The handle http://hdl.handle.net/1887/37134 holds various files of this Leiden University dissertation.

**Author:** Zhang, Yu  
**Title:** Functional analysis of Dof transcription factors controlling heading date and PPDK gene expression in rice  
**Issue Date:** 2015-12-22
Chapter 1

General Introduction

Y. Zhang\textsuperscript{1,2,3}, M. Wang\textsuperscript{1,4}, Z. Zhu\textsuperscript{2}, and P.B.F. Ouwerkerk\textsuperscript{1}

\textsuperscript{1} Department of Molecular and Developmental Genetics, Institute of Biology (IBL), Leiden University, Sylvius Laboratory, Sylviusweg 72, P.O. Box 9505, 2300 RA Leiden, The Netherlands

\textsuperscript{2} State Key Laboratory of Plant Genomics and National Center for Plant Gene Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, No.1 West Beichen Road, Chaoyang District, Beijing 100101, China

\textsuperscript{3} Graduate School of the Chinese Academy of Sciences, 100049 Beijing, China

\textsuperscript{4} SU BioMedicine/TNO Quality of Life, Zernikedreef 9, P.O. Box 2215, 2301 CE Leiden, The Netherlands
Introduction

Four millennia ago, people completed the first steps in the domestication of all major cereal crops including rice, barley, wheat and maize (Doebley et al. 2006). Rice amongst cereal grasses like maize, millet, barley, sorghum and wheat is the key food for more than half of mankind. The goal of modern rice breeding is to improve cultivars for yield and yield stability under adverse conditions and to improve resistance against biotic stresses. Rice yield is determined by several factors such as the number of tillers per plant, the number of grains per panicle, characteristics of grain filling and grain size. Since the 1960s, the productivity of rice has been largely improved by development of inbred cultivars, which was called the Green Revolution. In the 1970s, Longping Yuan developed the Chinese super-hybrid rice, which increased yield by 30% compared to the parental lines. This development is sometimes called the Second Green Revolution. Heterosis, or hybrid vigor, refers to the improved yield and performance of a hybrid offspring. The molecular genetics and biology behind this biological phenomenon needs to be further investigated in order to reveal the responsible genes.

Rice is an excellent system for plant molecular research as it represents a modest size genome of 430 Mb, containing about 50,000 genes (Goff et al. 2002; Tyagi et al. 2004). Many rice genomics studies have been initiated in the last decade, some of which took place as international projects like the rice genome research programme (RGP). The rice genome has been completely sequenced (Sasaki et al. 2005) for both the japonica (Goff et al. 2002) and indica (Yu et al. 2002) subtypes and recently low-coverage sequencing of a collection of 3,000 cultivars went public (Li et al. 2014). This is excellent material for allele mining and genome wide association studies. Also a large number of T-DNA insertion and transposon-tagged mutants have been produced (Miyao et al. 2007), and global gene expression profiles have been obtained using SAGE (Serial Analysis of Gene Expression) or microarray (Bao et al. 2005; Li et al. 2006). All this will provide useful information and act as a resource platform for further studies on the functional genomics level. However, for the majority of rice genes no functions have yet been demonstrated (Zhang et al. 2006). Several genetics studies especially in rice and maize revealed important roles for transcription factors (TFs) and also other regulatory proteins in QTLs determining plant architecture and yield, which were the bases for crop domestication in the past millenia. This chapter will briefly review the recent advances in understanding of the molecular mechanism and roles that TFs have in crop domestication such as apical dominance, seed shattering, tillering, grain size and flowering time.
Flowering is the transition from vegetative stage to reproductive stage, which is important during the rice life cycle. Both environmental signals and endogenous signals affected the transition. Transcription factors play important roles during the life cycle.

Flowering time is the transition from vegetative to reproductive development; it represents one of the major phase changes during the life cycle of a crop such as rice (Figure 1). This transition is initiated by both endogenous and environmental signals. In rice, Heading date genes play roles as endogenous factors controlling the flowering time.

**General features of plant transcription factors**
Transcription factors (TFs) are proteins or are part of protein complexes, which regulate transcription by binding to specific cis-elements in the promoters of target genes and affect the initiation of gene transcription (Ramirez and Basu 2009) and thereby control biological processes. A typical plant TF contains a DNA-binding domain, a transcription-regulation domain and a nuclear localization signal, while some of them also have domains enabling protein-protein interactions (Liu et al. 1999). TFs interact with DNA bases in cis-acting elements through their DNA binding domain, which in many cases has a basic character because of their amino acid content. Based on the DNA binding domain, the TFs can be divided into different families. For example, there are 57 TF families identified in sorghum and 58 in maize (Perez-Rodriguez et al. 2010) and 58 in Arabidopsis (Jin et al. 2014). TFs not only can bind to DNA but also regulate gene expression on the transcriptional level. Some TFs have a transcription repression domain by which they can repress target gene expression, whereas others have a transcription...
activation domain by which they can activate target gene expression. Another aspect of TF functionality is their intracellular localization. Translation of a TF-coding mRNA occurs in the cytoplasm, but transcription occurs in the nucleus, so the TFs have to enter the nucleus in order to be able to regulate the transcription of target genes. For entering the nucleus, TFs usually contain an NLS (Nuclear Localization Signal) which is characterized by a core peptide enriched in arginine and lysine (Boulikas 1994).

Many plant TFs form heterodimers or homodimers and as a consequence of the dimer interaction show more specificity in DNA recognition and functionality (Lu et al. 2002). Some TFs have oligomerization domains which are normally adjacent to or overlap with the DNA binding domain (Katagiri et al. 1992). The oligomerization domains and DNA binding domains display a tertiary structure which determines critical aspects of TF activity (Guiltinan and Miller 1994). The variations in oligomerization increase the versatility of the transcription machinery and of modulation of gene expression. Hence, studies of plant TFs are important not only for expanding our knowledge of plants but also for future challenges of generating genetically modified crops with superior characters.

**Transcription factors play important roles in crop domestication**

Over the past decade, several TFs have been identified, which control some of the important morphological changes associated with crop domestication. For instance, one of the biggest differences between maize and its progenitor, teosinte, is the difference in apical dominance (Doebly et al. 1997; Doebly 2004). Modern maize cultivars have a single stalk and one ear on the top of the plant, but wild type non-domesticated teosinte has multiple stalks and long branches. TEOSINTE BRANCHED1 (Tb1), a member of the TCP TF family, largely controls these differences between maize and teosinte. Tb1 can repress the outgrowth of the axillary meristems and branch elongation via a negative effect on the cell cycle (Dorweiler and Doebly 1997; Clark 2005; Camus-Kulandaivelu et al. 2008; Tsiantis 2011). Another well-described example of the role of a TF in crop domestication is Tga1. This is a member of the squamosa-promoter binding protein (SBP) TF family and was identified as a TF that played an important role in maize domestication (Wang et al. 2005). Tga1 might sit at the top of a cascade of transcription regulators, because some of the members of this family can regulate the expression of MADS TFs (Cardon et al. 1999). As major regulator of other TFs, the Tga1 gene affects diverse traits such as cell lignification, three dimensional organ growth, and organ size and silica deposition (Dorweiler and Doebly 1997). Other SBP genes with important roles in QTLs for yield components in rice are OsSPL14 and OsSPL16 which respectively control grain number and tiller formation.
(Miura et al. 2010; Liang et al. 2014; Wang et al. 2012). Likely these genes were also important for crop domestication.

Another important trait in cereal domestication is shattering of seeds. Normally seeds shatter easily from the rice panicle or the wheat spike but for agricultural purposes it is required that they stay attached during harvest. A number of QTLs controlling this process have been identified. Sh4 and qSH1 are two key genes required for reduced seed shattering during rice domestication. Sh4 encodes a protein with homology to Myb3, which is a TF from the MYB domain family and controls whether the seed fall off the plant as in wild rice or adhere to the plant like in cultivated types. Transgenic experiments showed that a single amino acid change converts rice from shattering to non-shattering types. The decrease in expression of the cultivated allele compared to the wild type allele is also important in the domestication (Li et al. 2006). qSH1 is a BEL1-type homeobox gene (Zhou et al. 2012). The single nucleotide change in cultivated rice interferes with a cis-regulatory element in the promoter required for the expression of qSH1 in the abscission layer at the base of the rice grain, which is needed for the grain to break away from the axis of the panicle (Konishi et al. 2006; Zhou et al. 2012).

In wheat, Q affects a range of traits including the tendency of the spike to shatter, the tenacity of the chaff surrounding the grain, and whether the spike is elongated as in wild wheat or compact like the cultivated forms after wheat domestication (Simons et al. 2006)

Tillering is also an important agronomic trait for determining rice yield, because it greatly affects the panicle number which is a key component of rice production and is one of three main yield components (Sakamoto and Matsuoka 2008; Xing and Zhang 2010). MOC1 (MONOCULM1) was identified as important gene in the control of rice tillering. The moc1 mutant has only one main culm without any other tillers because of a defect in the formation of tiller buds. MOC1 does not code for a TF but encodes a putative GRAS family protein and is expressed mainly in the axillary buds and functions to initiate axillary buds and to promote their outgrowth (Li et al. 2003). DLT (DWARF AND LOW-TILLERING) also encodes a GRAS family protein and the knockdown mutant showed a reduction in tiller numbers associated with altered expression of several brassinosteroid signalling genes (Li et al. 2010; Tong et al. 2012). Recently, it was found that OsMADS57 interacts with OsTB1 (TEOSINTE BRANCHED1) and targets D14 (Dwarf14) to control tillering (Guo et al. 2013).

In addition, grain size is important for rice yield potential, and grain size has been an important target in rice breeding (Takeda and Matsuoka 2008). Gene OsSPL16 controls grain size and shape in rice, which can improve grain quantity and yield (Wang et al. 2012). The over-expression of PGL1 (POSITIVE REGULATOR OF GRAIN LENGTH), which is a TF of the bHLH
family, increases grain length and weight due to increased cell length of lemma and palea. Moreover, APG (ANTAGONIST OF PGL) encodes a typical DNA-binding basic helix-loop-helix protein, which acts as an interaction partner of PGL1. Knock down of APG led to increased grain length, which suggested that APG was a negative regulator of grain length and weight, and PGL1 was an inhibitor of APG (Heang and Sassa 2012). Other genes involved in determining grain size are GS3 (Mao et al. 2010), OsPPKL1 (Zhang et al. 2012) and GS6 (Sun et al. 2013).

Another very important trait in rice domestication is the diversification of flowering time (FT). Rice is a typical short day (SD) plant and it needs a suitable photoperiod for flowering in the correct time which is important to get seeds. In rice, OsGI-Hd1-Hd3a is the major pathway controlling flowering time (Yano et al. 2000; Hayama et al. 2002; Kojima et al. 2002). The genes orthologous to CO and FT in rice were identified as flowering-time QTLs and named Hd1 and Hd3a accordingly. Under the SD condition, Hd1 has functions downstream of Hd3a (Kojima et al. 2002; Takahashi and Shimamoto 2011), encoding a mobile flowering signal and promotes floral transition (Tamaki et al. 2007). RFT1, a gene homologous to Hd3a, is also essential for flowering in rice and it promotes flowering in the absence of Hd3a (Komiya et al. 2008). Both RFT1 and Hd3a are floral activators under SD condition and they act upstream of OsMADS14 and OsMADS15 (Komiya et al. 2008). Early heading date 1 (Ehd1), a B-type regulator, promotes rice flowering under SD condition. No ortholog of Ehd1 was found in the Arabidopsis genome (Doi et al. 2004). Ehd2, encoding a putative TF with zinc finger motifs, acts as a flowering promoter by up-regulating Ehd1 under SD condition (Matsubara et al. 2008). OsMADS51 is also a short day flowering promoter which functions upstream of Ehd1, OsMADS14 and Hd3a (Kim et al. 2007). OsMADS50, which also belongs to the MADS TF family, also acts as an important flowering activator upstream of OsMADS1, OsMADS14, OsMADS15, OsMADS18 and Hd3a (Lee et al. 2004). Yeast two hybrid screening and pull down experiments showed that one of the MADS-box genes controlling flowering development in rice, OsMADS1, interacted with OsMADS14 and OsMADS15 together (Lim et al. 2000).

Data from the Arabidopsis genome project indicated that more than 5% of the genes encode TFs (Riechmann et al. 2000), and in rice about 4% of the genes encode TFs (Tyagi et al. 2004). Since the number of non-TF genes is 20-25 larger than the number of TFs, the conclusion is that a single TF is able to regulate the expression of large sets of other genes. The utilization of TFs provides advantages in genetic engineering in that a set of genes can be manipulated at the same time. Plant TFs might act as a hub in the complicated net of gene expression and regulation. Research on TFs can reveal the molecular mechanism of gene expression and regulation, and the
biological processes which are controlled by it. Furthermore, the study of transcription factors might lead us to meet future challenges in crop improvement.

In rice, in total 58 TF families are corresponding to 1,981 and 2,408 putative transcription factors in the indica and japonica subspecies respectively (Jin et al. 2014). One of these, the so-called Dof (DNA-binding with one finger), family is a plant specific group of zinc-finger TFs (Yanagisawa 1995). Although there are several reports about functions of Dof TFs in plants, the functions of most rice Dof genes is still far from clear, and they need to be further investigated and explored for their potential in plant breeding. The characteristics of this gene family and known functions are discussed in the next section of this chapter.

**Dof transcription factor genes**

The first protein identified with a Dof domain was MNB1a in maize which is identical to Dof1 and which can interact with the Cauliflower Mosaic Virus (CaMV) 35S promoter (Yanagisawa and Izui 1993; Yanagisawa 1995). Dof genes seem to be implicated in many different biological processes (Yanagisawa 1995, 1996, 1998). Dof domain proteins are TFs with a highly conserved DNA-binding domain, which presumably includes a single C2-HC zinc finger (Figure 2). They form an important plant-specific transcription factor family (Yanagisawa 1998, 2002, 2004). In Arabidopsis, there are 36 annotated TFs belonging to the Dof family, whereas there are 30 Dof proteins in the rice genome (Lijavetzky et al. 2003; Noguero et al. 2013). Although mainly found in higher plants, Dof proteins have also been discovered in the unicellular alga *Chlamydomonas reinhardtii*, mosses and gymnosperms. This indicates an ancient origin and the possibility of diversification throughout plant evolution. As we know, homologous sequences are paralogous if they were created by genome duplication, and they are orthologous if they are copies of a single gene in different species. Paralogs usually display different functions, while orthologs largely retain the same function (Tatusov et al. 1997). There is very little literature about rice Dof TFs, but a lot of data are available about Dof TFs in other species such as Arabidopsis and maize. The information about the Dof transcription factors from these other species will be helpful to assay the homologous Dof gene in rice.
Absolutely conserved cysteine residues that presumably coordinate zinc and other amino acid residues conserved among all known Dof domain proteins are shown in red and purple, respectively. Well-conserved residues are shown in blue, and redundant residues are shown in gray.

The highly conserved Dof zinc finger domain, normally locates at the N-terminus, and the amino acid sequences outside the Dof domain are very diverse and show little homology (Yanagisawa 2002). The specific DNA sequences to which the Dof proteins bind have been verified by in vitro and in vivo experiments. All Dof proteins analyzed so far recognize an AAAG motif or its revise complementary sequence, CTTT, in the in vitro DNA binding assays (Yanagisawa 2002). The only exception is the pumpkin protein AOBP (Kisu et al. 1998), which takes AGTA repeats as DNA-binding motif. The Dof domain alone is able to confer the sequence specific DNA interaction (Yanagisawa 1995; Kisu et al. 1998), so the regions outside the Dof domain do not play a major role in the interaction with DNA. It has been reported that maize Dof1 can bind an AAAG motif in the nucleosome reconstructed in vitro as well as the motif on naked DNA, although the binding was dependent on the position of the AAAG motif in nucleosome (Cavalar et al. 2003). Moreover, the specific interaction of Dof proteins with the AAAG motif has been also verified in vivo. Dof proteins can bind and activate transcription from a synthetic promoter containing an AAAG motif or certain native promoters embracing an AAAG motif in protoplasts isolated from plants (Yanagisawa and Sheen 1998; Yanagisawa 2000). Furthermore, mutational assays showed the physiological significance of the AAAG sequences in the putative target promoters, such as the promoter of the maize PEPC gene and the maize PPDK gene.
All Dof domains have four conserved cysteine residues, and metal chelators inhibit DNA-binding of maize Dof1 and some other Dof proteins in vitro (Yanagisawa 1995, 2002). Replacement of the conserved cysteine residues with alanine or serine completely abolished DNA-binding of the Dof proteins in vitro (Yanagisawa 1995) and in vivo (Mena et al. 1998). Although there are still no structural biological data, these results strongly indicate that the Dof domain is essential for DNA-binding.

The Dof domain was originally identified as a DNA-binding domain, but now it is regarded as a bifunctional domain for both DNA-binding and protein-protein interactions with other classes of transcription factors (basic domain-leucine zipper proteins, bZIP proteins) and non-histone nuclear proteins called high-mobility group (HMG) proteins. The first protein-protein interaction was found with an Arabidopsis Dof domain protein (OBP1), which was found interacting with bZIP proteins associated with stress responses (Zhang et al. 1995; Chen et al. 1996). Another example of a Dof-bZIP interaction was shown for a maize Dof domain protein (PBF) and a maize bZIP protein, Opaque2 (O2) which regulates endosperm-specific gene expression (Vicente-Carbajosa et al. 1997). The Dof domain also interacts with the conserved DNA-binding domain (the HMG-box) of HMG proteins, abundant non-histone nuclear proteins likely functioning in transcription and DNA replication. The binding of maize Dof1 and Dof2 to DNA was promoted by HMG protein (Yanagisawa 1997). Although protein-protein interactions mediated by the Dof domain were shown with only a few Dof proteins till now, all Dof domains might confer protein-protein interactions because the Dof domain is highly conserved.

As the amino acid sequences outside the Dof domain are divergent, the Dof domain proteins also play diverse roles in plants. These roles include responses to plant hormones such as gibberellin (Washio 2001; Mena et al. 2002) and auxin (DePaolis et al. 1996; Kisu et al. 1998), stress responses (Corrales et al. 2014), tissue specific expression (Yanagisawa 1998; Plesch et al. 2001) and photosynthesis (Yanagisawa and Sheen 1998; Yanagisawa 2000).

Maize Dof1 was associated with expression of multiple genes involved in carbon metabolism (Yanagisawa 2000). It can recognize the AAAG motif in the maize C4PPDK promoter and activates the expression of this gene (Yanagisawa 2000), whereas Dof2 acts as a tissue-specific repressor (Yanagisawa and Sheen 1998). Further in vitro experiments showed that the chromosomal high-mobility group (HMG) protein HMG5 can facilitate nucleosome binding of the transcription factor independent from the position of the recognition sites (Cavalar et al. 2003). Nowadays, Dof genes and functions have been described in many other plant species. Two soybean Dof genes, GmDof4 and GmDof11 can enhance the lipid contents in transgenic
Arabidopsis by up-regulating genes which are associated with the biosynthesis of fatty acids (Wang et al. 2007). SRF1, a Dof gene from sweet potato, modulates the carbohydrate metabolism in the storage roots by negative regulation of *lbbfruct2*, a vacuolar invertase gene (Tanaka et al. 2009). In *Triticum aestivum* (wheat), members of the Dof transcription factor family are associated with light-mediated gene regulation, including involvement in photosynthesis (Shaw et al. 2009), and TaDof1 might be related to the up-regulation of C4 pathway-related genes functioning in heterosis (Chen et al. 2005). In Arabidopsis, Dof genes play important roles in different biological courses, for instance *OBP1* in the cell cycle regulation (Skirycz et al. 2008) and *COG1* and *OBP3* in regulating of phytochrome signaling (Ward et al. 2005), the *CDF1*, *CDF2* and *CDF3* genes in photoperiodic control of flowering (Fornara et al. 2009), the *DAG1* and *DAG2* genes in seed germination (Gualberti et al. 2002).

The Dof TFs in the rice genome (Lijavetzky et al. 2003; Noguero et al. 2013) have been classified into four major clusters, but thus far only OsDof3 and OsDof12 have been well described by functional analysis (Washio 2001; Engelmann et al. 2003; Li et al. 2009). OsDof3 plays an important role in the gibberellin-mediated expression of genes involved in seed germination (Washio 2001; Engelmann et al. 2003), is expressed in maturing endosperm and coordinately expressed with seed storage protein genes. OsDof3 was also found involved in the quantitative regulation of genes which are expressed in the endosperm in cooperation with this *RISBZ1* gene (Yamamoto et al. 2006). OsDof12 is a promoter of the rice heading date by up-regulating the genes *Hd3a* and *OsMADS14* under Long Day (LD) condition (Li et al. 2008; 2009). OsDof9 and OsDof18 were reported to be expressed in rice aleurone layer of grains (Washio 2001). Except for the few reported Dof genes, most of the Dof transcription factor genes have not been investigated in rice till now. Functions of most of the rice Dof transcription factors need to be researched and unraveled. In the thesis we studied two of the rice Dof genes, *OsDof24* and *OsDof25*, which are related to *OsC4PPDK* transcription and regulation, heading date and seeds storage proteins in rice.

**Outline of this thesis**

The aim of this thesis was to make a start with studying the molecular basis for heterosis in Chinese Super hybrid rice by focusing on the transcriptional regulation of *C4PPDK* which is an important gene in photosynthesis. Heterosis, hybrid vigor, or outbreeding enhancement, is the improved or increased function of any biological quality in a hybrid offspring. Chinese Super hybrid rice shows typical heterosis than the parental lines, which offer a good system for investigation of heterosis. Previous studies showed that several genes involved in C4 photosynthesis including *C4PPDK*, were up-
regulated in hybrid rice (Bao et al. 2005) compared to the parental lines. Based on the hypothesis that certain transcription factors may be responsible for these expression polymorphisms, we developed experiments where we finally identified two Dof proteins as regulatory factors of rice C4PPDK.

Chapter 1 presents a general introduction to the factors involved in crop domestication and the specific role of transcription factors in this process. Chapter 2 describes the functional analysis of a Dof-type transcription factor, named OsDof24, from rice. Over-expressing OsDof24 in rice resulted in plants with a reduced plant height and delayed flowering time, but down-regulation using an RNAi approach did not render any obvious phenotype. Detailed transcriptional profiling of heading date-related genes in OsDof24-overexpressing transgenic rice showed that OsDof24 can function in the delay of the heading date by down-regulating transcription of genes Hd3a and RFT1, which are two critical genes controlling this process. Chapter 3 describes the role of OsDof25 in the regulation of OsC4PPDK. Interactions between OsDof25 and OsC4PPDK were confirmed in experiments involving EMSA, yeast one-hybrid analysis and transient expression in rice protoplasts. The regulatory role of OsDof25 in OsC4PPDK expression was confirmed using OsDof25 misexpression plants. The results suggested that OsDof25 is a trans-activator of the OsC4PPDK gene in rice. Chapter 4 describes the effects of OsDof24 and OsDof25 on the content of rice seeds storage proteins. First the interaction between both OsDof24 and OsDof25 with the GluB-1 promoter was investigated by a set of in vitro and in vivo experiments. Results showed that both OsDof24 and OsDof25 can affect seed storage protein content probably by activating the transcription of the GluB-1 promoter. Finally, in Chapter 5, a general discussion of the results from Chapters 2, 3 and 4 is presented.

References


Lee, S., Kim, J., Han, J.J., Han, M.J. and An, G. (2004) Functional analyses of the flowering time gene OsMADS50, the putative SUPPRESSOR OF


Takahashi, Y. and Shimamoto, K. (2011) Heading date 1 (Hd1), an ortholog of Arabidopsis CONSTANS, is a possible target of human selection during


