Plant Physiology and Biochemistry 71 (2013) 11–21

Contents lists available at SciVerse ScienceDirect

Plant Physiology and Biochemistry

journal homepage: www.elsevier.com/locate/plaphy

Research article

Genome-wide analysis and identification of stress-responsive genes of the NAM–ATAF1,2–CUC2 transcription factor family in apple

Hongyan Su [a,c,1], Shizhong Zhang [b,c,1], Xiaowei Yuan [b], Changtian Chen [b], Xiao-Fei Wang [b,c], Yu-Jin Hao [b,c,*]

[a] College of Agriculture, Ludong University, Yantai, Shandong 264025, PR China
[b] National Key Laboratory of Crop Biology, Shandong Agricultural University, Tai-An, Shandong 271018, PR China
[c] MOA Key Laboratory of Horticultural Crop Biology and Germplasm Innovation, College of Horticulture Science and Engineering, Shandong Agricultural University, Tai-An, Shandong 271018, PR China

ARTICLE INFO

Article history:
Received 15 February 2013
Accepted 20 June 2013
Available online 4 July 2013

Keywords:
Genome-wide analysis
NAC transcription factor
Gene family
Stress-responsive genes
Malus domestica

ABSTRACT

NAC (NAM, ATAF1,2, and CUC2) proteins constitute one of the largest families of plant-specific transcription factors. To date, little is known about the *NAC* genes in the apple (*Malus domestica*). In this study, a total of 180 *NAC* genes were identified in the apple genome and were phylogenetically clustered into six groups (I–VI) with the *NAC* genes from *Arabidopsis* and rice. The predicted apple *NAC* genes were distributed across all of 17 chromosomes at various densities. Additionally, the gene structure and motif compositions of the apple *NAC* genes were analyzed. Moreover, the expression of 29 selected apple *NAC* genes was analyzed in different tissues and under different abiotic stress conditions. All of the selected genes, with the exception of four genes, were expressed in at least one of the tissues tested, which indicates that the *NAC* genes are involved in various aspects of the physiological and developmental processes of the apple. Encouragingly, 17 of the selected genes were found to respond to one or more of the abiotic stress treatments, and these 17 genes included not only the expected 7 genes that were clustered with the well-known stress-related marker genes in group IV but also 10 genes located in other subgroups, none of which contains members that have been reported to be stress-related. To the best of our knowledge, this report describes the first genome-wide analysis of the apple *NAC* gene family, and the results should provide valuable information for understanding the classification and putative functions of this family.

1. Introduction

The plant-specific NAC (NAM, ATAF1/2, and CUC2) proteins constitute one of the largest TF (transcription factor) families and are characterized by a well-conserved N-terminal NAC domain [1,2]. Based on its motif distribution, the NAC domain, which comprises nearly 160 amino acid residues, can be divided into five subdomains (A–E). The highly conserved subdomains C and D may be responsible for binding to DNA, and subdomain A may be involved in homo- and heterodimerization, whereas the divergent subdomains B and E may be implicated in the functional diversity of the NAC proteins [3,4]. The TRR (transcriptional regulation region) of the NAC TFs, which are generally found in the highly divergent C terminus, can confer regulation diversity in the transcriptional activation activity. A common feature of the TRR is the frequent occurrence of group-specific motifs rich in serine and threonine, proline, glutamine, or acidic residues. At least ten motifs were identified in the TRR of the rice (*Oryza sativa*) NAC proteins, and these motifs have been found to be conserved for a given subgroup of NAC subfamilies but vary across different subfamilies [5].

The NAC TFs have been shown to regulate a number of biological processes, including shoot apical meristem formation and maintenance [6], floral development [7], control of flowering induction in response to stresses [8,9], embryo development [10], hormone signaling [11], and regulation of secondary cell wall synthesis [12]. In particular, the NAC TFs have received much attention as regulators in

Abbreviations: BLAST, the Basic Local Alignment Search Tool; GWD, genome-wide duplication; HMM, Hidden Markov Model; NAC, NAM, ATAF and CUC, transcription factor; MEGA, Molecular Evolutionary Genetics Analysis; MUSCEL, Multiple Sequence Comparison by Log-Expectation; NJ, the neighbor-joining; SMART, a Simple Modular Architecture Research Tool; TF, transcription factor; TRR, transcriptional regulation region; MEME, Multiple Em for Motif Elicitation; Apple GFDB database, Apple Gene Function and Gene Family Database; GDR, Genome Database for Rosaceae; Pfam, Protein family; GSDS, Gene Structure Display Server; pIs, isoelectric points; NCBI, National Center for Biotechnology Information.

* Corresponding author. National Key Laboratory of Crop Biology, College of Horticulture Science and Engineering, Shandong Agricultural University, Tai-An, Shandong 271018, PR China. Tel.: +86 538 8246692.
E-mail address: haoyujin@sdau.edu.cn (Y.-J. Hao).

[1] The two authors contributed equally to this work.

0981-9428/$ – see front matter © 2013 Elsevier Masson SAS. All rights reserved.
http://dx.doi.org/10.1016/j.plaphy.2013.06.022

both biotic and abiotic stress signaling pathways, and some of these have been considered potential targets for the engineering of plant tolerance [2,3,13–15]. For instance, *ANAC019*, *ANAC055*, and *RD26*/*ANAC072* are well-characterized stress-responsive members of the *Arabidopsis* NAC gene family, and the transgenic plants overexpressing these genes exhibit improved stress tolerance to different degrees compared with the wild type [3,11,16,17]. In rice, the *NAC* gene *SNAC1* is involved in the response to drought and salinity stresses, and its overproduction results in a significantly enhanced drought tolerance during anthesis under field conditions [18]. Moreover, the overexpression of *SNAC2*/*OsNAC6*, *OsNAC10*, *OsNAC045*, and *OsNAC063* enhances tolerance to multiple abiotic stresses [19–22].

Apple (*Malus domestica*) is one of the most widely cultivated fruit trees and is the most economically important woody plant in temperate regions. In contrast with the intensive research on NAC TFs in both model and crop plants, such as *Arabidopsis*, rice, soybean (*Glycine max*), there are only very limited reports on the characterization of the NAC TFs in apple [5,23,24]. Recently, the draft genome sequence of apple has been decoded, and this has provided an excellent opportunity for genome-wide analyses of all of the genes belonging to specific gene families [25]. Genome-wide analyses of the *RING finger* gene family, the *DREB* gene family, the *dehydrin* gene family, and the *Hsf* gene family have been reported in apple [26–29]. However, no genome-wide information on the apple *NAC* gene family is currently available.

Given the importance of the NAC TFs in diverse biological and physiological processes and their potential application in the development of improved stress-tolerant transgenic plants, we performed the first systematic analysis of the apple NAC TF family in the present study. The chromosome location, gene structure, and protein motifs of the putative *NAC* genes predicted by genome-wide surveys of the apple genomic sequences were carefully analyzed. Additionally, the putative apple *NAC* genes were subjected to phylogenetic analyses with their *Arabidopsis* and rice counterparts. These comparisons enabled the identification of gene orthologs and clusters of orthologous groups that can be functionally characterized further. Furthermore, we paid additional attention to the apple *NAC* genes associated with abiotic stresses. Given that the tolerance of apple tree to abiotic stresses mainly depends on the rootstock, we performed the expression analysis of the selected *NACs* in *Malus hupehensis*, an excellent apple rootstock widely used for grafting in China, to identify some stress-related candidate genes. To the best of our knowledge, this report presents the first genome-wide analysis of the apple NAC family, and the results will provide valuable information for understanding the classification and putative functions of apple NACs. Ultimately, these findings will lead to potential improvements in the stress resistance in apple through genetic engineering approaches.

2. Results

2.1. Identification and genome distribution of the NAC gene family in apple

To identify the NAC TFs-encoding genes in the apple genome, BLASTP searches of the apple genome database (GDR database) using 100 *Arabidopsis NACs* and 135 rice *NACs* as queries were first performed. Then, the HMM of the SMART and Pfam tools were exploited as queries to confirm the putative NAC genes. Finally, 180 typical *NAC* genes containing full ORFs were further manually analyzed using the InterProScan program to confirm the presence of the NAM domain and were used for further analysis below. To distinguish the remaining *NACs*, we provisionally named them *MdNAC1* through *MdNAC180* based on the order of the corresponding chromosomal locations identified from the apple genome browser (Supplementary Material 1). The identified *MdNAC* genes encode proteins that range from 146 (*MdNAC134*) to 1024 (*MdNAC136*) aa (amino acids) in length with an average of 372 aa and exhibit pIs that range from 4.58 (*MdNAC115*) to 9.52 (*MdNAC43*).

2.2. Genome distribution and gene duplication of the NAC gene family in apple

To determine the genomic distribution of the *MdNAC* genes, the DNA sequence of each *MdNAC* gene was used to search the apple genome database using BLASTN. Among these 180 *MdNAC* genes, 166 genes could be mapped on chromosomes 1 through 17, and 14 genes (*MdNAC167–MdNAC180*) could not be conclusively mapped on any chromosome. Although each of the 17 apple chromosomes contained some *MdNAC* genes, the distribution appeared to be uneven (Fig. 1a). The gene density per Chr (chromosome) ranged from 1.11% (2 *MdNAC* genes on Chr 4) to 11.67% (21 *MdNAC* genes on Chr 3), and relatively low numbers of *MdNAC* genes were observed in some chromosomes, including 5 each on Chrs 2, 8, and 17, and 4 each on Chrs 9 and 12 (Fig. 1b).

In this study, the large-scale segmental duplication events were investigated to elucidate the mechanism behind the expansion of the *MdNAC* gene family, which was thought to have occurred during the evolutionary process. In this study, multiple pairs linked each of at least 15 potential chromosomal segmental duplications (Fig. 1a, pairs of bars in blue areas in web version), such as the large sections of Chrs 3 and 11 and Chrs 4 and 12. It has been reported that a relatively recent (>50 million years ago) GWD (genome-wide duplication) has resulted in the transition of 9 ancestral chromosomes to 17 chromosomes in the *Pyreae* tribe [25]. Consistent with this finding, there were at least 70 *MdNAC* genes found on the GWD segment, which indicates a clear paralogous pattern of *NAC* gene divergence through gene duplication in apple.

2.3. Gene structure and protein motif analysis of the NAC gene family in apple

To gain further insights into the structural diversity of *MdNAC* genes, we analyzed the exon/intron organization in the coding sequences of the individual *MdNAC* genes in apple. As illustrated in Fig. 2b, among the 180 *MdNAC* genes, 17 had no intron, and the other genes had at least one intron. *MdNAC69*, which contained 14 introns, possessed the most introns. A separate phylogenetic tree was generated from the complete protein sequences of all of the *MdNAC* genes, which divided the *NAC* genes into 17 subgroups (Fig. 2a). Overall, the most closely related members in the same subfamilies share a similar exon/intron structure in terms of intron number and exon length. For instance, the majority of the *MdNAC* genes in subgroups I, N, P, and Q contained two or three introns, whereas all of the members in subgroup G possessed one intron with the exception of two members which harbored no intron. In contrast, the members of subgroups B, C, E, K, L, and O displayed a large variability in either the number or the distribution of introns.

To further reveal the diversification of the *MdNAC* genes, putative motifs were predicted by the program MEME, and 19 distinct motifs were identified. As expected, most of the closely related members in the phylogenetic tree exhibited common motif compositions, which suggests that there are functional similarities among the NAC proteins within the same subgroup (Fig. 2c).