Gene 465 (2010) 30–44

Contents lists available at ScienceDirect

Gene

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

Methods paper

Genome-wide analysis of NAC transcription factor family in rice

Mohammed Nuruzzaman [a], Ramaswamy Manimekalai [a], Akhter Most Sharoni [a], Kouji Satoh [a], Hiroaki Kondoh [a], Hisako Ooka [b], Shoshi Kikuchi [a,*]

[a] Plant Genome Research Unit, Division of Genome and Biodiversity Research, National Institute of Agrobiological Sciences (NIAS), Tsukuba, Ibaraki 305-8602, Japan
[b] Kurume National College of Technology, Komorino 1-1-1, Kurume, Fukuoka 830-8555, Japan

ARTICLE INFO

Article history:
Accepted 16 June 2010
Available online 26 June 2010

Received by C. Feschotte

Keywords:
Rice
Phylogenetic analysis
Microarray
Stress

ABSTRACT

We investigated 151 non-redundant *NAC* genes in rice and 117 in Arabidopsis. A complete overview of this gene family in rice is presented, including gene structures, phylogenies, genome localizations, and expression profiles. We also performed a comparative analysis of these genes in rice and Arabidopsis. Conserved amino acid residues and phylogeny construction using the NAC conserved domain sequence suggest that *OsNAC* gene family was classified broadly into two major groups (A and B) and sixteen subgroups in rice. We presented more specific phylogenetic analysis of OsNAC proteins based on the DNA-binding domain and known gene function, respectively. Loss of introns was observed in the segmental duplication. Homologous, paralogous, and orthologous searches of rice and Arabidopsis revealed that the major functional diversification within the *NAC* gene family predated the divergence of monocots and dicots. The chromosomal localizations of *OsNAC* genes indicated nine segmental duplication events involving 18 genes; 32 non-redundant *OsNAC* genes were involved in tandem duplications. Expression levels of this gene family were checked under various abiotic stresses (cold, drought, submergence, laid-down submergence, osmotic, salinity and hormone) and biotic stresses [infection with rice viruses such as RSV (rice stripe virus) and RTSV (rice tungro spherical virus)]. Biotic stresses are novel work and increase the possibilities for finding the best candidate genes. A preliminary search based on our microarray (22K and 44K) data suggested that more than 45 and 26 non-redundant genes in this family were upregulated in response to abiotic and biotic stresses, respectively. All of the genes were further investigated for their stress responsiveness by RT-PCR analysis. Six genes showed preferential expression under both biotic RSV and RTSV stress. Eleven genes were upregulated by at least three abiotic treatments. Our study provides a very useful reference for cloning and functional analysis of members of this gene family in rice.

1. Introduction

A number of transcription factors are involved to regulate the gene expression in living organism. As part of their regulatory function, transcription factors interact with plant-specific DNA sequences. Knowledge of the structure of the DNA-binding domains of transcription factors is essential for understanding their function and mechanism. The *NAC* gene family name was derived from the names of three transcription factors: (i) NAM (no apical meristem, Petunia), (ii) ATAF1–2, and (iii) CUC2 (cup-shaped cotyledon, Arabidopsis), all of which have the same DNA-binding domain (Souer et al., 1996; Aida et al., 1997). *NAC* genes encode plant-specific transcriptional regulators that constitute a large transcription factor family in plants (Olsen et al., 2005); this family was first identified by mutations (Souer et al., 1996; Aida et al., 1997). Transcription factor families are expressed at much higher levels in plants than in animals. In the plant kingdom, more than 50 families of different transcription factors have been identified by sequence analyses of model species such as rice (Xiong et al., 2005), and numerous reports suggest that transcription factors and *cis*-acting elements are involved in almost all aspects of cellular activity as part of their related roles in promoting gene expression (Xiong et al., 2005).

NAC protein family members are highly conserved at the N-terminal NAC binding domain and have a highly variable C-terminal domain that plays a major role in the regulation of transcription (Olsen et al., 2005). This unstable C-terminal domain of NAC proteins generally operates as a functional domain and acts as a transcriptional activator or repressor (Tran et al., 2004; Hu et al., 2006; Kim et al., 2007a). This variable C-terminal domain is very large and also has protein binding activity. Kim et al. (2007a) reported that the C-terminal domain of Arabidopsis calmodulin-binding (CB) NAC can bind with calmodulin proteins.

There are 105 redundant putative *NAC* genes in Arabidopsis (Ooka et al., 2003), and 140 putative *NAC* or *NAC*-like genes in rice (Fang

Abbreviations: NAC, NAMATAF1/2 and CUC transcription factor; RSV, Rice stripe virus; RTSV, Rice tungro spherical virus; RT-PCR, Reverse transcription-polymerase chain reaction; BAC, Bacterial artificial chromosome; Os, *Oryza sativa*; At, *Arabidopsis thaliana*; GA, Gibberellic acid; ABA, Abscisic acid.

* Corresponding author. Tel./fax: +81 29 838 7007.
E-mail address: skikuchi@nias.affrc.go.jp (S. Kikuchi).

0378-1119/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2010.06.008

et al., 2008). The NAC transcription factors appeared to control of biochemical and molecular pathways that can save plants under different stress conditions. The NAC transcription factors are multi-functional proteins with various roles in the plant life cycle, such as maintenance of the shoot apical meristem (Souer et al., 1996; Kim et al., 2007b), cotyledon development (Aida et al., 1997), lateral root development (He et al., 2005), flower formation (Sablowski and Meyerowitz, 1998), hormone signaling (Greve et al., 2003), response to pathogen infection (Xie et al., 1999; Ren et al., 2000; Olsen et al., 2005; Nakashima et al., 2007), plant organ senescence (Liu et al., 2009), embryo development (Duval et al., 2002), response to different abiotic stresses (Tran et al., 2004; He et al., 2005; Hu et al., 2006; Bhatnagar-Mathur et al., 2007; Nakashima et al., 2007; Yoo et al., 2007), formation of secondary walls (Zhong et al., 2007), cell division (Kim et al., 2006), fiber development (Ko et al., 2007), seed development (Sperotto et al., 2009), and senescence (Uauy et al., 2006). In addition, expression of a large array of genes is induced by a number of stress conditions, meaning that numerous proteins are produced to join the pathways that lead to synergistic improvement of stress tolerance (Seki et al., 2003).

NAC transcription factor production is induced by drought, as was first shown in Arabidopsis. Overexpression of three *Arabidopsis thaliana* (*At*) *NAC* genes (*ANAC019*, *ANAC055*, and *ANAC072*) in transgenic plants increased the stress tolerance of the plants and altered the expression of drought, salinity, and low-temperature-stress-inducible genes (Tran et al., 2004). Furthermore, some researchers have reported important functions of these genes in abiotic and biotic stress resistance during the plant life cycle. One field study showed that introduction of *SNAC1* into rice increased drought tolerance in the transgenic plants and gave 22% to 34% greater seed set than in the negative control population upon exposure to severe drought during flowering (Hu et al., 2006). *SNAC1* is also involved in cold and salt tolerance (Hu et al., 2006). *Oryza sativa* (*Os*) *NAC6*/*SNAC2* and *OsNAC10* are drought-tolerance gene that improves the expression of several protein-encoding genes under stress conditions (Nakashima et al., 2007; Jeong et al., 2010) and belongs to the ATAF subfamily (Kikuchi et al., 2000; Ooka et al., 2003). In Arabidopsis, *AtNAC2* expression is induced by salt and abscisic acid and amends salt tolerance and root development (He et al., 2005). Hegedus et al. (2003) have identified nine *NAC* genes that are upregulated in *Brassica napus* (rapeseed) by cold temperature. In wheat (*Triticum aestivum*) *NAC* gene, (*NAM-B1*) played an important role to move nutrient remobilization from leaves to developing grains (Uauy et al., 2006). *TaNAC4* gene functions as a transcriptional activator involved in wheat response to abiotic and biotic stresses (Xia et al., 2010). Lin et al. (2007) found that *OsNAC19* transcript was elevated by the infection of *Magnaporthe grisea*, suggesting that *OsNAC19* involved in rice defense response to *M. grisea* infection. NAC proteins are also involved in response to viral infection during vegetative development of plants (Xie et al., 1999; Ren et al., 2000).

Completion of the high-quality sequencing of the rice genome (International Rice Genome Sequencing Project 2005) has provided an excellent opportunity for genome-wide analysis of all the genes belonging to specific gene families. In rice, only a few *NAC* genes have been characterized, and the functions of most of them remain to be determined. Here, we identified 151 *OsNAC* genes in rice by database searches and classified these genes according to reported genes. The subgroups are more specific than previous work and biotic stresses are new work in this present study. *OsNAC* genes play an important role in the crosstalk of different kinds of stresses signaling. We analyzed the phylogenetic relationships of the *NAC* genes in rice and Arabidopsis, as well as the segmental and tandem duplications and exon and intron structures of *OsNAC* genes. We also studied the expression intensities of *OsNAC* genes under different abiotic and biotic stresses from our 22K and 44K microarray data. The data generated should be very helpful in studies of the biological functions of each *OsNAC* gene.

2. Materials and methods

2.1. Collection and classification of OsNAC gene family members

To identify members of the rice (*Oryza sativa* L. subsp. *japonica* cv.) NAC transcription factor gene family, multiple database searches were performed. The Database of Rice Transcription Factors (DRTF, http://drtf.cbi.pku.edu.cn; Gao et al., 2006), rice genome annotation (Machigan State University, MSU; http://rice.plantbiology.msu.edu/), Rice Transcription Factor Database (Rice-TFDB, http://ricetfdb.bio.uni-potsdam.de/v2.1/; Riano-Pachon et al., 2007), National Centre for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/), and the Knowledge-Based Oryza Molecular Biological Encyclopedia (KOME; http://www.cdna01.dna.affrc.go.jp/cDNA) were used to search for members of the *NAC* gene family. We then further explored the MSU database to gather a more complete collection of putative *NAC* genes in rice by using keywords (NAC, NAM) and a domain (PF02365) search. The BLASTP and TBLASTN search parameters of the three databases (MSU, NCBI, and KOME) were set as follows: maximum target sequences 350, and expected value less than 10. SMART (http://smart.embl-heidelberg.de/) and Pfam database (http://pfam.sanger.ac.uk/) searches were used to confirm and classify each predicted *OsNAC* gene. Information about gene structure, transcripts, length, chromosomal localization, full-length cDNA, BAC accessions for each gene, and the characteristics of the corresponding proteins were procured from MSU, KOME, and GRAMENE (Liang et al., 2008). Exon and intron structures were investigated by using the National Center of Plant Gene Research (NCPGR, http://gbrowse.ncpgr.cn/cgi-bin/gbrowse/japonica/) database.

2.2. NAC genes from Arabidopsis and other species

Various database searches were performed to collect data on all members of the Arabidopsis *NAC* gene family. We used the BLAST programs (TBLASTN and BLASTP) available in The Arabidopsis Information Resource (TAIR), MSU Arabidopsis databases and the NCBI Arabidopsis genome database. As a query sequence, we first used the amino acid sequence of the NAM domain from the rice NAC. To increase the extent of the database search results, we also performed a position-specific iterated BLAST search (Altschul et al., 1997) against the Arabidopsis database on the NCBI website. To confirm completion of the collection we also performed database searches using the amino acid sequences of the NAM domain of some members of the Arabidopsis NAC family as query sequences. We indentified non-redundant 117 *AtNAC* genes from database searches. We used the Plant Transcription Factor Data Base (PTFDB, http://plntfdb.bio.uni-potsdam.de/v3.0/fam_mem.php?family_id=NAC) for other species (sorghum, maize, and poplar).

2.3. Phylogenetic analysis and sequence alignment

Phylogenetic trees were constructed by using OsNAC domain sequences, and an unrooted tree was generated by ClustalX version 1.83 (Thompson et al., 1997) by the neighbor-joining method (Saitou and Nei, 1987) and bootstrap analysis (1000 replicates). The tree was analyzed and displayed by using MEGA software version 4 (Tamura et al., 2007). Another two unrooted trees were constructed by the same method to align OsNAC and AtNAC protein sequences and to align the NAC protein sequences of five monocot and dicot species (rice, sorghum, maize, poplar, and Arabidopsis). We defined two proteins with 100% support in the same species as homologous proteins and two with 100% support in different species as orthologous proteins. Multiple sequence alignments were performed with ClustalX version 1.83 (Thompson et al., 1997).