MdHOX*s mainly underwent purifying selection with restrictive functional divergence after the duplication events. The expression of *MdHOX*s has organ specific characteristics and were divided into seven different groups. Stress-related *cis*-acting elements were prevalent in the upstream sequence of *MdHOX*s by systematic analysis. To explore the response to abiotic stress, eight *MdHOX*s were randomly selected to investigate their expression using quantitative real-time polymerase chain reaction. Transcription levels of *MdHOX*s were upregulated in the leaves and roots under cold, osmotic, high salinity or exogenous ABA treatments, which suggested that they may take part in the plant response to abiotic stress. These results provided basic information of HOXs in apple and will further contribute to the functional research of *MdHOX*s, especially the response to abiotic stress.

Keywords. homeobox; apple; expression pattern; stress-related *cis*-acting element analysis; abiotic stress.

Introduction

Homeobox protein (HOX) contains a conserved 60 amino acid DNA-binding motif, known as homeodomain (HD). The HD consists of three alpha-helices, which can recognize and bind to specific DNA sequences to regulate the expression of target genes. The HOX encoding genes were first identified in *Drosophila melanogaster* (Burglin and Affolter 2016). More and more *HOX*s have been subsequently identified in various organisms, such as humans, fungi, nematode and plants (Maeda and Karch 2015). So far, it was found that *HOX*s were widely distributed in all eukaryotes.

*HOX*s constitute a large family in plants. There are ~110 members in *Arabidopsis thaliana*, 107 in rice and 89 in chickpea (Jain *et al.* 2008; Mukherjee *et al.* 2009; Bhattacharjee *et al.* 2015). According to the phylogenetic relationship and domain composition, the plant HOX members were classified into different groups. Earlier, HOX members in plants were divided into seven classes, including knotted-like homeobox (KNOX), BEL1-like

Qinghua Sun and Rong Li conceived and designed this research; Rong Li, Xin Liu, Li Yuan and Yaqing Dai performed the experiments; Rong Li, Qinghua Sun and Hongjuan Ge analysed the data and Xiaoyun Wang, Qinghua Sun and Rong Li wrote the paper. All authors read and approved the manuscript.

Electronic supplementary material: The online version of this article (https://doi.org/10.1007/s12041-018-1049-y) contains supplementary material, which is available to authorized users.

homeobox (BEL), ZM-HOX, HAT1, HAT2, ATHB8 and GL2 (Mukherjee *et al.* 2009). Because the latter four groups (HAT1, HAT2, ATHB8 and GL2) were all characterized by a leucine-zipper (LZ) motif downstream of the HD, they have been renamed homeobox domain-leucine zipper (HD-ZIP) I, HD-ZIP II, HD-ZIP III and HD-ZIP IV, respectively (Katiyar *et al.* 2001; Ariel *et al.* 2007). Later, additional three groups (Wuschel-like homeobox (WOX), plant-specific zinc-finger homeobox (PLINC) and plant homeobox domain (PHD)) were classified in rice (Jain *et al.* 2008). Recently, a comprehensive analysis of plant HOXs was performed and they were divided into 14 classes, including several new classes (NDX, DDT, LD, SAWADEE and PINTOX) and an unclassified group (Chan *et al.* 1998).

As transcription factors (TFs), plant HOXs participate in various developmental processes, hormone signalling and stress response (Tsuda *et al.* 2014). For example, many members of HD-ZIP III and HD-ZIP IV subfamilies are reported to be involved in shoot meristem formation, vascular development, photomorphogenesis, flowering and fruit ripening (Liu *et al.* 2013; Chen *et al.* 2014), while the members of HD-ZIP I and HD-ZIP II are primarily associated with environmental response (Jain *et al.* 2008). The members of WOX subfamily play a key role in controlling the development of reproductive organ, such as root, shoot and floral meristem formation in rice (Cheng *et al.* 2014; Costanzo *et al.* 2014). The members of KNOX and PLINC subfamilies are reported to be implicated in the development of shoot apical meristem and floral developmental processes in *Arabidopsis* and tobacco (Tan and Irish 2006; Hake 2016). The members of BEL subfamily are known to interact with KNOX proteins to regulate various hormone homeostasis and participate in abiotic stress response in *Medicago truncatula* and potato (Sharma *et al.* 2014; Shu *et al.* 2015a, b). The members of PHD subfamily are reported to regulate the expression of certain gene encoding histone methyltransferase and participate in abiotic stress in *Populus trichocarpa* (Wu *et al.* 2016). All these findings exhibit the functional diversity of HOXs in different plants. Nevertheless, the function and regulatory mechanism of HOX genes in apple remain largely unknown.

Apple is one of the most economically important perennial fruit crops in temperate regions and the growth is seriously affected by adverse environmental conditions, such as drought, high salinity and low temperature. Discovering novel genes involved in abiotic-stress resistance and application of molecular breeding are considered as an attractive method to improve the resistance to stress. Although some HOX encoding genes in apple have been reported in the Plant Transcription Factor Database v4.0 (PlantTFDB, http://planttfdb.cbi.pku.edu.cn/index.php?sp=Mdo), only four groups (HD-ZIP, WOX, PLINC and other HD) have been classified, which is much lower than that of plant HOX family with about 14 groups. The sequencing of the whole genome of *M. domestica* makes genomic analyses available (Velasco *et al.* 2010). In the present study, genomewide screen of apple HOX gene family was performed and more members have been identified. The functional domains, gene structure, chromosomal localization, promoter analysis and organ-specific expression patterns of *MdHOX*s were analysed. Moreover, eight *MdHOX*s were randomly selected to investigate their expression patterns under different abiotic stress. The results present more useful information about HOX genes.

Materials and methods

Identification of HOXs in apple

To identify the HOX encoding genes in apple (*M. domestica* Borkh.), two approaches were implemented in this study. In the first method, the sequences of HOX protein previously identified from *Arabidopsis* and rice were used as queries to perform a local BLASTp against the complete apple genomics database (genome v1.0, http://www.rosaceae.org/) by the BioEdit program with an *E*-value cutoff of 0.001 (Cui *et al.* 2015). In the second method, the hidden Markov model (HMM) profile of homeobox domain (Pfam number: PF00046) obtained from the Pfam website (http://pfam.xfam.org/) was used as a query to search against the above apple database with HMMER 3.0 (Eddy 1998). All the potential candidates of HOXs obtained by the two methods were then submitted to the InterPro Database (http://www.ebi.ac.uk/interpro/) and SMART database (http://smart.embl-heidelberg.de/) respectively, to confirm the presence of HD domains. To obtain a more complete HOX family of *M. domestica*, all members identified by the two above approaches were compared with the confirmed *Md*HOXs in the database of PlantTFDB. The redundant sequences with the same gene ID were removed. The sequence logos of HD domains were generated using the MEME website database (http://meme-suite.org). Phosphorylation sites of *Md*HOXs were predicted by Netphos 2.0 serve (http://www.cbs.dtu.dk/services/NetPhos-2.0/).

Identification of additional domains other than HD in MdHOXs and analysis of the structure of their encoding genes

The full-length sequences of *Md*HOXs were submitted to the InterPro website (http://www.ebi.ac.uk/interpro/) to search any additional domains other than the HD domain and they were divided into different groups according to their functional domains, combined with the similarity of the sequences of HD. To explore the diversity of intron–exon organization, the gene structures of *MdHOX*s were determined by submitting the predicted coding sequences and the corresponding genomic sequences to the gene

structure display server (GSDS, http://gsds.cbi.pku.edu.cn/) software (Hu *et al.* 2012; Xu *et al.* 2016a, b).

Chromosomal localization of MdHOXs and gene duplication analysis

Chromosomal location data of *MdHOX*s were retrieved from apple genome annotations downloaded from the Genome Database for *Rosaceae* (GDR, http://www.rosaceae.org/) and then mapped to the chromosomes using software MapDraw (Xu *et al.* 2016a, b; Liu *et al.* 2017). Tandem duplications were checked, if the distance between two adjacent genes were within 30 kb and their similarity was more than 70% (Du *et al.* 2013). Segmental duplications were examined using SyMAP (Synteny Mapping and Analysis, http://www.symapdb.org/) database. Before chromosomal localization, we tried to check whether alternative splicing forms from one gene locus exist using the *M. domestica* Genome v1.0 genome annotation download from *Rosaceae* database.

To explore the mechanism of gene divergence after duplication, the K_a/K_s ratio between paralogues was calculated to show the positive selection (with the ratio >1), negative or purifying selection (the ratio <1) and neutral selection (the ratio=1), respectively. While, the K_a (nonsynonymous nucleotide substitutions) and K_s (synonymous nucleotide substitutions) were analysed using the K-Estimator 6.1 V program. To examine the occurrence time of duplication events, K_s value was used to calculate the duplication time as $T = K_s/2\lambda \times 10^{-6}$ million years ago (Mya) ($\lambda = 6.5 \times 10^{-9}$) (Quraishi *et al.* 2011).

Organ-specific expression patterns of MdHOXs

To explore the expression patterns of *MdHOX*s in different organs, the expression profile of apple (GSE42873) was obtained from the Gene Expression Omnibus database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE42873). This profile (GSE42873) contained the expression levels of eight apple organs, and the gene names in GSE42873 were consistent with the GDR database. According to their gene names, the expression data of the identified *MdHOX*s were extracted from the two datasets (Cui *et al.* 2015). The hierarchical clustering was performed with software Cluster 3.0 (http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm) (Larkin *et al.* 2007). The heatmap and clustering tree were constructed and shown with software Java Treeview (http://sourceforge.net/projects/jtreeview/?source=typ_redirect). Organ-specific gene expression was also checked by quantitative real-time polymerase chain reaction (qRT-PCR) using the samples from 6-year-old trees of apple growing at the experiment station of Shandong Agricultural University (Taian, Shandong, China).

Searching for cis-acting elements in the upstream sequence of MdHOXs

The promoter sequences of *MdHOX*s, ~1500 bp of genomic DNA sequences upstream of the transcriptional start site (ATG) were obtained from the apple genome. Further these sequences were submitted to the PlantCARE website (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) to investigate the presence of various *cis*-acting elements, respectively.

Plant growth, stress treatments and total RNA extraction

To investigate the response of *MdHOX*s under various types of abiotic stress, 1 year seedlings of Golden Delicious growing in Hoagland solution at normal conditions (25 ± 1°C, 16-h light / 8-h dark) were used for stress treatment as done in similar investigations. Exogenous ABA (100 μM) was directly sprayed on the whole seedlings, while 20% PEG and 250 mM NaCl were added to the Hoagland nutrient solution, respectively (Zhao *et al.* 2012; Liu *et al.* 2017). The concentrations of ABA, NaCl and PEG were selected based on the similar investigations. For chilling treatment, the whole seedlings were placed in a 4°C environment (Xu *et al.* 2016a, b). Each sample was from three different seedlings and collected at 1, 3, 6, 9 and 12 h, then frozen into liquid nitrogen and stored at −70°C till use. The total RNA was isolated using the CTAB procedure (Gasic *et al.* 2004). The RNA concentrations and quality were determined by a NanoDrop Spectrometer (ND-1000 Spectrophotometer, Peqlab).

Quantitative real-time PCR (qRT-PCR) analysis

cDNA fragments were synthesized from total RNA using the 5× All-In-One RT MasterMix with AccuRT Genomic DNA Removal kit (Applied Biological Materials (abm) Inc., Vancouver, Canada). To ensure the results of qRT-PCR were credible, DREB gene (MDP0000147009) was selected as positive control, whose expression has been demonstrated to be upregulated under stress treatments (Zhao *et al.* 2012). The apple *actin* gene (GeneID: 103453508) was selected as an internal standard and the relative expression level of each gene was calculated according to the $2^{-\Delta\Delta CT}$ method (Xu *et al.* 2016a, b). The gene-specific primers were designed based on target gene sequences using the software Beacon Designers 7.91, and the sequences of primer pairs are listed in table 1 in electronic supplementary material at http://www.ias.ac.in/jgenet/. The relative expression levels of *MdHOX*s in stressed samples were compared to the controls (0 h) with Student's *t*-test at significance levels of $*P<0.05$ and $**P<0.01$. Each reaction was carried out in three biological replicates. This article does not contain any studies on animals performed by any of the authors.

Table 1. Distribution of the stress responsive *cis*-acting elements in 11 subfamilies of *MdHOX*s.

Stress responsive *cis*-acting elements	Subfamily (numbers of *MdHOX*s containing the relative *cis*-acting elements)											
	HD-ZIP I (42)	HD-ZIP II (11)	HD-ZIP III (8)	HD-ZIP IV (20)	BEL (23)	PHD (4)	KNOX (23)	WOX (23)	SAWADDE (2)	PLINC (33)	DDT (3)	Total (192)
MBS	27	11	6	17	17	3	16	17	2	24	3	143
CGTCA-motif	24	11	6	14	12	4	16	19	2	23	3	134
TGACG-motif	28	9	5	14	13	4	15	18	2	23	3	134
TCA-element	32	9	4	14	16	2	14	16	2	21	2	132
TC-rich repeats	23	10	5	11	14	1	13	19	1	25	3	131
HSE	29	8	5	12	14	3	16	16	2	25	1	131
ABRE	21	6	5	8	14	2	14	15	2	22	2	111
GARE-motif	20	10	7	11	13	4	9	13	1	19	3	110
TGA-element	31	7	5	3	7	2	6	14	1	16	0	92
LTR	20	3	2	7	8	1	9	15	1	15	1	82
ERE	16	3	3	4	6	1	6	11	1	14	1	66
WUN	12	3	1	2	1	0	4	10	1	12	0	46
TATC-box	19	0	0	3	2	1	2	9	1	0	2	39

MBS, MYB binding site involved in drought-inducibility; CGTCA-motif, *cis*-acting regulatory element involved in the methyl jasmonate-responsiveness; TGACG-motif, *cis*-acting regulatory element involved in the methyl jasmonate-responsiveness; TCA-element, *cis*-acting element involved in salicylic acid responsiveness; TC-rich repeats, *cis*-acting element involved in defense and stress responsiveness; HSE, *cis*-acting element involved in heat stress responsiveness; ABRE, *cis*-acting element involved in the abscisic acid responsiveness; GARE-motif, gibberellin-responsive element; TGA-element, auxin-responsive element; LTR, *cis*-acting element involved in low-temperature responsiveness; ERE, ethylene-responsive element; WUN, wound-responsive element; TATC-box, *cis*-acting element involved in gibberellin-responsiveness.

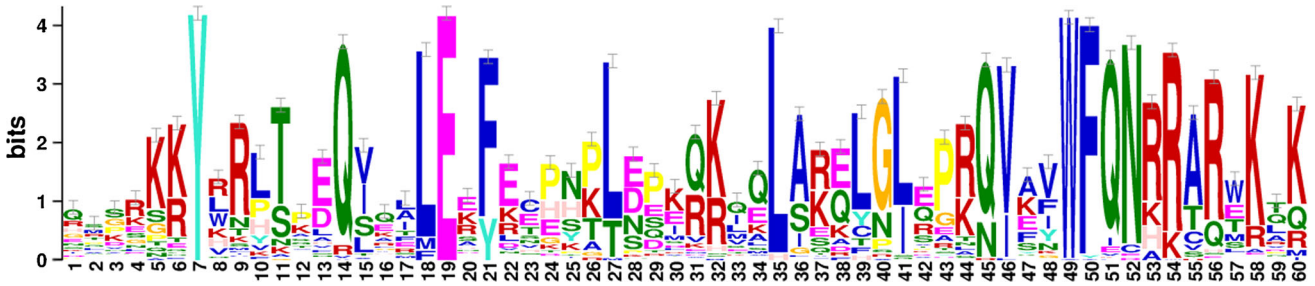

Figure 1. Sequence logos of HD domains generated from all *Md*HOXs. The *x*-axis represents the relative positions of the HD motifs. The *y*-axis represents the information content as measured in bits.

Subfamily	Numbers	Representative	Structure
HD-ZIP I	42	*Md*HOX19	HD LZ
HD-ZIP II	11	*Md*HOX24	ZIBEL HD LZ
HD-ZIP III	8	*Md*HOX9	HD LZ START MEKHLA
HD-ZIP IV	20	*Md*HOX103	HD LZ START
KNOX	23	*Md*HOX166	KNOX1 KNOX2 ELK HD
BEL	23	*Md*HOX122	BELa BELb HD
PLINC	33	*Md*HOX18	PLINC HD
PHD	4	*Md*HOX13	PHD HD
SAWADEE	2	*Md*HOX35	HD SAWADEE GroES-like
WOX	23	*Md*HOX17	HD WUS BOX
DDT	3	*Md*HOX45	HD DDT WHIM-1 WHIM-2
UNCLASSIED	15	*Md*HOX48	HD
Total	207		

Figure 2. Structure and domain composition of representative *Md*HOXs from each subfamily.

Result

Identification, classification and gene structure analysis of MdHOXs in apple

To obtain the credible results without missing the potential candidates of the genes that encode HOX, the *E*-value cutoff was set at 0.001 similar to the previous studies (Cui *et al.* 2015; Xu *et al.* 2016a, b). Using this method, ~1500 putative HOX sequences were identified. The HMM Profile of HD (PF00046) from plants and animals was used as a query to search against the complete apple genome and about 200 putative HOX sequences were identified. All the potential candidates obtained by the above two approaches were further checked for the presence of HD domains and then compared with the confirmed *Md*HOXs in PlantTFDB (http://planttfdb.cbi.pku.edu.cn/index.php?sp=Mdo). Finally, 207 genes encoding HOXs with HD domain were confirmed as *Md*HOXs and named according to their chromosomal distributions. The detailed information of each *MdHOX* is shown in table 2 in electronic supplementary material at http://www.ias.ac.in/jgenet/, including the gene name, gene ID, chromosome location, predicted phosphorylation sites and the length of protein sequence. The logo of HD domain of *Md*HOXs with 60 amino acids was constructed and shown in figure 1.

In addition to HD domain, most of the members contained diverse characteristic domains, while fewer than 10 of them had only HD domain. Based on the presence of characteristic domains in the confirmed HOXs from other plants (Angela and Miltos 2010), all the identified members were categorized into 11 known subfamilies and one 'unclassified' group with unknown functional domain (figure 2). A total of 81 *Md*HOXs were grouped into the HD-ZIP I-IV subfamilies by the presence of a LZ domain downstream HD. Members in HD-ZIP I had no other functional domains. Compared with the subfamily of HD-ZIP I, the HD-ZIP II contained an additional conserved ZIBEL domain, by which domain, the members in HD-ZIP II can interact with the target HOXs containing BEL domain. HD-ZIP III and HD-ZIP IV were characterized by the steroidogenic acute regulatory protein-related lipid transfer (START) domain,

with HD-ZIP III having an extra MEKHLA domain which is speculated to be involved in reaction of oxygen redox.

The KNOX subfamily contained additional domains of KNOX1, KNOX2 and ELK, which were reported to be involved in nuclear localization, suppressing target gene expression and homo-dimerization, respectively (Chan *et al.* 1998). The subfamily of PLINC harboured 33 *Md*HOXs, characterized by a conserved PLINC domain, a zinc-finger-like domain upstream of HD. PLINC domain may participate in homodimerization and heterodimerization (Thomas *et al.* 2015). The BEL subfamily contained two conserved domains (BEL_a and BEL_b) and PHD subfamily was a well-characterized zinc-finger resembling a RING domain upstream of HD. WOX subfamily contained a WUS box domain and the DDT subfamily harboured a DDT domain downstream of the HD. Only two *Md*HOXs were grouped in the SAWADEE subfamily by the SAWADEE domain, with a site of metal coordination by conserved cysteine and histidine residues (Doerks *et al.* 2001).

The intron–exon structure of encoding genes of *MdHOX*s were further analysed by the software GSDS (http://gsds.cbi.pku.edu.cn/). Among 207 *MdHOX*s, 24 members contained no introns and 179 contained 1–30 introns (table 2 in electronic supplementary material). The *MdHOX*s within the same group displayed similar intron–exon structures in intron number and distribution (figure 1 in electronic supplementary material). However, the number of introns in each group varied. For example, a majority of members in HD-ZIP II subfamily contained fewer than three introns, while most genes in HD-ZIPIII subfamily harboured more than 10 introns (table 2 in electronic supplementary material).

Chromosomal distribution and gene duplication of MdHOXs

For the genomic distribution of *MdHOX*s, total 201 of 207 *MdHOX* genes were successfully mapped to all 17 apple chromosomes (figure 3). Only six *MdHOX*s failed to be mapped to any chromosomes, which could be distributed in the scaffolds of unsuccessful assembled genomic sequences like another family (Li *et al.* 2011). The number of *MdHOX*s on different chromosomes was uneven. Chromosome 15 contained the most with 20 *MdHOX* genes and chromosome 4 harboured the least with only three *MdHOX*s. In most chromosomes, some *MdHOX*s tend to form clusters and showed high similarity within the clusters. When the similarity of two genes was more than 70% and the distance between them was within 30 kb distance, it was defined as tandem duplication. About one-third *MdHOX*s belonged to tandem duplication that is marked in green colour in figure 3, which represented 28 individual duplication events. In addition to tandem duplication, the genes with high similarity also appeared in different chromosomes, which suggested that segmental duplication has occurred (table 3 in electronic supplementary material). Total 15 pairs of *MdHOX*s were involved in the segmental duplications, and the highest frequency of segmental duplication events occurred between chromosomes 9 and 17 (table 3 in electronic supplementary material). All the above results suggested that tandem gene duplication and segmental gene duplication resulted in the expansion of *MdHOX*s. Besides, before determining the genomic distribution of the *MdHOX*s, the *M. domestica* Genome v1.0p annotation was checked and no alternative splicing forms from the same gene locus were found.

To investigate the different selective constrains on duplicated *MdHOX* genes, the K_a/K_s ratio for each pair of duplicated *MdHOX*s was calculated. The results indicated that the majority of *MdHOX* paralogous pairs (35 of 51) underwent negative selection during evolution, since the ratios of K_a/K_s were less than 1, while the other 16 duplicated pairs might undergo positive selection because their K_a/K_s ratios were more than 1 (table 4 in electronic supplementary material). Based on the K_s value, the time of 51 duplicated events was calculated with a substitution rate of 6.5×10^{-9} substitutions per site per year (Wang *et al.* 2015), which varied from 0.264 to 214.57 Mya (table 5 in electronic supplementary material). The tandem duplication pairs occurred from 0.29 to 103.5 Mya, while the segmental duplication was formed from 7.2 to 112 Mya. This result indicated that the segmental duplication events occurred before the formation of tandem duplication.

The organ-specific expression pattern of MdHOXs

Previous studies suggested that *HOX*s are widely involved in organ development in plant, such as cotyledon development, flowering and fruit ripening (Chen *et al.* 2014). To further examine the organ-specific expression of *MdHOX*s, the expression levels of *MdHOX*s identified above were extracted from the expression profile of apple (GSE42873). All of the 207 *MdHOX*s could be found in this expression profile. According to their specific expression pattern in different organs of apple, all 207 *MdHOX*s were divided into seven groups and named as group I, II, III, IV, V, VI_A and VI_B (figure 4). In group I, five genes were grouped together because of their higher expression levels in fruits and stems. In group II, 31 genes exhibited similar expression patterns with high levels in flowers and seeds. Group III members (128 genes) showed higher expression in the leaves, flowers and fruits, and lower expression in other organs. Group IV contained 28 genes, and most of the genes showed lower expression in seeds and seedlings, but higher expression in other organs. In group V, all the members had higher expression in fruits and lower expression

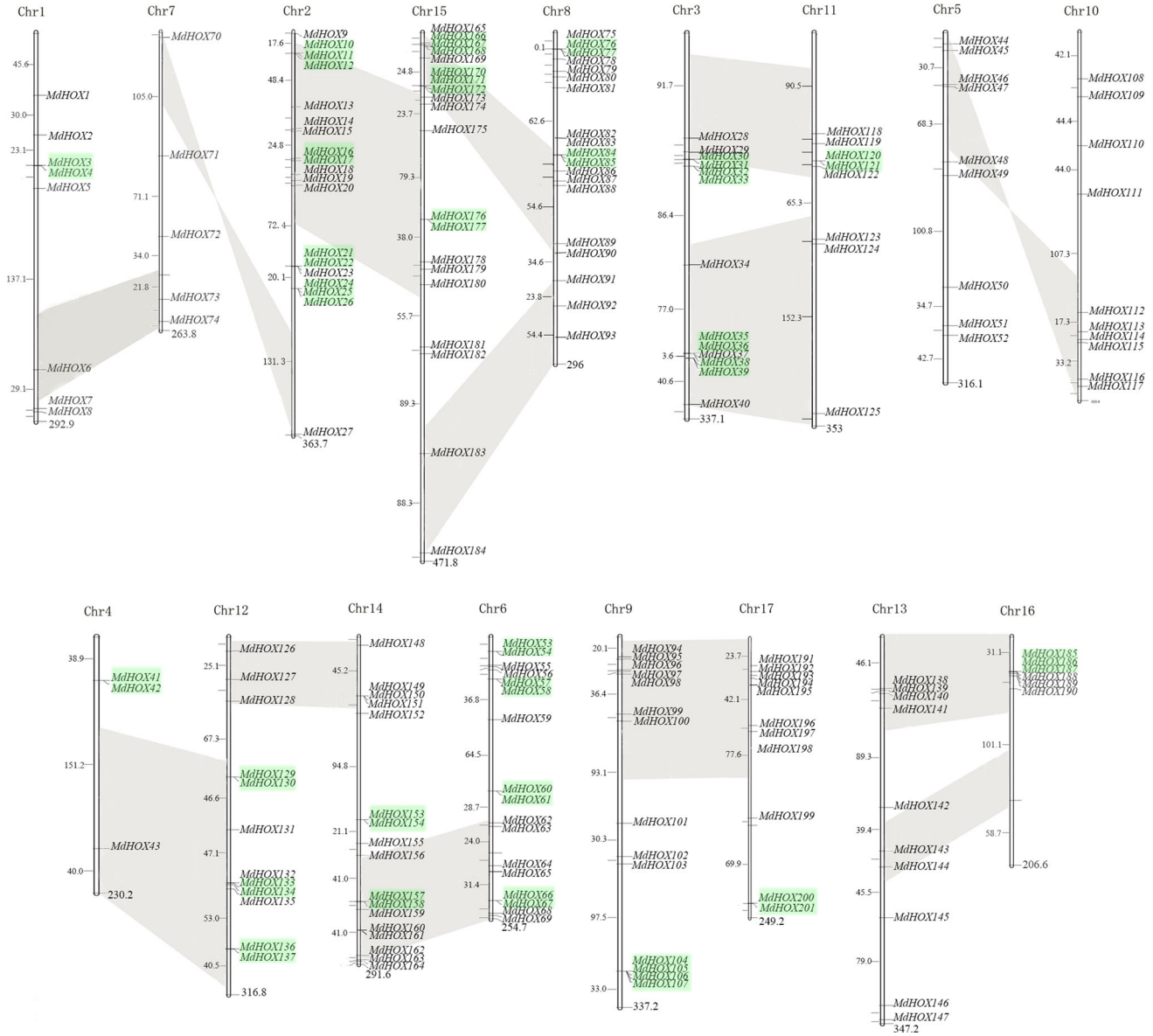

Figure 3. Distribution of the *MdHOX*s in 17 apple chromosomes. The numbers on the left side of each chromosome represent a megabase between two *MdHOX*s. Tandem duplications are indicated by light green. Shaded parts represent homologous regions on chromosomes.

in seedlings. Groups VI_A and VI_B exhibited the common characteristic of higher expression in stems, leaves and flowers and lower expression in fruits and seeds, but they showed completely opposite expression in seedlings. Generally, *MdHOX*s were ubiquitous in the organs of apple, which may participate in the plant growth and development.

To verify the reliability of the expression profiles (GSE42873) data, qRT-PCR analysis was performed in 23 *MdHOX* genes randomly selected from *MdHOX* genes. The results were found to be consistent with the data of GSE42873 (figure 2 in electronic supplementary material).

Analyses of stress-responsive cis-elements in upstream sequence of MdHOXs

HOXs act as TFs, regulating the expression of their downstream genes, while the expression of HOXs was related to their *cis*-acting sequences in promoter regions by interaction with other TFs. To investigate putative *cis*-acting elements related to abiotic stresses, 1500 bp upstream sequences of the *MdHOX*s were scanned with the PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Higo *et al.* 1998). Among 207 *MdHOX*s, 192 upstream sequences were found in the database of apple genome and total 86 types of elements

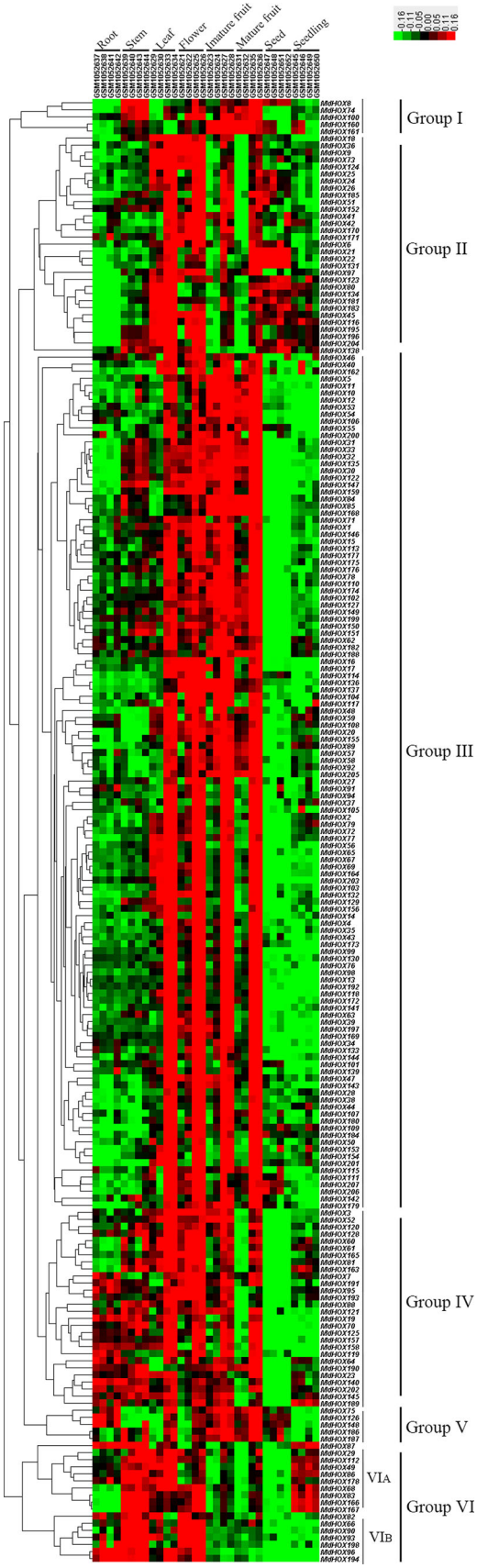

Figure 4. The expression pattern of *MdHOX*s in different organs.

were detected. The *cis*-acting elements involved in stress-response existed widely in the promoters of *MdHOX*s (table 1). More than 100 of 192 genes contained heat stress response elements (HSE), MYB-binding site (MBS), methyl jasmonate-response elements (CGTCA-motif and TGACG-motif), stress response elements (TC-rich repeats) and salicylic acid response elements (TCA-element). More than 90 genes harboured ABA-response element (ABRE) and gibberellin-response element (GARE). In addition, one or more types of stress-responsive *cis*-elements were found (table 5 in electronic supplementary material). For example, *MdHOX*144 harboured one MBS, three CGTCA-motif, two HSE and four ABRE. The presence of stress-responsive *cis*-elements indicates that *MdHOX*s may take part in the response to various types of abiotic stress in apple.

The expression of MdHOXs under abiotic stress

Systematic analysis of promoter suggested that stress-related *cis*-elements were widely existed in that of *Md*HOXs. To further investigate the potential functions of *MdHOX*s under different abiotic stress treatments, eight genes were randomly selected as experimental gene. One year seedlings were exposed to abiotic stress and exogenous ABA treatments, eight genes expression was measured using qRT-PCR, which is highly sensitive and widely used. To detect the quality of cDNA samples, the DREB gene (MDP0000147009) was selected as marker gene. The data showed that its expression was significantly upregulated under different stresses in apple roots (figure 3 in electronic supplementary material), which is similar with the previous report (Zhao *et al.* 2012). Under high salinity stress, two of eight genes were upregulated obviously, especially *MdHOX8*, whose expression increased nearly 6-fold and 35-fold in leaves and roots, respectively. Six of eight in leaves and four in roots of *MdHOX*s were upregulated significantly under PEG treatment, *MdHOX*5 were upregulated ∼5-fold in leaves and 40-fold in roots (figure 5). In response to chilling stress, four of eight *MdHOX*s in leaves and six in roots exhibited obvious upregulation. Among them, the expression of *MdHOX8* increased obviously both in roots and leaves, reaching more than 2-fold and 8-fold, which was consistent with the stress responsive *cis*-acting elements of *MdHOX8* (two ABRE, two CGTCA-motif, one low-temperature responsive (LTR), 2TCA-elements, two TGACG-motif, one TGA-motif). Under exogenous ABA treatment, the expressions of *MdHOX8* and *MdHOX169* were significantly upregulated both in roots and leaves, reaching ∼16-fold in leaves. In general, majority of eight *MdHOX*s were sensitive to various types of abiotic stress. Response to stress by *MdHOX*s is related to stress responsive *cis*-acting elements of the upstream sequence. All the results indicated that *MdHOX*s may play an important role in the regulation of the abiotic stress signalling pathway.

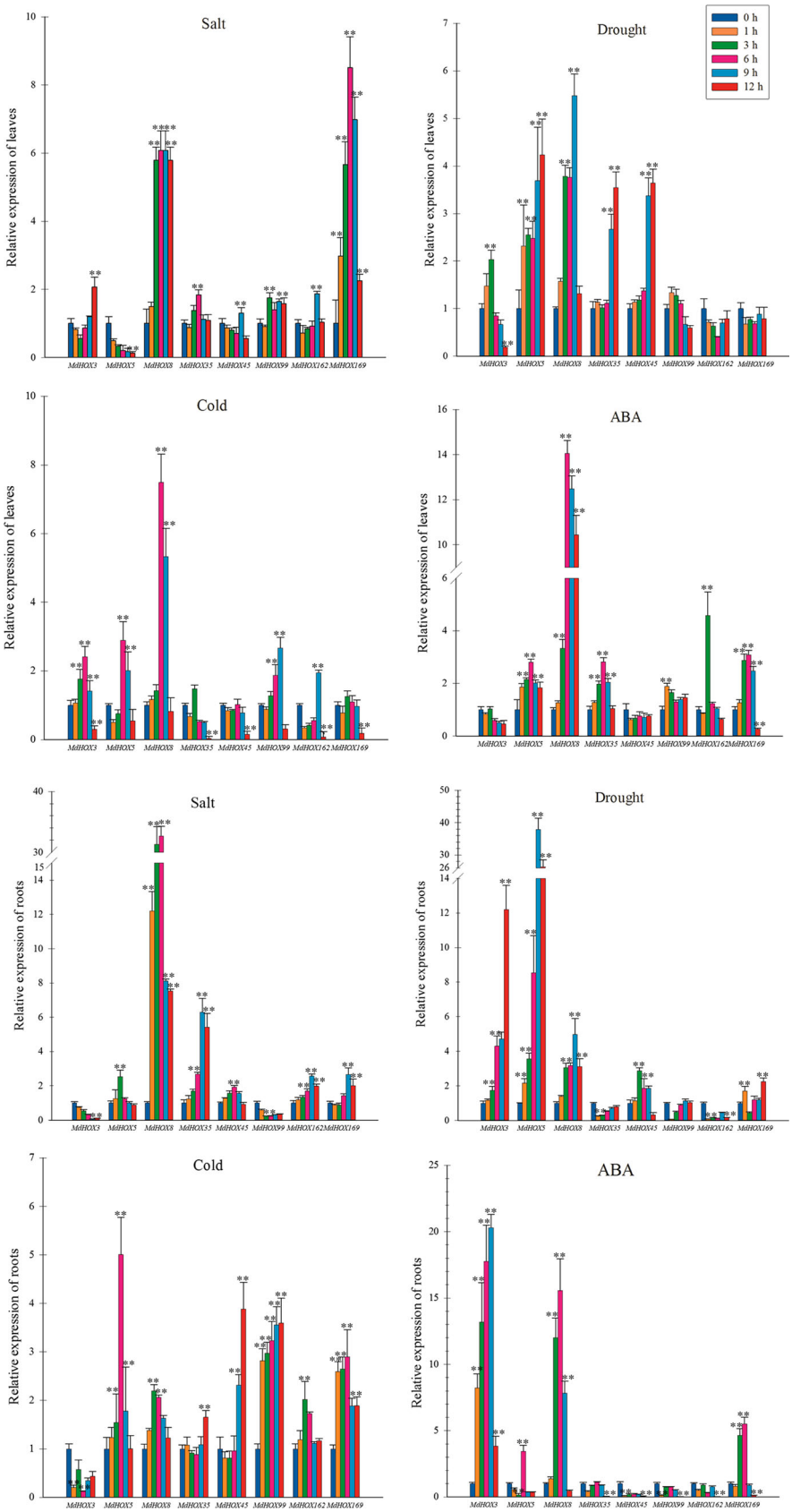

Figure 5. The expression of *MdHOX*s in leaves and roots under different types of abiotic stress. The mean expression value was calculated from three independent replicates. Vertical bars indicate the standard error of mean. $^{**}P \leq 0.01$ and $^{*}P \leq 0.05$ compared with 0 h.

Discussion

As one of the largest known families of regulatory proteins, HOXs are widely distributed in eukaryotes. The genomic analysis of plant HOX family has been performed in several species. There are 110 members predicted in *A. thaliana*, 107 in the rice, 89 in chickpea and 137 in pigeon pea (Jain *et al.* 2008; Mukherjee *et al.* 2009; Bhattacharjee *et al.* 2015). In this study, a total of 207 genes encoding HOXs were identified in apple, which was much more than that of other plants. Most of *MdHOX*s belonged to tandem duplication (marked in green colour in figure 3) and some *MdHOX*s were involved in segmental duplications (table 3 in electronic supplementary material). In addition, a genomewide duplication event was reported to have occurred in apple (Velasco *et al.* 2010). All these results suggested that gene tandem duplication, segmental duplication and genomewide duplication may all contribute to the expansion of the *MdHOX* gene family members. Similar to other species, the gene duplication in apple is derived from the long evolutionary history (Dong *et al.* 2000) and the presence of additional *HOX*s in the apple genome indicates that these genes are indispensable for the growth and development of apple.

Previous studies suggested that plant *HOX* genes from different species were classified into 14 classes based on phylogenetic relationship and domain composition (Bhattacharjee *et al.* 2015). In this study, 192 *Md*HOXs were distributed into 11 distinct known classes, and 15 members were divided into 'unclassified' class because they had no known functional domain other than HD (table 2 in electronic supplementary material). Compared with the number (total 156) of *MdHOX*s reported in the PlantTFDB database before, more members of *MdHOX*s were identified in this study and detailed classification has been performed. Notably, *Md*HOX58 and *Md*HOX3 (MDP0000206914 and MDP0000238771) were originally classified into HD-ZIP subfamily in PlantTFDB. However, no functional domains other than HD were identified, thus they were classified into unclassified group in this investigation.

Some HOX proteins have been involved in stress response in plants. *At*HB7 and *At*HB12, the members of HD-ZIP I subfamily were strongly induced by ABA and drought (Liu *et al.* 2013). *Gm*SBH1 (the member of KNOX subfamily) was found to respond to high temperature and humidity (Shu *et al.* 2015a, b). Although some *HOX*s have been found to participate in various types of abiotic stress, a comprehensive genomic analysis of *HOX*s in response to abiotic stress has not been carried out in apple. In this study, all the upstream sequences of 207 *MdHOX*s were analysed. The results revealed that the *cis*-acting elements related to abiotic stress widely existed in the promoter regions of 192 *MdHOX*s. More than half of 192 *MdHOX*s contained MBS, HSE, TC-rich repeats and ABRE elements, which were mainly involved in drought, high temperature and exogenous ABA response. Each gene of 192 *MdHOX*s harboured at least one of the *cis*-elements related to abiotic stress response, which suggested that some *MdHOX* genes may have potential ability to respond to abiotic stress. Further, the qRT-PCR results showed that the majority of *MdHOX*s responded sensitively to stress treatments; especially the expression of *MdHOX8* was increased significantly both in roots and leaves under all abiotic stress and exogenous ABA treatment. This finding indicated that *MdHOX* may play putative role in abiotic stress signalling and response to exogenous hormone regulation.

In conclusion, the *HOX* gene family consists of a large group of genes with different functions. In the present study, the genomewide analysis of the *HOX* gene family was performed in apple for the first time. Further, according to systematic analysis of stress-responsive *cis*-acting elements in promoter of *MdHOX*s, their expression profiles were examined in various types of abiotic stress. The results could provide useful information for further studies on *MdHOX*s, particularly for the members that may have potential ability to respond to abiotic stress.

Acknowledgements

This work was supported by the Special Research Fund of Public Welfare of China Agricultural Ministry (201303093) and China Natural Foundation (31872042).

References

Angela H. and Miltos T. 2010 KNOX genes: versatile regulators of plant development and diversity. *Development* **137**, 3153–3165.

Ariel F. D., Manavella P. A., Dezar C. A. and Chan R. L. 2007 The true story of the HD-Zip family. *Trends Plant Sci.* **12**, 419–426.

Bhattacharjee A., Ghangal R., Garg R. and Jain M. 2015 Genome-wide analysis of homeobox gene family in legumes: identification, gene duplication and expression profiling. *PLoS One* **10**, e0119198.

Burglin T. R. and Affolter M. 2016 Homeodomain proteins: an update. *Chromosoma* **125**, 497–521.

Chan R. L., Gago G. M., Palena C. M. *et al.* 1998 Homeoboxes in plant development. *Biochim. Biophys. Acta* **1442**, 1–19.

Chen X., Chen Z., Zhao H., Zhao Y., Cheng B. and Xiang Y. 2014 Genome-wide analysis of soybean HD-Zip gene family and expression profiling under salinity and drought treatments. *PLoS One* **9**, e87156.

Cheng S., Huang Y., Zhu N. and Zhao Y. 2014 The rice WUSCHEL-related homeobox genes are involved in reproductive organ development, hormone signaling and abiotic stress response. *Gene* **549**, 266–274.

Costanzo E., Trehin C. and Vandenbussche M. 2014 The role of WOX genes in flower development. *Ann. Bot.* **114**, 1545–1553.

Cui H. R., Zhang Z. R., Lv W., Xu J. N. and Wang X. Y. 2015 Genome-wide characterization and analysis of F-box protein-encoding genes in the *Malus domestica* genome. *Mol. Genet. Genomics* **290**, 1435–1446.

Dong Y. H., Yao J. L., Atkinson R. G., Putterill J. J., Morris B. A. and Gardner R. C. 2000 MDH1 an apple homeobox gene belonging to the BEL1 family. *Plant Mol. Biol.* **9**, 42–49.

Du D., Zhang Q., Cheng T., Pan H., Yang W. and Sun L. 2013 Genome-wide identification and analysis of late embryogenesis abundant (LEA) genes in *Prunus mume*. *Mol. Biol. Rep.* **40**, 1937–1946.

Eddy S. R. 1998 Profile hidden Markov models. *Bioinformatics* **14**, 78–96.

Gasic K., Hernandez A. and Korban S. S. 2004 RNA extraction from different apple tissues rich in polyphenols and polysaccharides for cDNA library construction. *Plant Mol. Biol. Rep*. **22**, 437–444.

Hake S. 2016 Homeobox gene in the functioning of plant meristem. *Biol. Sci*. **350**, 45–51.

Higo K., Ugawa Y., Iwamoto M. and Higo H. 1998 Place: a database of plant cis-acting regulatory DNA elements. *Nucleic Acids Res.* **26**, 358–359.

Hu R., Chi X., Chai G., Kong Y., He G., Wang X. *et al.* 2012 Genome-wide identification, evolutionary expansion, and expression profile of homeodomain-leucine zipper gene family in poplar (*Populus trichocarpa*). *PLoS One* **7**, e31149.

Jain M., Tyagi A. K. and Khurana J. P. 2008 Genome-wide identification, classification, evolutionary expansion and expression analyses of homeobox genes in rice. *FEBS J*. **275**, 2845–2861.

Katiyar S., Till E. A. and Lennarz W. J. 2001 Studies on the function of yeast protein disulfide isomerase in renaturation of proteins. *Biochim. Biophys. Acta* **15**, 22–31.

Larkin M. A., Blackshields G., Brown N. P., Chenna R., McGettigan P. A., McWilliam H. *et al.* 2007 Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948.

Li Y., Wu B., Yu Y., Yang G., Wu C. and Zheng C. 2011 Genome-wide analysis of the RING finger gene family in apple. *Mol. Genet. Genomics* **286**, 81–94.

Liu W., Fu R., Li Q., Li J., Wang L. and Ren Z. 2013 Genome-wide identification and expression profile of homeodomain-leucine zipper Class I gene family in *Cucumis sativus*. *Gene* **531**, 279–287.

Liu X., Li R., Dai Y., Chen, X. and Wang X. 2017 Genome-wide identification and expression analysis of the B-box gene family in the Apple (*Malus domestica* Borkh.) genome. *Mol. Genet. Genomics* **293**, 303–315.

Maeda R. K. and Karch F. 2015 The open for business model of the bithorax complex in *Drosophila*. *Chromosoma* **124**, 293–307.

Mukherjee K., Brocchieri L. and Burglin T. R. 2009 A comprehensive classification and evolutionary analysis of plant homeobox genes. *Mol. Biol. Evol.* **26**, 2775–2794.

Quraishi U. M., Abrouk M., Murat F., Pont C., Foucrier S., Desmaizieres G. *et al.* 2011 Cross-genome map based dissection of a nitrogen use efficiency ortho-metaQTL in bread wheat unravels concerted cereal genome evolution. *Plant J.* **65**, 745–756.

Sharma P., Lin T., Grandellis C., Yu M. and Hannapel D. J. 2014 The BEL1-like family of transcription factors in potato. *J. Exp. Bot.* **65**, 709–723.

Shu Y., Liu Y., Zhang J., Song L. and Guo C. 2015 Genome-wide analysis of the AP2/ERF superfamily genes and their responses to Abiotic stress in *Medicago truncatula*. *Front. Plant Sci.* **6**, 12–47.

Shu Y., Tao Y., Wang S., Huang L., Yu X., Wang Z. *et al.* 2015 GmSBH1, a homeobox transcription factor gene, relates to growth and development and involves in response to high temperature and humidity stress in soybean. *Plant Cell Rep.* **34**, 1927–1937.

Tan Q. K. and Irish V. F. 2006 The Arabidopsis zinc finger-homeodomain genes encode proteins with unique biochemical properties that are coordinately expressed during floral development. *Plant Physiol.* **140**, 1095–1108.

Tsuda K., Kurata N., Ohyanagi H. and Hake S. 2014 Genome-wide study of KNOX regulatory network reveals brassinosteroid catabolic genes important for shoot meristem function in rice. *Plant Cell* **26**, 3488–3500.

Velasco R., Zharkikh A., Affourtit J., Dhingra A., Cestaro A., Kalyanaraman A. *et al.* 2010 The genome of the domesticated apple (Malus × domestica Borkh.). *Nat. Genet.* **42**, 833–839.

Wang X. L., Zhong Y., Cheng Z. M. and Xiong J. S. 2015 Divergence of the bZIP gene family in strawberry, peach, and apple suggests multiple modes of gene evolution after duplication. *Int. J. Genomics* **536**, 39–43.

Wu S., Wu M., Dong Q., Jiang H., Cai R. and Xiang Y. 2016 Genome-wide identification, classification and expression analysis of the PHD-finger protein family in *Populus trichocarpa*. *Gene* **575**, 75–89.

Xu J. N., Xing S. S., Zhang Z. R., Chen X. S. and Wang X. Y. 2016 Genome-wide identification and expression analysis of the Tubby-like protein family in the *Malus domestica* genome. *Front. Plant Sci.* **7**, 64–69.

Xu J. N., Xing S. S., Cui H. R., Chen X. S. and Wang X. Y. 2016 Genome-wide identification and characterization of the apple (*Malus domestica*) HECT ubiquitin-protein ligase family and expression analysis of their responsiveness to abiotic stresses. *Mol. Genet. Genomics* **291**, 635–646.

Zhao T., Liang D., Wang P., Liu J. and Ma F. 2012 Genome-wide analysis and expression profiling of the DREB transcription factor gene family in Malus under abiotic stress. *Mol. Genet. Genomics* **287**, 423–436.

Corresponding editor: U. C. Lavania