Plant Growth Regul
DOI 10.1007/s10725-017-0268-1

ORIGINAL PAPER

Preferential retention, expression profile and potential functional diversity analysis of HD-Zip gene family in *Brassica rapa*

ZanGe Jing[1,2] · WeiKe Duan[1,3] · XiaoMing Song[1,4] · Peng Wu[1] · Jun Tang[1] · Zhen Wang[1] · YanPing Wei[1] · Cheng Wang[1] · XiLin Hou[1]

Received: 20 July 2016 / Accepted: 25 March 2017

Abstract Homeodomain-Leu zipper (HD-Zip) gene family performs important biological functions related to organ development, photomorphogenesis and abiotic stress response in higher plants. However, systematic analysis of HD-Zip genes in *Brassica rapa* has not been performed. In the present study, a bioinformatics approach was used to identify and characterize the *BraHD-Zip* gene family in *B. rapa*. A total of 88 members were identified. All putative *BraHD-Zip* proteins contained a clear HD and LZ combined domain. Eighty-seven *BraHD-Zips* were non-randomly located on ten chromosomes. This gene family was mainly expanded following the whole genome triplication event and was preferentially over-retained relative to its neighboring genes in *B. rapa*. On phylogenetic analysis, the *BraHD-Zip*s could be categorized into four distinct major groups (I–IV). Each group exhibited variant gene structures and motif distributions. Some syntenic orthologous gene pairs presented diverse expression profiles, which indicate that these gene pairs may be involved in the development of new functions during evolution. In summary, our analysis provided genome-wide insights into the expansion, preferential retention, expression profiles and functional diversity of *BraHD-Zip* genes following whole genome triplication in *B. rapa*.

Keywords HD-Zip genes · *Brassica rapa* · Genome-wide analysis · Expression profiles · Functional diversity

Introduction

As a transcription factor (TF) gene family, Homeodomain-Leu Zipper (HD-Zips) participates in various biological and physiological processes, including in the regulation of organ development, flowering time, hormone-mediated signaling and stress responses (Ariel et al. 2007; Chen et al. 2014). This superfamily can be divided into four distinct classes (I–IV) (Sessa et al. 1997). Much of the data pertaining to functions of HD-Zip superfamily has emanated from studies conducted in plant *Arabidopsis thaliana* (*A. thaliana*). Members of the HD-Zip subfamily I are mainly involved in abiotic stress response, photomorphogenesis and organ development (Ariel et al. 2007). *A. thaliana* homeobox protein 12 (*ATHB12*) and its paralogous gene homeobox protein 7 (*ATHB7*) showed similar up-regulation patterns under conditions of water deficit and also on exogenous treatment with abscisic acid (ABA). Over-expression of *ATHB7* and/or *ATHB12* was shown to result in delayed elongation of inflorescence stem and increased branching

Electronic supplementary material The online version of this article (doi:10.1007/s10725-017-0268-1) contains supplementary material, which is available to authorized users.

✉ XiLin Hou
hxl@njau.edu.cn

[1] State Key Laboratory of Crop Genetics and Germplasm Enhancement/Key Laboratory of Biology and Germplasm Enhancement of Horticultural Crops in East China, Ministry of Agriculture, Nanjing Agricultural University, Nanjing 210095, People's Republic of China

[2] Zhenan Key Laboratory of Crop Breeding, Wenzhou Vocational College of Science and Technology, Wenzhou 325006, People's Republic of China

[3] School of Life Science and Food Engineering, Huaiyin Institute of Technology, Huaian 223003, People's Republic of China

[4] Center of Genomics and Computational Biology, College of Life Sciences, North China University of Science and Technology, Tangshan 063000, People's Republic of China

Published online: 24 May 2017

in *A. thaliana* (Olsson et al. 2004). *A. thaliana* homeobox protein 16 (*ATHB16*) was shown to negatively regulate leaf cell expansion and flowering time sensitivity to photoperiod (Wang et al. 2003). The expressions of HD-Zip subfamily II genes were shown to be regulated by the quality of light and hormone to affect organ development. Function of five members were regulated by light quality involved during photomorphogenesis (Carabelli et al. 2013). Further, *A. thaliana* homeobox protein 4 (*ATHB4*) and Histone acetyltransferase 3 (*HAT3*) were shown to affect the establishment of dorso-ventral axis to regulate leaf development in *A. thaliana* (Torrent et al. 2012). *A. thaliana* homeobox protein 17 (*ATHB17*) as a positive regulator of ABA response, affected growth of seedlings during a narrow developmental stage (Park et al. 2013).

HD-Zip subfamily III genes was shown to play crucial roles in the differentiation of vascular development, organ polarity, apical embryo, polar auxin transport and meristem function (Prigge et al. 2005). *REVOLUTA* loss-of-function mutations were shown to cause defects in vascular/leaf development and auxin transport (Prigge et al. 2005; Agalou et al. 2008). Over-expression of *A. thaliana* homeobox protein 8 (*ATHB8*) caused precocious secondary growth in *Arabidopsis* (Baima et al. 2001). HD-Zip IV genes are mainly involved in epidermal trichome differentiation, anthocyanin accumulation, and root development (Hu et al. 2012). Several members established cell fates in epidermal layer (Ariel et al. 2007). Anthocyaninless 2 (*ANL2*) affected anthocyanin accumulation in leaf epidermis, and was also involved in the determination of root cell identity (Kubo et al. 1999).

Whole genome duplication (WGD) events widely occurred during history of plant evolution. *A. thaliana* underwent at least three ancient polyploid events which were shared with other members of *Brassicales* (Duan et al. 2015). After divergence from *A. thaliana* lineage, occurrence of an addition whole-genome triplication (WGT) event has been confirmed in *Brassica rapa*. The density of orthologous genes in three subgenomes of *B. rapa* were shown to be dramatically divergent as compared to that in *A. thaliana* (Project 2011). The proportion of genes retained in least fractionated subgenome (LF) was much higher than that in the medium fractionated subgenome (MF1) and the most fractionated subgenome (MF2) (Cheng et al. 2012a). Recent studies have confirmed that transcription factors are preferentially retained in a dosage-sensitive relationship following diploidization of ancient WGD events (Lou et al. 2012). These retained orthologous TF genes may have provided raw material for functional innovation as part of the evolutionary process.

In the present study, 88 BraHD-Zip transcription factor putative candidate genes were comprehensively identified in *B. rapa*. Further, we analyzed their genome organization, conserved motifs and sequence phylogeny. After examining their expression profiles in different organs/tissues, we also determined the expression profiles of selected duplicated genes in group I under different abiotic stress treatments using qRT-PCR. Our results will be helpful for further functional analyses of *BraHD-Zip*s and would also help identify candidate genes to improve stress-resistance in *B. rapa*.

Results

Identification and characterization of HD-Zips

After careful survey of the entire *B. rapa* database, a total of 88 HD-Zip homologs were identified by BLASTp program (Table S1). All candidate proteins contained a conserved HD-Zip combined domain (Table S2). All *BraHD-Zip*s except *BraHD-Zip*88 (Bra040941) were mapped on ten chromosomes (Fig. S1). All members of *HD-Zip* family were located on the segmental duplicated regions (Fig. S1). We also identified 30 HD-Zips in *Carica papaya*, 63 HD-Zips in *Populus trichocarpa*, 33 HD-Zips in *Vitis vinifera* and 21 HD-Zips in *Amborella trichopoda* (Fig. 1; Table S3).

Predictions of physical and chemical characterization and signal peptide of a gene product can provide additional functional information. Our results revealed that four

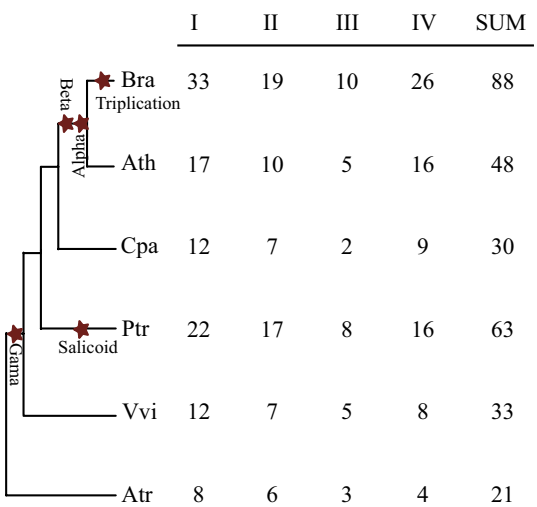

Fig. 1 Copy number variation of HD-Zip transcription factors family in eudicots. The species tree is shown at the *left*. The α, β, γ and salicoid duplications and Brassica-specific triplication are indicated on *branches of the trees* according to the Plant Genome Duplication Database. *Numbers* are genes numbers of each group in *Brassica rapa* (Bra), *Arabidopsis thaliana* (Ath), *Carica papaya* (Cpa), *Populus trichocarpa* (Ptr), *Vitis vinifera* (Vvi), and *Amborella trichopoda* (Atr)

subgroups showed diverse physical and chemical characterizations in each selected plant species (Fig. S2). The results of signal peptide prediction indicated that all *BraHD-Zip*s did not possess signal sequences for targeting the secretory pathway (Table S1).

Differential retention of *BraHD-Zip* genes

The dosage-sensitive gene duplicates are over-retained to maintain the stability of transcriptional and signal networks during the evolutionary process (Freeling and Thomas 2006; Lou et al. 2012). The retention of four gene sets were compared, including the HD-Zip gene set, the set of 960 neighboring genes (ten on either side) flanking 48 *AtHD-Zip* genes, set of 458 core eukaryotic genes and 458 random selected genes from the microsyntenic regions corresponding to the *HD-Zip* genes (Table S4). Approximately 60% of *HD-Zip* genes retained two or three copies, much more than the comparison group of neighboring gene set (Fig. 2a). Of note, only 4 (8.51%) *BraHD-Zip* genes were completely lost, which was less than that in the neighboring and core comparisons (Fig. 2a). The proportion of retained *HD-Zip* genes among LF, MF1 and MF2 subgenomes was significantly higher than the neighboring gene set (Fig. 2b). In the LF subgenome, 74.47% of *BraHD-Zip* genes were retained, whereas MF1 and MF2 sub-genomes retained substantially lower proportions of *BraHD-Zip* genes (48.94 and 40.43%, respectively) (Fig. 2b, c).

Analysis of synonymous substitution rates

Using syntenic orthologous *HD-Zip* gene pairs identified between *B. rapa* and *A. thaliana*, the Ka, Ks and Ka/Ks ratios were calculated. The Ka/Ks ratios of all HD-Zip syntenic orthologs were ≤1. This distribution (Ka/Ks < 1) represented that the genes within each group were subject to strong purifying selection after divergence (Table S5). The Ks values of the *HD-Zip* genes ranged from 0.2667 (8.89 Mya) to 0.8376 (27.92 Mya) with average of 0.4405 (~14.68 Mya) (Fig. 3a). Pair-wise divergences were indistinguishable from each other among the three sub-genomes (Fig. 3b). Investigation of the divergence amongst the four HD-Zip groups showed less divergence in the HD-Zip III group as compared to that in the other three subgroups (Fig. 3c). The divergence of this group occurred later and the genes underwent very strong purification selection pressure (Fig. S3).

Phylogenetic analyses of *BraHD-Zip* genes

To understand the evolutionary relationship of *HD-Zip*s in *B. rapa* and *A. thaliana*, a phylogenetic maximum likelihood tree was constructed using full-length HD-Zip protein

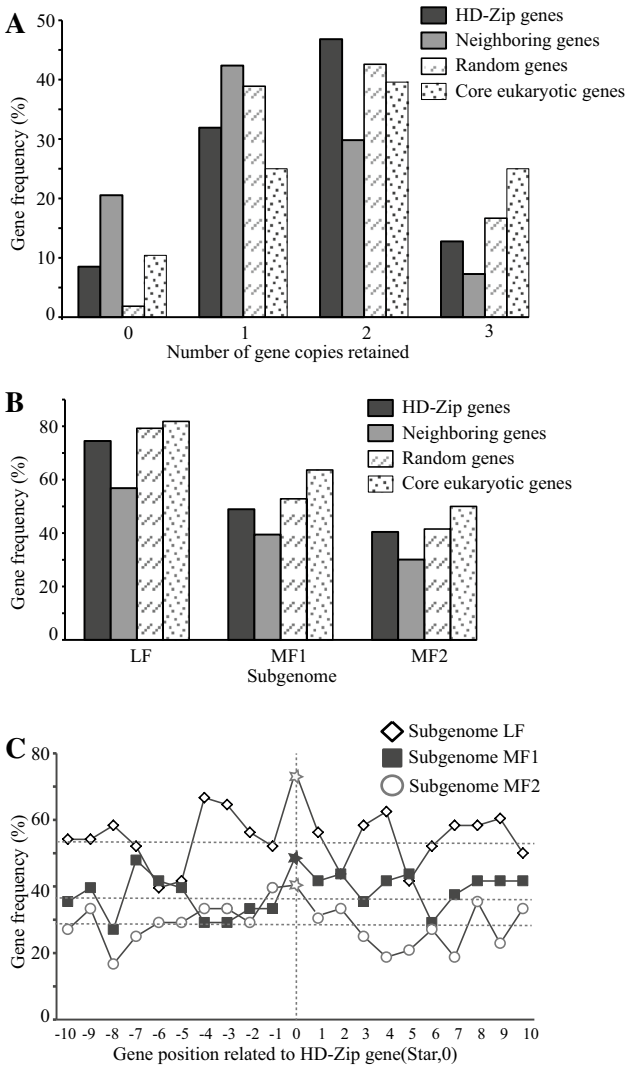

Fig. 2 Retention of HD-Zip genes, neighboring genes (ten flanking genes on either side of HD-Zip gene), 458 randomly selected genes, and 458 core eukaryotic genes in the syntenic region after whole-genome triplication in *Brassica rapa*. **a** Retention by number of homoeologous copies in the syntenic region. **b** Retention of homoeologs among three subgenomes of *B. rapa*. **c** Retention of homoeologs among the HD-Zip genes (*stars*) and their neighbors in three subgenomes of *B. rapa. Skew open squares* indicate LF, *filled squares* indicate MF1, and *open circles* indicate MF2

sequences. It can be seen that the BraHD-Zips clustered into four major groups (I–IV) (Fig. 4). Group I comprised of 33 genes and they could be further clustered into multiple subgroups (Henriksson et al. 2005). Group II contained 19 *BraHD-Zip* genes that could be clustered into six subgroups. Amongst these, only three genes corresponded to a single Arabidopsis gene, including *BraHD-Zip29*/*ATHAT9, BraHD-Zip86*/*ATHAT14* and *BraHD-Zip06*/*ATHB2*. Group IV contained twenty-six HD-Zip genes and Group III, which was the smallest group, contained ten *BraHD-Zip* genes. We also identified 31 sister pairs with high bootstrap

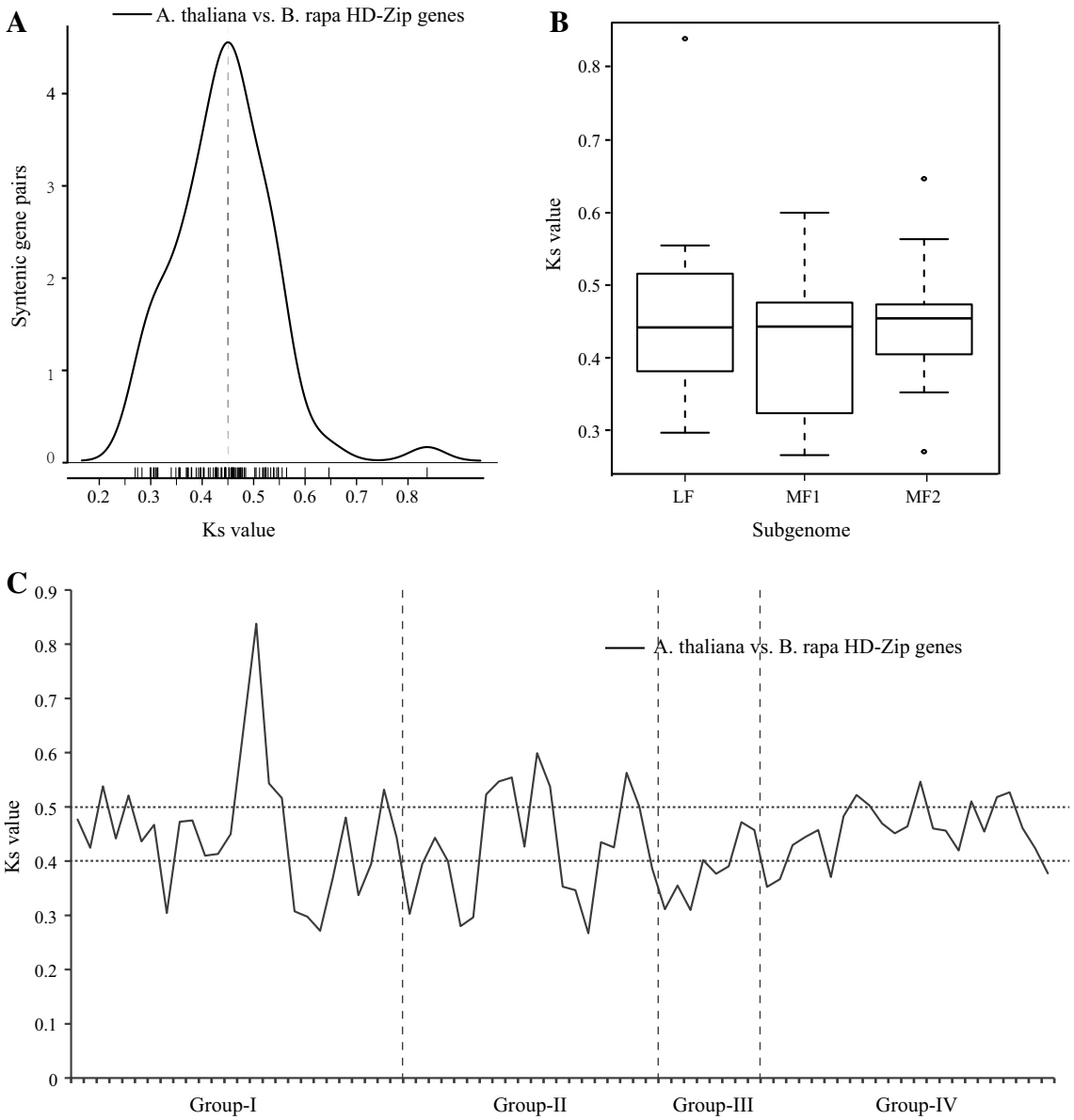

Fig. 3 Pairwise comparison of Ks values for HD-Zip genes between *Brassica rapa* and *Arabidopsis thaliana*. **a** The distribution of Ks values for HD-Zip genes. The *dashed line* indicates the divergence time (15 Mya). **b** The distribution of Ks values for HD-Zip genes in three subgenomes. **c** The distribution of Ks values for four HD-Zip groups. The *horizontal dashed line* indicates the main concentrated area of the Ks value (0.4–0.5)

support. More than half of HD-Zip sister pairs corresponded to two or three orthologous genes between *B. rapa* and *A. thaliana*.

The exon/intron organizations were compared to examine the structural gene diversity among different subfamilies. The intron numbers dramatically varied from zero to seventeen. Most closely related *BraHD-Zip* sister pairs shared similar intron numbers and exon/intron structures (Fig. S4). The *BraHD-Zip* genes were subjected to MEME program to detect their conserved motifs. Twenty conserved motifs were identified in total (Table S6). The distribution of motifs in the four groups was noticeably diverse. These conserved motifs, like HD-Zip, CPSCE, START and MEKHLA, showed specific distribution pattern (Fig. S4).

Expression profiles and potential functional diversity of duplicated *BraHD-Zip* pairs

In this study, we analyzed *BraHD-Zip*s expression characteristics using Illumina RNA-seq data (Tong et al. 2013). Among 88 *BraHD-Zips*, 59 members were expressed at

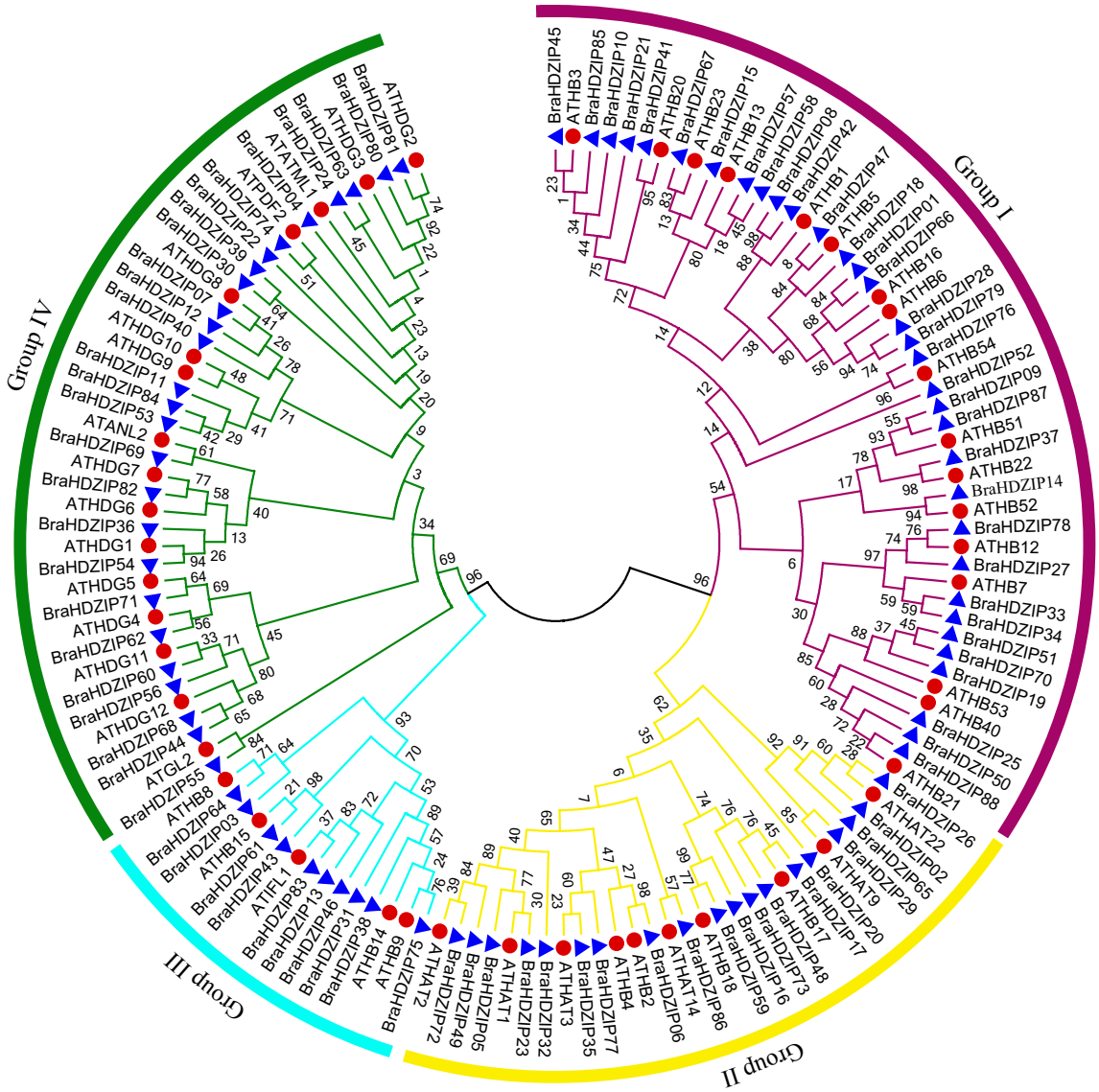

Fig. 4 Phylogenetic analysis of HD-Zip proteins between *Brassica rapa* and *Arabidopsis thaliana*. Protein sequences were aligned by ClustalW2. A phylogenetic tree was constructed by maximum likelihood (ML) and bootstrap values were calculated with 1000 replications using MEGA5.0. Group I, II, III and IV HD-Zip genes are shown in *different color*. (Color figure online)

relatively high levels (FPKM > 5) in at least one organ/tissue (Table S7). On silico analysis, the expression profiles of syntenic orthologous gene pairs were compared in three subgenomes of *B. rapa* (Fig. 5; Table S7). These gene pairs exhibited diverse expression profiles. The first type revealed a paralogous gene with higher expression levels than that in its corresponding gene in one or more organs/tissues, while its corresponding gene showed higher expression in some other organs/tissues. The second type was the paralogous gene, which was expressed at higher levels than the others in all tissues. For the paralogous pairs in which all three copies were detected, their expression patterns showed a tendency to be a mixture of the above two types.

Expression analysis and potential functional diversity of duplicated *BraHD-Zip* subfamily I gene pairs under abiotic stress

HD-Zip subfamily I is mainly involved in abiotic stress response to environmental factors in higher plants (Chen et al. 2014). Six duplicated gene pairs of BraHD-Zip group I showed comprehensive expression profiles in response to NaCl, PEG, heat and cold treatments (Tables S8, S9). *BraHD-Zip*s showed diverse expression profiles when subjected to the above treatments (Fig. 6). The change in gene expression profiles when subjected to NaCl, PEG, and heat treatments were more obvious than that under cold

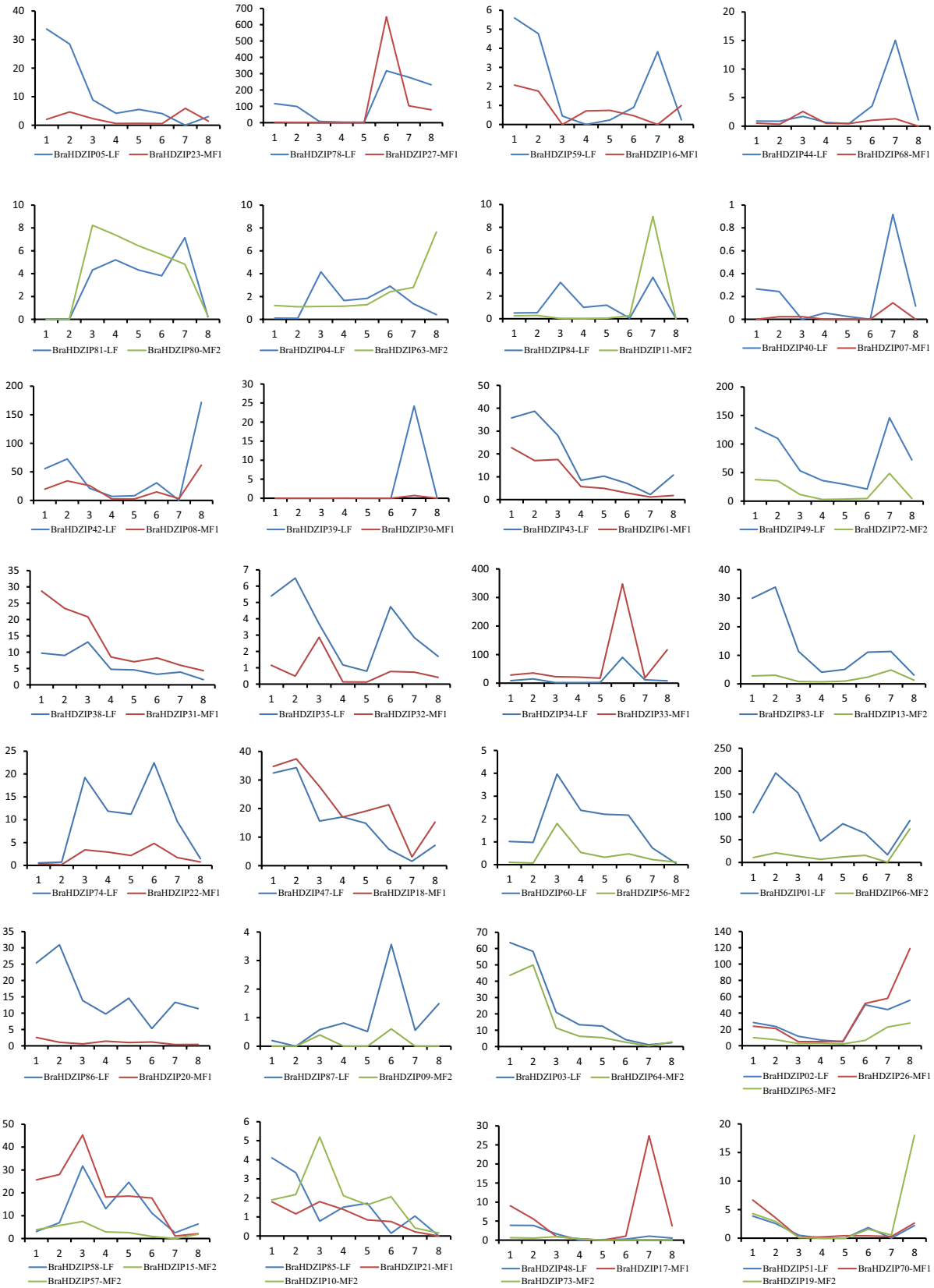

Fig. 5 Expression patterns of syntenic orthologous BraHD-Zip gene pairs. *X-axis* represents the different tissues (1 root-1, 2 root-2, 3 stem, 4 leaf-1, 5 leaf-2, 6 flower, 7 silique, 8 callus). *Y-axis* represents the FPKM expression value (supplementary Table S7)

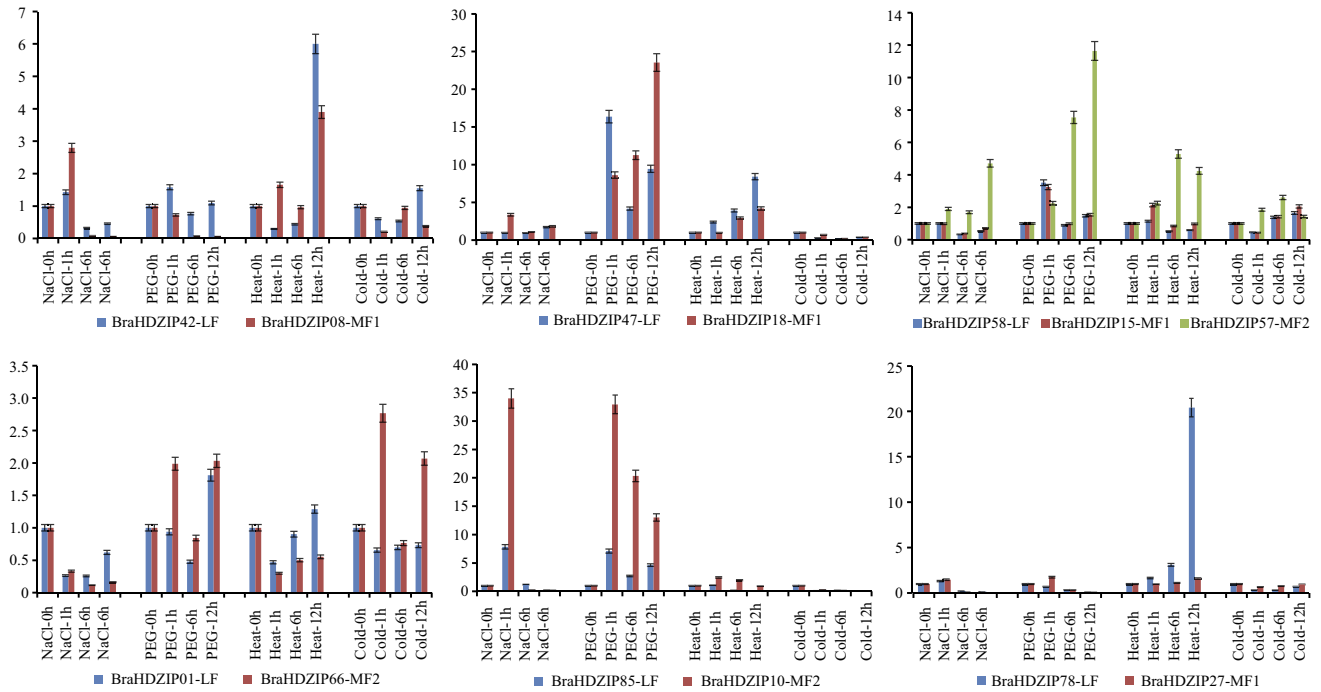

Fig. 6 Expression patterns of BraHD-Zip duplicated genes in response to various abiotic stresses, including 200 mM NaCl, 20% polyethylene glycol (PEG)-6000, heat (38 °C) and cold (4 °C). The *Y-axis* is the scale of relative expression levels. The *X-axis* represents the different simples under abiotic treatments. All values are the means of three measurements

treatment. *BraHD-Zip27/78* exhibited distinct response models under heat stress. The relative abundance of *BraHD-Zip10* dramatically increased under NaCl and PEG treatment as compared to *BraHD-Zip85*. Among the three duplicates of syntenic orthologous pair, *BraHD-Zip58/15* presented similar expression pattern, while *BraHD-Zip57* exhibited more intense response to stress.

Discussion

Comprehensive identification of gene family is essential to study its evolution and function. A large number of HD-Zip members have been investigated in several plants; however, systematic studies of *HD-Zip*s in *B. rapa* have been largely lacking. In this study, we comprehensively analyzed a list of 88 *BraHD-Zips*. *B. rapa* has a relatively large genome size when compared with other plants. Considering, the number of *HD-Zips* in other plants, it is likely that most, if not all, of *HD-Zips* have been identified in *B. rapa*. The chromosomal locations of *BraHD-Zips* were distributed with highly variable density in the ten chromosomes. Contrary to expectations, no *BraHD-Zip* genes involved in cluster were considered as tandem duplicated genes. This data suggested that the major contribution of segmental duplications is to expand the *HD-Zip* family. The distribution may reflect the variation in *BraHD-Zips* introduced via segmental duplication during its evolution. Recent work suggests that the segmental and tandem duplications caused a substantial expansion of the gene family during the process of genome evolution (Zhao et al. 2011). This amplification reflects the evolutionary particularity of the *HD-Zip* gene family in *B. rapa*.

Genome evolution and preferential retention of the BraHD-Zip transcription factor family

By detecting the footprint of *HD-Zip* genes in eudicots, 283 HD-Zip genes were compared in 6 selected species. *B. rapa* contained a great amount of family members of HD-Zip genes (88 members), which is notably more than that of the other five species [*A. thaliana* (47), *C. papaya* (30), *P. trichocarpa* (63), *V. vinifera* (33), and *A. trichopoda* (21)]. Through comparative genomics analysis, we found that this family was expanded mainly by genome duplication, accompanied by substantial gene loss over the evolutionary process. The biased gene fractionation occurred typically following the polyploid formation in the eukaryotes (Lim et al. 2010). A large number of gene families exhibit a similar pattern of evolutionary dynamics in plants (Lou et al. 2012; Duan et al. 2015). The HD-Zip gene family also followed this common evolutionary mechanism for dynamic expansion of the gene family (Lou et al. 2012; Duan et al. 2015).

On phylogenetic analysis, *BraHD-Zip* genes were found clustered into four major subfamilies. Further, the majority of subfamilies could be further divided into smaller clades. Within each clade, one to three similar orthologous genes emerged between *B. rapa* and *A. thaliana*. Therefore it is entirely conceivable that the *BraHD-Zip* genes mainly expand following the WGT event after divergence from the A. thaliana lineage.

Whole genome duplication (WGD) events has been a significant part in the evolution of complexity and facilitated speciation during the long history of plant lineages. After duplication events occurred, the fates of gene duplicates were distinct during diploidization. The gene balance hypothesis predicts that dosage-sensitive gene duplicates were preferentially retained to maintain the stability of transcriptional networks (Freeling and Thomas 2006). This hypothesis was also confirmed by comparative genomics between *B. rapa* and *A. thaliana*. Genes with Gene Ontology terms associated with response to environmental factors and plant hormones were also over retained (Project 2011). Thus, *HD-Zip* genes that are strongly involved in transcriptional and/or signaling networks should exhibit preferential retention in *B. rapa*. In this study, approximately 59.57% HD-*Zip* genes were retained in the form of two or three duplicates after whole-genome triplication in *B. rapa*. For their neighboring genes, only 37.09% genes could be detected as multi duplicates.

Expression profiles and potential functional diversity of duplicated pairs of *BraHD-Zip* genes

After a comprehensive investigation of *BraHD-Zip* gene expression characteristics (Tong et al. 2013), we found that most HD-Zips were expressed at relatively high levels, and *AtHD-Zip* genes also exhibited similar characteristics (Table S10; Fig. S5). This characteristic was the hallmark of the survived *HD-Zip*s, which were likely played a key role in their growth and development. During plant evolution, orthologous genes appear to have conserved expression profiles and functions. According to the functions of their orthologous genes, the possible functions of *HD-Zip*s could be hypothesized in *B. rapa* by comparing with orthogous genes found in *A. thaliana*. For example, the orthologue of *BraHD-Zip01/66* is *ATHB16*, which could negatively regulate leaf cell expansion and flowering time sensitivity to photoperiod in Arabidopsis. Therefore, *BraHD-Zip01/66* may have a similar function in the regulation of leaf development and flowering time. Interestingly, *BraHD-Zip01/66* exhibited diverse expression profiles in leaf, flower and callus. *BraHD-Zip01* had a relatively high expression in leaf and flower, as compared to its paralogous genes *BraHD-Zip66*. Both these had a very high expression level in the callus. This diverse expression characteristic suggests that paralogous gene pair *BraHD-Zip01/66* may have prevented potential harmful mutations or innovate new functions during its evolution.

Duplicate genes provide raw material for functional innovation over evolutionary time. The potential paralogous relationships facilitated us to understand possible functional redundancy and diversity of *HD-Zips* in *B. rapa*. *ATHAT1* was involved in meristematic activity and was reported to be important for fruit development in Arabidopsis (Victor et al. 2012). Being its orthologou, *BraHD-Zip05* and *BraHD-Zip23* exhibited diverse expression in silique. This diversity may demonstrate that *BraHD-Zip05* may have lost regulatory functions in silique development. *BraHD-Zip*10, *BraHD-Zip*78 and *BraHD-Zip*57 showed discrepancy in their responses under specific abiotic stress as compared to that in their syntenic orthologous genes. The functional divergences among these duplicate genes can originate from mutations, subfunctionalization and selection for changes in gene dosage (Conant and Wolfe 2008). *B. rapa* have been confirmed to undergo multiple WGD and one WGT events, which is typically followed by biased gene fractionation. These functional innovation of over-retained duplicate genes can cause remarkable morphological variants and enhance environmental adaptability (Project 2011).

Methods

Identification of HD-Zip transcription factors and comparison gene sets

The coding sequences of the 48 *AtHD-Zip* genes were used as a set of reference genes and were retrieved from the Arabidopsis TAIR10 database (http://www.arabidopsis.org) (Lamesch et al. 2012), according to previous research (Henriksson et al. 2005; Prigge et al. 2005; Nakamura et al. 2006; Ciarbelli et al. 2008). Homologous Sequences were retrieved from the BRAD database (http://brassicadb.org/brad/) based on a BLASTp search with e-value $\leq$ 1e−10 and identity $\geq$60% (Cheng et al. 2011). FGENESH (http://linux1.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind) were used for manual re-annotation with parameters optimized for Arabidopsis, following which the sequences were verified in the NCBI database (http://www.ncbi.nlm.nih.gov/) (Duan et al. 2015).

Homologs of *AtHD-Zip*s in *A. trichopoda, V. vinifera, C. papaya*, and *P. trichocarpa* were retrieved from Phytozome v10.1 (http://www.phytozome.net/) (Goodstein et al. 2012).

The data sets of Core eukaryotic genes and random genes were downloaded from CEGMA (http://korflab.ucdavis.edu/Datasets/cegma) (Parra et al. 2007), and were used to select homologs from microsyntenic regions

corresponding to HD-Zips using SynOrths tool in Brassica database (Lou et al. 2012).

Synteny analysis of HD-Zip genes in *A. thaliana* and *Brassica rapa*

Each HD-Zip homolog between *A. thaliana* and *B. rapa* was extracted from the BRAD database (http://brassicadb.org/brad/searchSynteny.php) (Cheng et al. 2012b).

Prediction analysis of protein characterization

The value of theoretical isoelectric point (pI) and molecular weight (Mw) were predicted using Compute pI/Mw software (http://www.expasy.ch/tools/pi_tool.html). SignalP (http://www.cbs.dtu.dk/services/SignalP/) and PredoTar (http://www.genscript.com/wolf-psort.html) were used for signal peptide and subcellular localization prediction of HD-Zip in *B. rapa*.

Ka and Ks analysis

The homologous alignments for protein-coding sequence between *A. thaliana* and *B. rapa* were constructed using a parallel tool-ParaAT (Zhang et al. 2012). Ka (nonsynonymous substitution rates), Ks (synonymous substitution rates) values and Ka/Ks ratio of these alignments were calculated by KaKs_calculator using the method of Nei and Gojobori (Zhang et al. 2006).

Phylogenetic analysis, exon/intron structure and conserved motifs distribution

The identified BraHD-Zip and AtHD-Zip protein sequences were aligned by ClustalW2 (Thompson et al. 1997). Subsequently, the phylogenetic trees were constructed by maximum likelihood (ML) method using MEGA5.0 (Tamura et al. 2011). The exon–intron structure was determined by aligning each cDNA sequence of *BraHD-Zip* with its corresponding genomic sequence in the BRAD database (http://brassicadb.org/brad/). The gene visualized structure was drawn by the Gene Structure Display Server (http://gsds.cbi.pku.edu.cn/index.php) (Hu et al. 2015). To identify the conserved BraHD-Zip motifs and its distribution, the protein sequences were analyzed by Multiple EM for Motif Elicitation program (http://meme-suite.org/meme_4.10.1/tools/meme) (Bailey et al. 2009). The selection parameters were set to motif 20, minimum width 20 and maximum width 150.

Plant material, growth conditions and abiotic stress treatment

The accession Chiifu-401 was used in this experiment. Seeds were sown in small pots containing a turf: vermiculite mixture (1:1) in a controlled-environment phytotron. The parameters was programmed for light 16 h/24 °C, dark 8 h/20 °C and 80% relative humidity (Duan et al. 2015). 200 mM NaCl, 20% polyethylene glycol (PEG)-6000, heat (38 °C) and cold (4 °C) were treated as abiotic stresses at five-leaf stage. The third leaves were collected at 0, 1, 6 and 12 h after treatment.

Organ expression and quantitative RT-PCR analysis

Six tissues of *B. rapa* (Chiifu-401) were prepared for transcriptome sequencing. The FPKM values were calculated to construct expression profiling of *BraHD-Zip*s (Tong et al. 2013). The developmental expression profiling of *AtHD-Zip* was analyzed by AtGenExpress Visualization Tool with mean-normalized values (Schmid et al. 2005; Duan et al. 2015). Data of qPCR were analyzed via $2^{-\Delta\Delta C_T}$ method. The expression profiling was clustered by Cluster 3.0 program (Michael et al. 1998), and subsequently the heatmap was visualized in Java Tree View (http://jtreeview.sourceforge.net/).

Acknowledgements This work was supported by State Key Special Program "Seven Main Crops Breeding" (2016YFD0101701), National Science and Technology Support Program (2012BAD02B01), Independent Innovation of Agricultural Science and Technology Program of Jiangsu Province [CX(13)2006], Science and technology support program of Jiangsu Province (BE2013429), A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions, National Natural Science Foundation of China (No. 31330067).

References

Agalou A, Purwantomo S, Overnas E, Johannesson H, Zhu X, Estiati A, de Kam RJ, Engstrom P, Slamet IH, Zhu Z, Wang M, Xiong L, Meijer AH, Ouwerkerk PB (2008) A genome-wide survey of HD-Zip genes in rice and analysis of drought-responsive family members. Plant Mol Biol 66(1–2):87–103

Ariel FD, Manavella PA, Dezar CA, Chan RL (2007) The true story of the HD-Zip family. Trends Plant Sci 12(9):419–426

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37:W202–208

Baima S, Possenti M, Matteucci A, Wisman E, Altamura MM, Ruberti I, Giorgio M (2001) The Arabidopsis ATHB-8 HD-zip protein acts as a differentiation-promoting transcription factor of the vascular meristems. Plant Physiol 126:643–655

Carabelli M, Turchi L, Ruzza V, Morelli G, Ruberti I (2013) Homeodomain-Leucine Zipper II family of transcription factors to the

limelight: central regulators of plant development. Plant Signal Behav 8 (9):e25447
Chen X, Chen Z, Zhao H, Zhao Y, Cheng B, Xiang Y (2014) Genome-wide analysis of soybean HD-Zip gene family and expression profiling under salinity and drought treatments. PLoS One 9(2):e87156
Cheng F, Liu S, Wu J, Fang L, Sun S, Liu B, Li P, Hua W, Wang X (2011) BRAD, the genetics and genomics database for Brassica plants. BMC Plant Biol 11:136
Cheng F, Wu J, Fang L, Sun SL, Liu B, Lin K, Bonnema G, Wang XW (2012a) Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa*. PLoS One 7 (5):e36442
Cheng F, Wu J, Fang L, Wang X (2012b) Syntenic gene analysis between *Brassica rapa* and other Brassicaceae species. Front Plant Sci 3:198
Ciarbelli AR, Ciolfi A, Salvucci S, Ruzza V, Possenti M, Carabelli M, Fruscalzo A, Sessa G, Morelli G, Ruberti I (2008) The Arabidopsis homeodomain-leucine zipper II gene family: diversity and redundancy. Plant Mol Biol 68(4–5):465–478
Conant GC, Wolfe KH (2008) Turning a hobby into a job: how duplicated genes find new functions. Nat Rev Genet 9(12):938–950
Duan W, Song X, Liu T, Huang Z, Ren J, Hou X, Du J, Li Y (2015) Patterns of evolutionary conservation of ascorbic acid-related genes following whole-genome triplication in *Brassica rapa*. Genome Biol Evol 7(1):299–313
Freeling M, Thomas BC (2006) Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. Genome Res 16(7):805–814
Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40(D1):D1178–D1186
Henriksson E, Olsson AS, Johannesson H, Johansson H, Hanson J, Engstrom P, Soderman E (2005) Homeodomain leucine zipper class I genes in Arabidopsis. Expression patterns and phylogenetic relationships. Plant Physiol 139(1):509–518
Hu R, Chi X, Chai G, Kong Y, He G, Wang X, Shi D, Zhang D, Zhou G (2012) Genome-wide identification, evolutionary expansion, and expression profile of homeodomain-leucine zipper gene family in poplar (*Populus trichocarpa*). PLoS One 7(2):e31149
Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G (2015) GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics 31(8):1296–1297
Kubo H, Peeters AJ, Aarts MG, Pereira A, Koornneef M (1999) ANTHOCYANINLESS2, a homeobox gene affecting anthocyanin distribution and root development in Arabidopsis. Plant Cell 11(7):1217–1226
Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, Karthikeyan AS, Lee CH, Nelson WD, Ploetz L, Singh S, Wensel A, Huala E (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res 40(D1):D1202–D1210
Lim SD, Yim WC, Moon JC, Kim DS, Lee BM, Jang CS (2010) A gene family encoding RING finger proteins in rice: their expansion, expression diversity, and co-expressed genes. Plant Mol Biol 72(4–5):369–380
Lou P, Wu J, Cheng F, Cressman LG, Wang X, McClung CR (2012) Preferential retention of circadian clock genes during diploidization following whole genome triplication in *Brassica rapa*. Plant Cell 24(6):2415–2426
Michael B, Eisen PTS, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci 95(25):14863–14868
Nakamura M, Katsumata H, Abe M, Yabe N, Komeda Y, Yamamoto KT, Takahashi T (2006) Characterization of the class IV homeodomain-Leucine Zipper gene family in Arabidopsis. Plant Physiol 141(4):1363–1375
Olsson A, Engström P, Söderman E (2004) The homeobox genes ATHB12 and ATHB7 encode potential regulators of growth in response to water deficit in Arabidopsis. Plant Mol Biol 55(5):663–677
Park MY, Kim SA, Lee SJ, Kim SY (2013) ATHB17 is a positive regulator of abscisic acid response during early seedling growth. Mol Cells 35(2):125–133
Parra G, Bradnam K, Korf I (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23(9):1061–1067
Prigge MJ, Otsuga D, Alonso JM, Ecker JR, Drews GN, Clark SE (2005) Class III homeodomain-leucine zipper gene family members have overlapping, antagonistic, and distinct roles in Arabidopsis development. Plant Cell 17(1):61–76
Project BrGS (2011) The genome of the mesopolyploid crop species *Brassica rapa*. Nat Genet 43(10):1035–1039
Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU (2005) A gene expression map of *Arabidopsis thaliana* development. Nat Genet 37(5):501–506
Sessa G, Morelli G, Ruberti I (1997) DNA-binding specificity of the homeodomain-leucine zipper domain. J Mol Biol 274(3):303–309
Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28(10):2731–2739
Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25(24):4876–4882
Tong CB, Wang XW, Yu JY, Wu J, Li WS, Huang JY, Dong CH, Hua W, Liu SY (2013) Comprehensive analysis of RNA-seq data reveals the complexity of the transcriptome in *Brassica rapa*. BMC Genomics 14(1):689
Torrent BJ, Martret SM, Brandt R, Musielak T, Palauqui JC, Martinez-Garcia JF, Wenkel S (2012) ATHB4 and HAT3, two class II HD-ZIP transcription factors, control leaf development in Arabidopsis. Plant Signal Behav 7(11):1382–1387
Victor M, Zúñiga M, Martinez NM, Folter SD (2012) The class II HD-ZIP JAIBA gene is involved in meristematic activity and important for gynoecium and fruit development in Arabidopsis. Plant Signal Behav 7(11):1–3
Wang Y, Henriksson E, So¨derman E, Henriksson KN, Sundberg E, Engstro¨m P (2003) The Arabidopsis homeobox gene, ATHB16, regulates leaf development and the sensitivity to photoperiod in Arabidopsis. Dev Biol 264:228–239
Zhang Z, Li J, Zhao X-Q, Wang J, Wong GK-S, Yu J (2006) KaKs_ Calculator: calculating Ka and Ks through model selection and model averaging. Genomics Proteomics Bioinform 4(4):259–263
Zhang Z, Xiao J, Wu J, Zhang H, Liu G, Wang X, Dai L (2012) ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments. Biochem Biophys Res Commun 419(4):779–781
Zhao Y, Zhou YQ, Jiang HY, Li XY, Gan DF, Peng XJ, Zhu SW, Cheng BJ (2011) Systematic analysis of sequences and expression patterns of drought-responsive members of the HD-Zip Gene family in maize. PLoS One 6(12):e28488