

DOF transcription factors in developing peanut (*Arachis hypogaea*) seeds

Haiyan Yan¹, Jiaquan Huang², Boshou Liao², Xianqing Lan¹, Qiuting Luo¹, Junlong Tang¹

¹Life Sciences College, South-Central University for Nationalities, Wuhan, China

²Oil Crops Research Institute of Chinese Academy of Agricultural Sciences, Key Laboratory of Oil Crop Biology of the Ministry of Agriculture, Wuhan, China

Email: haiyuan988@yahoo.com

Received 25 October 2011; revised 24 November 2011; accepted 3 December 2011

ABSTRACT

DNA binding with one finger (DOF) transcription factors play important roles in storage material accumulation and morphogenesis of developing seeds. Oil and protein contents varied in different cultivars in important oil crop peanut. DOF proteins have not been studied in this crop. In this paper, we analyzed all the DOF genes expressed in developing seeds from a cDNA library with 20,000 transcripts, cloned and compared similar genes of GW391729 from eight peanut cultivars, and analyzed similar genes expressed in root and leave with control and inoculated with *Ralstonia solanacearum*. The results indicate that total eight types of DOF genes were expressed in developing seeds of cultivar 063103. Most of DOF transcription factors expressed involved in developmental process in a complicated way. Among them, GW391729 is possible related to the seed number in fruit, and also is possible related to leafspot resistance. Detailed function of these DOF proteins need to be further studied.

Keywords: DOF; Leafspot; Peanut; Seed Development; Seed Number

1. INTRODUCTION

DOF family proteins are plant specific with the structure feature of C2-C2 formed Zn-finger binding to DNA motif T/AAAAG at their N-terminal, which possess functions in various biological process in plants [1-3]. The DOF proteins have been systematically analyzed in arabidopsis, rice (*Oryza sativa* subsp. Japonica), poplar (*Populus trichocarpa*) and soybean (*Glycine max*) [1,2,4,5]. DOF proteins have been classified into seven groups [1], and each group contains certain number of conserved sequences [2]. The function of proteins are composed of subfunctions, which are determined by the responsive subdomains in the proteins. The conserved DOF zinc

finger domains in DOF proteins can bind to cis DNA core elements AAAG [6], and may also interact with other proteins including DOF itself [7]. The differences in DOF zinc finger domains determined its binding affinity with specific cis AAAG motif, which could locate in different genes or in different elements of one promoter [8]. The region outside DOF zinc finger domain could have different function, such as modifying structure of the DOF zinc finger or interacting with other proteins, etc [9,10]. Thus it is possible to bioinformatically analyze the function of unknown DOF proteins by paired comparison of similar amino acid sequences of conserved domains.

DOF proteins play different regulation role in various developmental stages in plants. PBF (P box binding factor) is the DOF protein studied in a great detail in the process of seed development and formation of storage material. PBFs expressed in endosperm in maize (*Zea mays*), barley (*Hordeum vulgare*), rice and wheat (*Triticum aestivum*), can bind to cis-elements in endosperm specific promoters to regulate the expression of endosperm specific genes as well as typical storage protein genes [11,12]. The wheat PBFs (WPBF), however, have been reported to be expressed in all types of tissues, and interacts with a TaQM protein from root [13]. PBFs can bind to dof box in the promoter of storage protein gene, and also interact with other transcription factors which binding with the same promoter in the adjacent region such as PBF with O2 [9], BPBF with GAMYB [10] and OBF with OBP [14]. It has also been demonstrated that in some genes there are more than one dof boxes in promoters [15]. For example, ZmDof1 can binds to multiple sites in the promoter of *CYPPDK1* and *PEPCZM2A* genes [15]. In promoters of major storage proteins *ARAH1* and *ARAH3* genes of peanut, there are also multiple dof boxes, which even takes up more than one third of known cis-elements of promoter [16-18].

A genome-wide analysis of genes expressed in arabidopsis revealed as many as 18 DOF genes expressed in

different stages and positions of the developing seeds [19]. Some of them have function in morphogenesis process, such as *At2g28510*, *At5g60200*, *At5g66940*, *At3g55370* and *At2g28810* in vascular differentiation [20-26], and *At5g62430* in flower initiation [27-29]. Some has function as stress response regulation factor, such as *At3g47500* [30,31] and *At5g39660* [32]. Some has function in several types of activities such as *At5g39660* [27,30,32,33], *At3g47500* [30,31,34] and *At2g28510* [20-22,35]. In legume plant *G. max*, 28 *GmDof* cDNA have been cloned, and their expression patterns have been studied with RT-PCR [5]. *GmDof4* and *GmDof11* can increase seed lipid content by network regulation in transgenic arabidopsis plants [5]. For *DOF* genes in developing seed of *A. hypogaea*, little information is available except a few EST sequences from several cDNA libraries. The polymorphism of *DOF* genes expressed in these plants and their possible roles remain to be investigated.

In a cDNA library of developing seeds from peanut variety 063103, eight clones of *DOF* cDNAs from over 20,000 clones have been found. These clones were sequenced, translated and compared with *AtDofs*, *GmDofs* and *ZmPBF*. The *AhDof* proteins are structurally classified and their functions are suggested via bioinformatic and theoretic analysis.

2. MATERIAL AND METHODS

2.1. Materials

Eight *DOF* cDNA clones are from a cDNA library of developing seeds of peanut line 06-3103, which is leaf spot disease resistance. Additional *DOF* cDNA sequences similar to *DF* cDNAs from peanut seed are screened from cDNA libraries of leave and root of control and inoculated with *Ralstonia solanacearum*. The vector is pDNR-LIB (CLONTECH). The plasmids are purified and the inserts are examined by PCR with M13F and M13R primers. The *DOF* clones are sequenced with M13R and T7 primers by BGI (Huada Gene) Company in Wuhan. The *DOF* ESTs from peanut cultivars VBL6, GT-C20 of developing seeds are obtained with blastn-est program from GeneBank [36]. The cDNA sequences are translated into protein sequence with AUGUSTUS (version 2.1) [37]. The *AhDof* ests from 06-3103 are deposited in GeneBank as: *GW391722*, *GW391723*, *GW391724*, *GW391725*, *GW391726*, *GW391727*, *GW391728*, *GW391729*.

Other proteins and Ests from peanut listed in text are from GeneBank and Swiss-Prot. DNA extracted from Material listed in **Table 3**.

2.2. Cloning of *DOF* Genes from Genomic DNA of Nine Peanut Cultivars

DNA was extracted with CTAB method [38] and used as

template in PCR cloning of *DOF* genes.

PCR primers for *GW391729* are: S37: 5'GAAGCA-GAAGGAGGAAGG 3', A1133: 5'ACAACAAAGCA-AAGTAATAAA 3'.

PCR reaction mix was 25 μ L in total, contained ddH₂O 18.3 μ L 10x PCR buffer 2.5 μ L, dNTP (each 10 μ M) 1 μ L, forward and reverse primers (10 μ M) each 1 μ L, DNA template 1 μ L, Taq polymerase (Takara) 0.2 μ L. The PCR reaction for *GW391729* is: 95°C 2 min, then the cycle 95°C 40 s, 49.5 (each cycle down 0.4)°C 30 s, 72°C 40 s, for 20 cycles, next cycle, 95°C 40 s, 41.5°C 30 s, 72°C 40 s for 18 cycles, then 72°C for 7 mins. The specific products are purified and constructed into pMD18T vectors, the positive clones are sequenced by BGI (Huada) genomic company and Ebioe (Shenggong) company.

2.3. Sequence Comparison and Analysis

DOF cDNAs from peanut line 06-3103 were batch blasted with blastx-nr and blast-swiss to get mostly similar *DOF* proteins from arabidopsis and soybean, maize PBF was used as a reference [39]. Then the selected seed *DOF* proteins were pairwise compared with GeneDoc software. The conserved domains are searched with conserved motifs listed by Lijavetzky *et al.* [2]. The *DOF* proteins are classified by the similarity comparison with that of arabidopsis according to Yanagisawa [1].

2.4. Bioinformatic Analysis on Function of *DOF* Proteins

The information of expression patterns of *AtDofs* similar to *Ahdofs* from developing seeds of 06-3101 were collected from iHOP (Information Hyperlinked over Proteins), <http://www.ihop-net.org/> [40,41]. The information about functions of related *DOF* proteins are also picked from literature by searching key word "DOF" and genebank association numbers in NCBI Pubmed, and by searching TAIR and linked information.

3. RESULTS

3.1. Eight Different *DOF* Genes Expressed in Developing Peanut Seed

Eight different *DOF* genes were found from a cDNA library from developing seed with over 20,000 clones of cultivar 063103. The resulted *DOF* proteins belong to *DOF* type II, III, IV, V, and VII respectively according to Lijavetzky *et al.* [2] (**Table 1**). *GW391725*, *GW391726* and *GW391727* are similar to three CYCLING *DOF* FACTORS (*CDF*) genes (type II), *AtCDF3*, *AtCDF2* and *AtCDF1* in arabidopsis respectively (**Figure 1(a)**). *GW391722* belong to type III, similar to *GmDof11* and *AtDof2.1* (**Figure 1(b)**). *GW391723* is similar to *AtDof2.2*

Table 1. Comparison of the conserved domains between AhDofs, GmDof and AtDofs besides DOF domains.

Group ¹	GeneBank Association Number	Location/Domain number ¹	Conserved sequences
II CDFs	GW391725 GW391726 GW391727 GO332017 At5g39660 At3g47500 At5g62430 DQ857278	3N	KDPAIK(T)LFGKTIPV(L/F)P
		4N	K(G)T(A/K)L(V)KKPD(T)K(V)ILP
		14C	NXXTV(L)LT(S/A)FGS(L)D(E)XXLCESMA(T)SV(G)L NLXEK(D)
		2C	E(E)EAAKSSIWTLGIK
		23C	Y(C)F(I)PGP(V/A)P(S)XWT(P)F(Y)P(T)W
		18C	PFYXPYWG
		5C	LQANPAALSRS(XX)Q(N)(S/C)FHE(GS)
		9C	SPTLGKHS(P)RD(E)E
		30C	C(S)L(V)W(L)VPKTLRIDDPN(E)EAAKSSIWT(E)TLG IK
		12N	ERKP(A)RPQ(H)P(K/V)E(D)QALK
III	GW391722 DQ857261 DQ857263	AhGm N	MDPSSGQxQ(E)MSSXQS(G)L(V)E
		13C	S(I)DLTLAL(V)AXXRLQKQS
		11C	K(R)VL----GFPW—N-D-N-G-D-N-GE-D-S-R
		20C	G--D—DS—R--W
IV	GW391724 At5g66940 DQ857251	AhGmAtN	MPS—S-ESRR—K-P-----AP—QE-L
		33C cp AhAtC	GPGG-F-----FGLGLG-G---E-V—G----W-F-----A- -----TWQFEG-----GFVSGD
V	GW391723 At2g28810	17N	MVF-S—P-YLDPP
VII	GW391728 DQ889513 DQ857255	N	H(K)---V(L/I)QQQQ-----E(A)Q(L)P(K)R)
		C	L(V/A)K(S)R--SS-VL(I)--SSS--D(H)---SS--H--

¹amino acids inside the brackets are alternatives in different DOFs 1 [1].

(**Figure 1(c)**), belong to type V. GW391724 is similar to AtDof3.4 and GmDof2 (**Figure 1(d)**), belong to type IV. GW391728 is similar to AhDof3 (ACF74278.1) and GmDof5 (**Figure 1(e)**), belong to type VII; GW391729 is similar to but at less similar level to AtDof2.2 and GmDof12 (**Figure 1(f)**). Among these similar AtDof and GmDof proteins, All DOFs are expressed in developing seed [19,40].

Different types of DOF proteins not only have their distinct sequences in N and C terminal regions (**Table 1**), but also have their specific characteristics in conserved DOF finger regions (**Figure 2**). The key structure in zinc finger domain is two CXXC bridges and a C residue surrounded by some conserved amino acids between two CXXCs. Four C residues in two CXXCs binding with one zinc ion cause the formation of a finger like conformation in this region. There are four variable motifs in zinc finger domain among AhDofs, GmDof and AtDofs. The first motifs is N (or K/D/E) SM (orT/P/S) D (E/N) which locates right after CPRC (**Figure 2**). The second is NVN (NIN/NFS/NYS/SLS/SLT), which is after KFCY-YNNY, the middle conserved domain containing middle C residue (**Figure 2**). The third variable region cover the second CXXC, H (or/Y) FACKN (or/A/S/T/K) CQ

(or/R/K) RYWTS (or/A/H/K/R) (**Figure 2**). The fourth variable region is AGR (/GGS/GVS/GGC/GSY) (**Figure 2**), right before the possible NLS sequence RKN—KR [42,43]. The N terminal CXXC is conserved as CPRC in AhDofs listed, the second CXXC in AhDofs listed have variation as CKNC (CDFs), CKSC (GW391724, GW391722), CKAC (GW391729, GW391728, and PBF of Maize) and CKTC (GW391723) (**Figure 2**).

It is obvious that in all CDFs there is coherence among four variable regions and consistency among different CDFs in the same region (**Figure 2**). It is coherent that the four variable regions are all relatively conserved in CDFs. They are N (or K) SMD (for peanut) (E for arabidopsis), NV (or I) N, HFCKN (for peanut and soybean) (A or K for arabidopsis) Q—A(or S), AGR. The CDFs are consistent in that the conserved combination of amino acids are rich of N or Q (only one) which are polar amino acid residues without charge. In the key variable region, the variable amino acids have the same physiochemical property, such as D and E in the first region, V and I in the second region (**Figure 2**). Some are restrictly the same, such as M before D or E, N-N, Q after the second CXXC, AGR before RKN. The common property in CDF zinc DOF domain right before nuclear

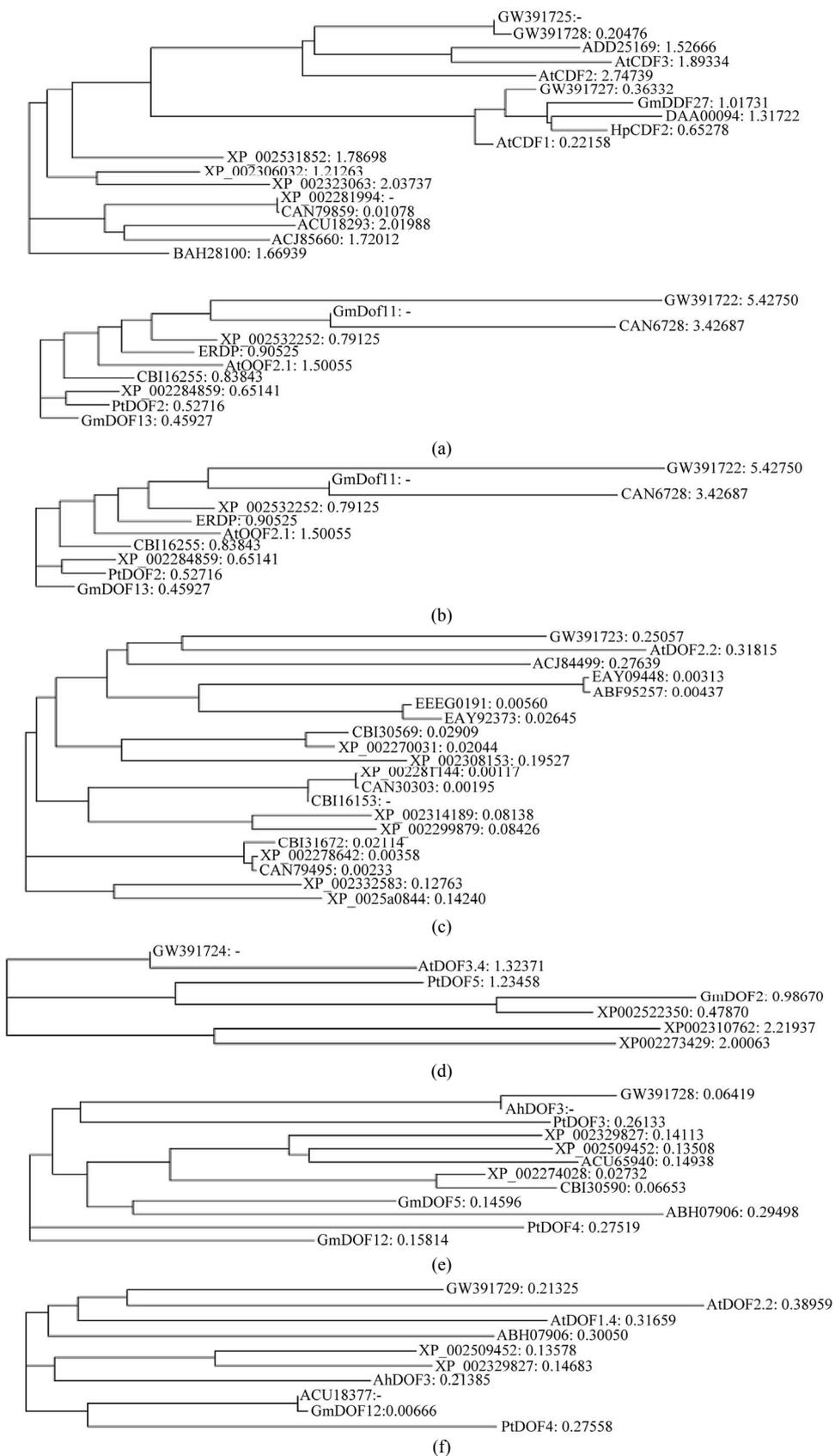


Figure 1. Phylogenetic trees of AhDofs with their similar DOF proteins.

	I	II	III	IV				
GW391726	: --QDKTLKKPKDKILP	CPRCNSM	DAKFCYNNY	NVNVQPRHFC	CKNCQRYWTAGG	TRNVVPGAGRR	RKNKN----	
GW391725	: --QDKTLKKPKDKILP	CPRCNSM	DTKFCYNNY	NVNVQPRHFC	CKNCQRYWTAGG	TRNVVPGAGRR	RKNKN----	
GO332017	:							
GW391727	: --QEKAVKKPKDKVLP	CPRCNSM	DTKFCYNNY	N-----				
AT5G39660	: --QEGKLLKPKDKILP	CPRCNSM	ETKFCYNNY	NVNVQPRHFC	CKCQRYWTAGG	TRNVVPGAGRR	RKNKS----	
GMDOF27_L	:	PCPRCNSM	DTKFCYNNY	NVNVQPRHFC	CKNCQRYWTAGG	TRNVVPGAGRR	RKNKN----	
AT3G47500	: ----KTLKKPTKILP	CPRCKSM	ETKFCYNNY	NINQPRHFC	CKACQRYWTAGG	TRNVVPGAGRR	RKNKS----	
AT5G62430	: ----DKTLKKPTKILP	CPRCNSM	ETKFCYNNY	NVNVQPRHFC	CKACQRYWTAGG	TRNVVPGAGRR	RKNKN----	
EE125478	: ----SSS--EP--MK	CPRCES	STNTKFCYNNY	NKSPRYLC	RTHWTKVD--	LRNVVPGG--	RKNKR----	
GW391724	: -----AEQEHL	CPRCDS	STNTKFCYNNY	NFSQPRHFC	CKSCRRYWTHGG	TLRDI	PVGGGSR	RKNAKR----
AT5G66940	: -----DQEQQLS	CPRCES	STNTKFCYNNY	NFSQPRHFC	CKSCRRYWTHGG	TLRDI	PVGGGSR	RKSSKR----
GMDOF2	: -----PEQENL	CPRCDS	STNTKFCYNNY	NYSQPRHFC	CKSCRRYWTHGG	TLRDI	PVGGGSR	RQNAKR----
GW391722	: --RKPRPQPE-QALK	CPRCDS	STNTKFCYNNY	SLSQPRYF	CKSCRRYWTKGG	TLRNVVPGGG	CRKN--KR----	
GO339137	:							
GMDOF13	: --RKPTPQPE-QALK	CPRCDS	STNTKFCYNNY	SLSQPRYF	CKSCRRYWTKGG	TLRNVVPGGG	CRKN--KR----	
GMDOF11	: --KKPKPHPE-QALK	CPRCDS	STNTKFCYNNY	SLSQPRYF	CKSCRRYWTKGG	TLRNVVPGGG	CRK--KR----	
AT2G28510	: KKKVSRGELGGEAQN	CPRCES	STNTKFCYNNY	SLSQPRYF	CKSCRRYWTKGG	TLRNVVPGGG	CRRN--KR----	
AT5G60200	: TRPQP---PEL-ALR	CPRCDS	STNTKFCYNNY	SLTQPRYF	CKSCRRYWTKGG	TLRNVVPGGG	CRKN--KR----	
GO324523	: -----LRCPRCDS	STNTKFCYNNY	NKSPRHF	CKACRHWTKGG	TLRNVVPGGG		RKNKR----	
GW391728	: EHVQQQQQQQAL	CPRCDS	STNTKFCYNNY	SLSQPRHFC	CKACKRYWTRGG	TLRNVVPGGG	CRKNTKRLK	RQ
EE124644	: EHVQQQQQQQ--AL	CPRCDS	STNTKFCYNNY	SLSQPRHFC	CKACKRYWTRGG	TLRNVVPGGG	CRKNTKRLK	RQ
O24463	: GERKPRPQLP-EAL	CPRCDS	STNTKFCYNNY	SMSQPRYF	CKACRRYWTHGG	TLRNVVPGGG	CRKNKHASR	EV
GMDOF5_L	: DQLVQQQQQQEAP	CPRCDS	STNTKFCYNNY	SLSQPRHFC	CKACKRYWTRGG	TLRNVVPGGG	CRRN--KRVK	RP
AT1G28310	: QQQQQQPQPQPQL	CPRCDS	STNTKFCYNNY	SLSQPRHFC	CKACKRYWTRGG	TLRNVVPGGG	SYRKN--KRVK	RP
GW391723	: -----PRSGSEGAL	CPRCES	STNTKFCYNNY	SLSQPRHFC	CKRRYWTRGGAL	LRNVVPGGG	FRRNKK--TKRS	
AT2G28810	: -----ALKCPRCDS	ANTKFCYNNY	NLTQPRHFC	CKACRRYWTRGGAL	LRNVVPGGG		CRRNKK--GKSG	
GO332293	:							
GW391729	: S*SRMEKQKEEGV	KCPRCDS	STNTKFCYNNY	SLSQPRHFC	CKACKRYWTRGG	TLRNVVPGGG	SRKNKRR----	
GMDOF12_L	: DLLQQQQQ---AL	KCPRCES	STNTKFCYNNY	SLSQPRHFC	CKACKRYWTRGG	TLRNVVPGGG	CRKN--KRVK	RP

Figure 2. Comparison of DOF domain sequences between similar AhDofs, AtDofs and GmDofs.

localization sequences possibly provided the structure basis for their common specific DNA binding property related to the function of CDF types of DOF proteins.

The second similarity group of DOFs contains GW391724, At5g66940 and GmDof2. They have variable I region as D(E)ST, variable II region as NFS, the third variable region as SCR—H, the fourth variable region as G(V)SR (**Figure 2**).

The specific conserved domain in GW391722 similarity group is in region III, YFCKSCRRYWTK (**Figure 2**).

GW391722,23,28,29 all contained a same N-glycosylation site overlay the DOF middle conserved domain and variable II region (NYSL). All three CDF and GW391724 do not have N-glycosylation site in conserved DOF domain, but three overlaid N-glycosylation site on the N-terminal outside of conserved DOF domain.

From ESTs deposited in GeneBank database derived from various cDNA libraries of developing seed of peanut, partial sequences of four *DOF* cDNAs (*GO332017*, *GO339137*, *GO324523*, *GO332293*) from VBL6 cDNA library, two *DOF* cDNAs (*EE124644*, *EE125478*) from Luhua4 cDNA library, and one (*ES703897*) from Tifton C20R5 cDNA library were obtained, translated and compared with DOF proteins from 06-3103. *GO332017* is a partial sequence with 215 amino acids, highly similar to

C terminal domain of AhCDF3s (GW391726, GW391725) (**Figure 3(c)**). Within its N terminal 205 amino acids residues, only 13 amino acids are different from that of GW391726, within its N terminal 86 amino acids residues, only 9 amino acids are different from that of GW391725 (**Figure 3(c)**), which is different between species. *GO339137* is highly similar to a 67 aa at C terminal of GW391722, with only 6 aa residues difference between them (**Figure 3(a)**). These differences are possibly the result from differences between species. *GO324523*, *GO332293* have little similarities to other domains outside of DOF zinc finger domain with known AhDofs.

3.2. GW391729 Was Expressed in Peanut Leave Inoculated with *Ralstonia solanacearum*

DOF genes expressed from cDNA libraries of roots and leave with control and inoculated with *Ralstonia solanacearum* from same cultivar 063103 were examined. *GW391729* was found expressed only in leave inoculated with *Ralstonia solanacearum*, not in controlled leave (**Figure 4**), this indicates that *GW391729* is possibly related to the leaf resistant to *Ralstonia solanacearum*. In leave of both controlled and inoculated with *Ralstonia solanacearum*, similar but different ests to *GW391724*

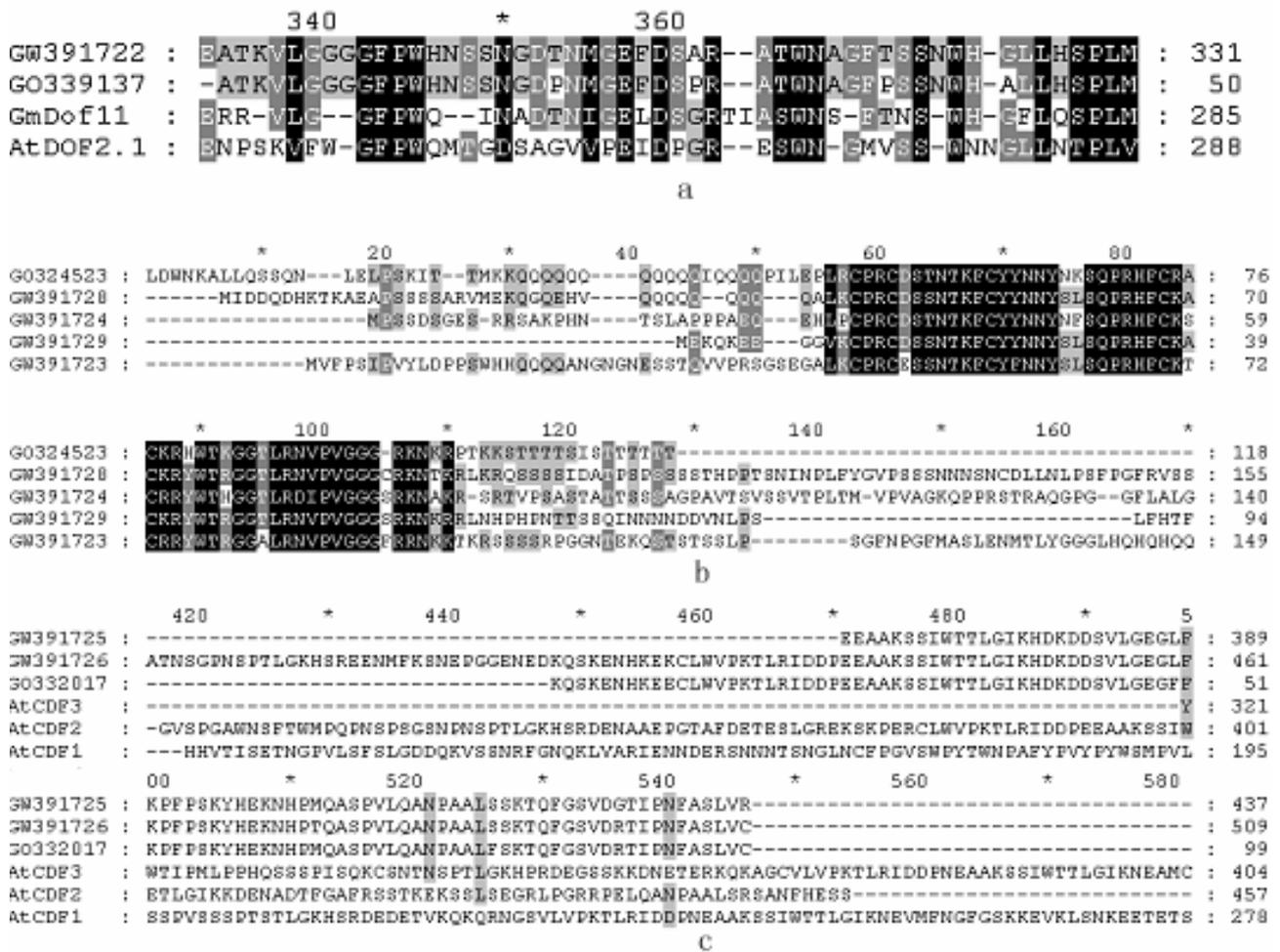


Figure 3. Alignment of AhDOFs and AtDOFs.

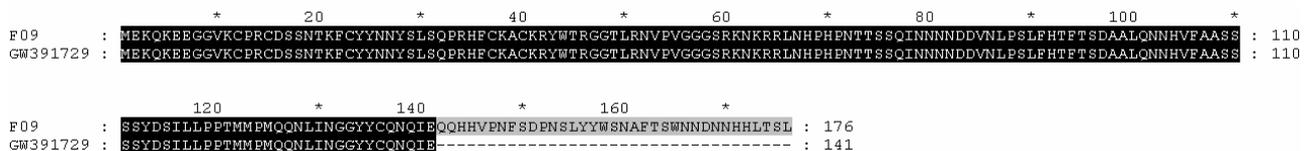


Figure 4. Comparison of GW391729 and cDNA F09 from leaf inoculated with *Ralstonia solanacearum*.

were expressed (**Figure 5**), these ests are not the same DOF proteins as GW391724, possibly related to the leaf development. In root inoculated with *Ralstonia solanacearum*, an est similar to but different from GW391722 was expressed (**Figure 6**), its function is possibly related to root response to *Ralstonia solanacearum*. As for function of GW391724 and GW391722 expressed in seed and different from those in leaf and root respectively, need further investigation.

3.3. Evolution Origin of AhDofs from Developing Seeds

The blastn results using the est sequences of AhDofs from

developing seeds of 063103 with Genebank tools show that the most similar *AhDofs* are from wild peanut species *A. ipaensis* (**Table 2**), which is the source of BB genome of cultured peanut specie [44]. This indicates that all the eight *AhDofs* are possible from peanut BB genome.

Among *AhDofs*, ESTs from wild *Arachis* species similar to GW391722, GW391725, GW391726, GW391727 are only from *A. ipaensis* (**Table 2**), indicates these four genes are possible locate in only BB genome. For GW391723, GW391728, GW391729, beside that the most similar ests are from *A. ipaensis*, there are two similar ests to each from seed of *A. duranensis* (**Table 2**), which is the origin of AA genome of cultured peanut

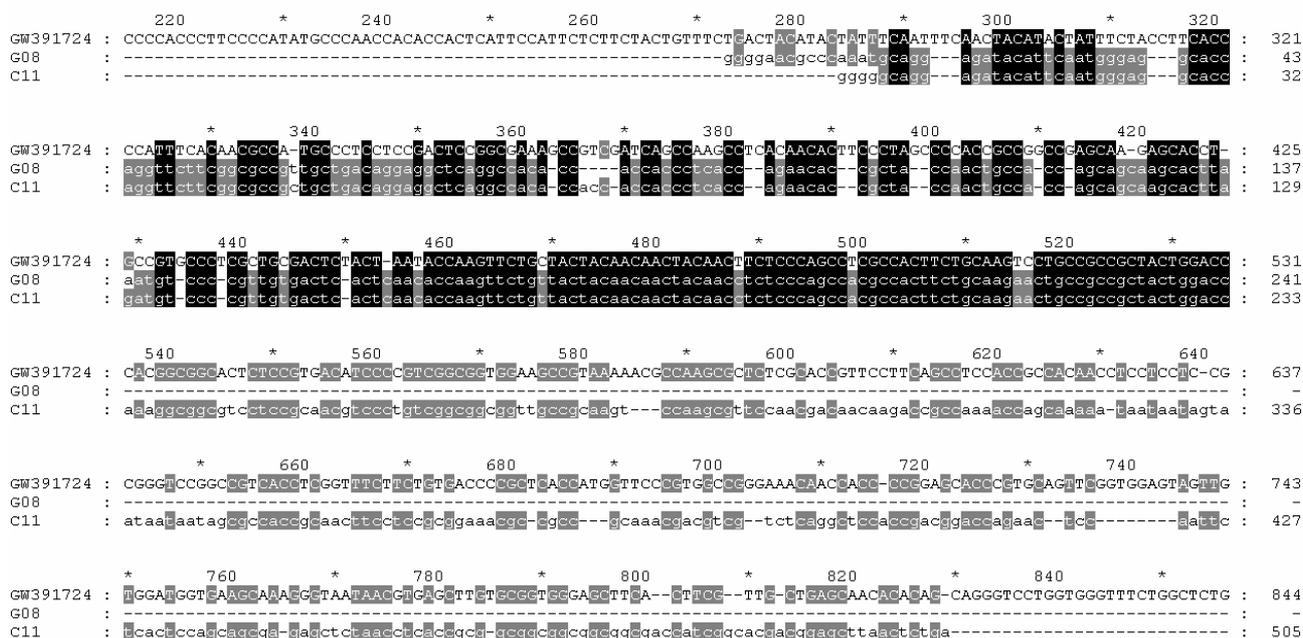


Figure 5. Comparison of *GW391724* and similar cDNAs from leaf inoculated with *Ralstonia solanacearum*.

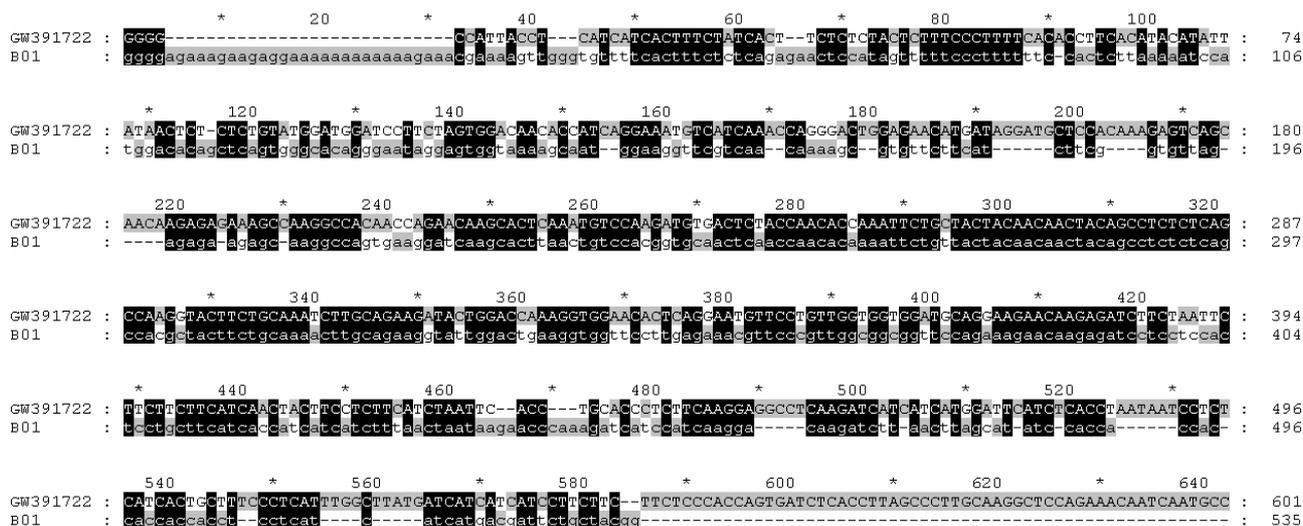


Figure 6. Comparison of *GW391722* and similar cDNAs from root inoculated with *Ralstonia solanacearum*.

specie [44]. This reveals that in AA genome there are possible similar alle genes.

3.4. GW391729 Are Different between Strains with Different Seed Numbers in a Fruit

Sequences of GW391729 from 8 cultivars were compared and analysed. Most of cultivars with 3 seeds per fruits have distinctive sequences, which is different from those with 2 seeds per seeds (Table 3). But it seems that size of seeds and resistance to intrusion of seed do not related to GW- 391729.

4. DISCUSSION

The integrated function of a protein is determined by its integration of structure and sequence. The different domain of a protein may have different activities. DOF proteins have relatively conserved zinc finger domain to bind to DNA promoter sequence, and variable C terminals differing at amino acid composition and lengths. The different structure in zinc finger domain determined their specific binding targets [7]. Specific C terminal sequences and structure determined their specific reaction other than binding to DNA. Thus the specific function of

Table 2. Similarity analysis of DOF est to those from original species in *Arachis* genus.

GeneBank Acc. No.	Species	Similarity Score	Similar DOF	Genome
GW391722	<i>A. hypogaea</i>			
GW974537	<i>A. ipaensis</i>	638	<i>GmDof13</i>	BB
GO339137	VBL6	335	<i>Dof11</i>	
GW391723	<i>A. hypogaea</i>			
GW973881	root <i>A. ipaensis</i>	987	<i>OBP3</i>	BB
GW947099	Seed <i>A. duranensis</i>	869	<i>OBP3</i>	AA
GW946556	Seed <i>A. duranensis</i>	339	<i>At2g37590</i>	AA
GW391725	<i>A. hypogaea</i>			
GO332017	VBL6	1079	<i>AT5g39660 CDF2</i>	
GW983610	<i>A. ipaensis</i>	843	<i>Dof27</i>	BB
GW965234	<i>A. ipaensis</i>	822	<i>AT5g39660 CDF2</i>	BB
GO259985	VBL1	459		
GW391726	<i>A. hypogaea</i>			
GO332017	VBL6	1094	<i>AT5g39660 CDF2</i>	
GW965234	<i>A. ipaensis</i>	843		BB
GW983610	<i>A. ipaensis</i>	830	<i>Dof27</i>	BB
GO259985	VBL1	468		
GW978277	<i>A. ipaensis</i>	287	<i>CDF3</i>	BB
GW391727	<i>A. hypogaea</i>			
GW978277	<i>A. ipaensis</i>	287	<i>CDF3</i>	BB
GW391728	<i>A. hypogaea</i>			
EE124644	Luhua14	1051	<i>Dof5</i>	
GW989805	<i>A. ipaensis</i>	1014		BB
GW950436	Seed <i>A. duranensis</i>	422		AA
GW948904	Seed <i>A. duranensis</i>	363		AA
GW935134	root <i>A. ipaensis</i>	283	<i>Dof8</i>	BB
GW977281	<i>A. ipaensis</i>	276	<i>Dof12</i>	BB
GW391729	<i>A. hypogaea</i>			
GW982725	<i>A. ipaensis</i>	1155	<i>Dof12</i>	BB
GW980043	<i>A. ipaensis</i>	665	<i>Dof12</i>	BB
GW967091	root <i>A. ipaensis</i>	422	<i>Dof12</i>	BB
GW953727	Seed <i>A. duranensis</i>	318	<i>Dof5</i>	AA
GW936909	root <i>A. duranensis</i>	316	<i>Dof5</i>	AA

Table 3. Differential sites of GW391729 among from 8 cultivars.

Property	Cultivar			Sites in protein								
	58	79	82	96	99	104	120 ~ 122	146	154	174	176 ~ 178	
1	3	C	T	H	H	P	A	AST	P	P	P	~~~
2	2s	C	T	H	H	L	A	AST	P	P	P	~~~
3	2RI	C	T	H	H	L	A	AST	P	P	P	~~~
4	2L	S	N	D	D	L	V	AST	P	A	P	~~~
5	3s	S	N	D	D	L	V	~~~	H	P	H	SLI
6	3	S	N	D	D	L	V	~~~	H	P	H	SLI
7	3RI	S	N	D	D	L	V	~~~	H	P	H	SLI
8	2RI	S	N	D	D	L	V	~~~	H	P	H	SLI

(notes: 2, 3-4-seeds per fruit; s-small; RI-resistant to intrusion of *Aspergillus flavus*, L-large). 1-Indored; 2-Danfeng; 3-laofandou; 4-xh3; 5-Meiying8; 6-NcA927; 7-ICG12370; 8-ICG4750.

DOF proteins can be deduced by their sequence similarities to each structure domain with known function.

GW391725, GW391726 and GW391727 are three highly similar DOFs within AhDofs. They are mostly similar to AtCDF1 (At5g62430), AtCDF2 (At5g39660) and AtCDF3 (At3g47500). They contain all the conserved motifs of group II DOF proteins [1] (**Table 1**). CDF genes were first identified in Arabidopsis with yeast two-hybrid screen for proteins interacting with C terminal kelch repeats of FKF1, as well as that of LKP2, which can degrade CDFs, but not that of ZEITLUPE (ZTL) which functions in different signal transduction pathways [45,46]. Three AtCDFs were isolated as AtCDF1 (At5g62430), AtCDF2 (At5g39660), AtCDF3 (At3g47500) [45,46]. AtCDF1 was further found to bind to the promoter of *CO* gene as a key negative factor in flowering pathway. By degrading the CDF1, FKF1 can release the inhibition of CDF1 on *CO* expression and thus release the inhibition of flowering [45,46]. Thus AtCDF1 is an inhibitor for flowering. The highest expression of *AtCDF1* is in cauline senescent leave, stem between 1st and 2nd internode, the intermediate expression level is detected in vegetative rosette, early embryo stages 3 to 5, and the pods in the same stages, each flower organs of flower stages 12 to 15, as well in mature pollen. AtCDF1 was also reported restricted in vasculature [29]. AtCDF2 and AtCDF3 do not have effect on flowering [45], while another study indicates AtCDF2 also function in repression of photoperiod induced flowering [47]. *AtCDF2* and *AtCDF3* expressed more specifically in developing seeds, the expression level increased from torpedo stage till the highest level at curling cotyledon stage 9 [40,41]. The expression of *AtCDF2* is more specific in developing seed and dry seeds than *AtCDF3*. Besides in developing seeds, *AtCDF2* is also expressed in mature pollen and cauline senescent leave. *AtCDF3* expressed at higher level in cauline senescent leave, stem between 1st and 2nd internode than *AtCDF2*, it also expressed in vegetative leave, stage 15 flower/pedicel, stage 15 petal and stamen, lower expression level in sepal, early stage of surface of pods [40,41]. All the three *AtCDFs* (*At3g47500*, *At5g39660* and *At5g62430*) are expressed in sperm cells and they are the only *DOF* genes expressed there of the 32 *DOF* transcription factors [27]. In cold acclimations, *AtCDF2* and *AtCDF3* are two of five *DOF* genes upregulated [29,30]. AtCDF3 has been known as H-protein promoter binding factor-2a. Since three AtCDFs all interact with FKF1 and LPK, and they all have long N terminal domains before DOF zinc domains which located just before NLS and binding to DNA sequences on promoters of genes, thus may cause large bulk at binding sites and possibly hinder the interaction of other transcription factors with the promoters, so that the function of CDFs may be con-

served in higher plants, and they act as inhibitors for expression of certain genes.

GW391722 has a similarity of 94% and 91% to At5g60200 and At2g28510 within 56 aa residues in DOF conserved domain respectively, 85% to GmDof13 in DOF domain with 97 aa residues, 54% to GmDof11 in DOF zinc finger containing domain with 145 aa residues, it also have DOF group III conserved motif 13, 11, 20 [1], belonging to group III. Both group III DOF At5g60200 and At2g28510 were reported to regulate vascular formation process [19]. *At2g28510* was upregulated by cytokinin [21] and GA4 [20] stimulation. Compared to *At2g28510*, more expression of *At5g60200* was detected in shoot meristem [39,40]. The expression of *At2g28510* enriched in 24 h imbibed seeds. In *G. max*, under control of 35S CaMV promoter, *GmDof11* increase the lipid content in seeds of transgenic Arabidopsis plants by binding to a P-box like cis DNA element and also regulate the expression of a set of genes [4]. This indicates that GW391722 may have possible function in both vascular differentiation and metabolism by binding to promoters of a set of genes.

GW391724 shared sequence similarity of 81% to GmDof2 zinc finger domain with 64 aa residues, 73% to AtDof3.4 (At3g50410, OBP1) zinc finger domain with 63 aa residues, 72% to AtDof5.8 (At5g66940) zinc finger domain with 66 aa residues. It has a DOF group IV conserved motif 33 and a undetermined motif among GW391724, AtDof5.8, and GmDof2 (**Table 1**), so possibly belong to group IV [1]. Group IV DOF At5g66940 is highly expressed in shoot apex inflorescence, vegetative, and transition stage, and in a patchy pattern in young floral organ primordia, latter stages limited to stamens and carpels, and early stages of developing seeds [39,40]. *At5g66940* was down regulated in the strubbelig-like mutant (slm) in which plant organogenesis controlled by SLM genes, was one of genes responsive to SLM genes [23]. AtOBP1 was recently reported upregulating cell cycle specific genes, shortening G1 phase and overall length of the cell cycle, function as specific cell division regulator [38].

Among AhDofs, GW391728 shared the highest similarity to PBF at N terminal sequences including zinc finger domain (65%, 55 out of 84 aa), which belong to group VII [1], thus GW391728 is also possibly a cis DNA element P box-like binding factor. Besides, GW391728 has the highest similarity to ACF74278/ DQ-889513 from peanut line Luhua14 with sequence identity of 169/181 aa (93%), but is much less similar to protein sequence of GO324523 from peanut line Tifrunner (**Figure 2(b)**), and did not found similar ests in peanut GT-C20R or other cultivars in genebank database. Comparison of protein sequences from eight cultivars reveals that GW391729 is possible related to the seed number in

fruits. But it is only display such phenomenon among 8 evolutionally closed cultivars. GW391729 is also has possible function on leaf disease resistance, thus its function may related to development and possible cooperate with other DOF proteins.

The protein sequences of GW391723 and GW391729 shared similarities to that of At2G28810 with 48% at 76 aa/157aa and 59% at 61aa/103 aa respectively. Amino sequence of GW391723 has a conserved group V motif 17 at N-terminal domain (**Table 1**) and possibly belong to group V [1], but which of GW391729 does not have similar sequence in other domains besides DOF zinc finger domain with known DOF proteins. Group V *At2G28810* expressed in all stages but not in late pod stages [39,40]. It is preferred expressed in phloem companion cells. Thus, it has a possible role in differentiation of vascular tissue [24].

5. ACKNOWLEDGEMENTS

This paper was supported by Natural Science Foundation of Hubei province (Bzy05011) and 863 project (2006AA10A114).

REFERENCES

- [1] Yanagisawa, S. (2002) The Dof family of plant transcription factors. *Trends in Plant Science*, **7**, 555-560. [doi:10.1016/S1360-1385\(02\)02362-2](https://doi.org/10.1016/S1360-1385(02)02362-2)
- [2] Lijavetzky, D., Carbonero, P. and Vicente-Carbajosa, J. (2003) Genome-wide comparative phylogenetic analysis of the rice and Arabidopsis Dof gene families. *BMC Evolutionary Biology*, **3**, 17-27. [doi:10.1186/1471-2148-3-17](https://doi.org/10.1186/1471-2148-3-17)
- [3] Umemura, Y., Ishiduka, T., Yamamoto, R. and Esaka, M. (2004) The Dof domain, a zinc finger DNA-binding domain conserved only in higher plants, truly functions as a Cys2/Cys2 n finger domain. *Plant Journal*, **37**, 741-749. [doi:10.1111/j.1365-313X.2003.01997.x](https://doi.org/10.1111/j.1365-313X.2003.01997.x)
- [4] Yang, X., Tuskan, G.A. and Cheng, Z. (2006) Divergence of the Dof gene families in poplar, arabidopsis, and rice suggests multiple modes of gene evolution after duplication. *Plant Physiology*, **142**, 820-830. [doi:10.1104/pp.106.083642](https://doi.org/10.1104/pp.106.083642)
- [5] Wang, H., Zhang, B., Hao, Y., Huang, J., Tian, A., Liao, Y., Zhang, J. and Chen, S. (2007) The soybean Dof-type transcription factor genes, GmDof4 and GmDof11, enhance lipid content in the seeds of transgenic Arabidopsis plants. *Plant Journal*, **52**, 716-729. [doi:10.1111/j.1365-313X.2007.03268.x](https://doi.org/10.1111/j.1365-313X.2007.03268.x)
- [6] Yanagisawa, S. (1995) A novel DNA-binding domain that may form a single zinc finger motif. *Nucleic Acids Research*, **23**, 3403-3410. [doi:10.1093/nar/23.17.3403](https://doi.org/10.1093/nar/23.17.3403)
- [7] Yanagisawa S. (1997) Dof DNA-binding domains of plant transcription factors contribute to multiple protein-protein interactions. *European Journal of Biochemistry*, **250**, 403-410. [doi:10.1111/j.1432-1033.1997.0403a.x](https://doi.org/10.1111/j.1432-1033.1997.0403a.x)
- [8] Yanagisawa, S. and Schmidt, R.J. (1999) Diversity and similarity among recognition sequences of Dof transcription factors. *Plant Journal*, **17**, 209-214. [doi:10.1046/j.1365-313X.1999.00363.x](https://doi.org/10.1046/j.1365-313X.1999.00363.x)
- [9] Vicente-Carbajosa, J., Moose, S.P., Parsons, R.L., Parsons R.L. and Schmidt R.J. (1997) A maize zinc-finger protein binds the prolamins box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator Opaque2. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 7685-7690. [doi:10.1073/pnas.94.14.7685](https://doi.org/10.1073/pnas.94.14.7685)
- [10] Diaz, I., Vicente-Carbajosa, J., Abraham, Z., Martinez, M., Isabel-La, Moneda, I. and Carbonero, P. (2002) The GAMYB protein from barley interacts with the DOF transcription factor BPBF and activate endosperm-specific genes during seed germination. *Plant Journal*, **29**, 453-464. [doi:10.1046/j.0960-7412.2001.01230.x](https://doi.org/10.1046/j.0960-7412.2001.01230.x)
- [11] Albani, D., Hammond-Kosack, M.C.U., Smith, C., Conlan, S., Colot, V., Holdsworth, M. and Bevan, M.W. (1997) The wheat transcription activator SPA. A seed-specific bZIP protein that recognizes the GCN4-like motif in the bifactorial endosperm box of prolamins genes. *Plant Cell*, **9**, 171-184.
- [12] Vicente-Carbajosa, J., Onate, L., Lara, P., Diaz, I. and Carbonero, P. (1998) Barley BLZ1: A bZIP transcriptional activator that interacts with endosperm-specific gene promoters. *Plant Journal*, **13**, 629-640. [doi:10.1111/j.1365-313X.1998.00068.x](https://doi.org/10.1111/j.1365-313X.1998.00068.x)
- [13] Dong, G.Q., Ni, Z.F., Yao, Y.Y., Nie, X.L. and Sun, Q.X. (2007) Wheat Dof transcription factor WPBF interacts with TaQM and activates transcription of an alpha-gliadin gene during wheat seed development. *Plant Molecular Biology*, **63**, 73-84. [doi:10.1007/s11103-006-9073-3](https://doi.org/10.1007/s11103-006-9073-3)
- [14] Zhang, B., Chen, W., Foley, R.C., Buttner, M. and Singh, K.B. (1995) Interaction between distinct types of DNA binding proteins enhance binding to ocs element promoter sequences. *Plant Cell*, **7**, 2241-2252.
- [15] Yanagisawa, S. (2000) Dof1 and Dof2 transcription factors are associated with expression of multiple genes involved in carbon metabolism in maize. *Plant Journal*, **21**, 281-288. [doi:10.1046/j.1365-313x.2000.00685.x](https://doi.org/10.1046/j.1365-313x.2000.00685.x)
- [16] Viquez, O.M., Konan, K.N. and Dodo, H.W. (2003) Structure and organization of the genomic clone of a major peanut allergen gene, *Ara h 1*. *Molecular Immunology*, **40**, 565-571. [doi:10.1016/j.molimm.2003.09.002](https://doi.org/10.1016/j.molimm.2003.09.002)
- [17] Viquez, O.M., Konan, K.N. and Dodo, H.W. (2004) Genomic organization of peanut allergen gene, *Ara h 3*. *Molecular Immunology*, **41**, 1235-1240. [doi:10.1016/j.molimm.2004.06.033](https://doi.org/10.1016/j.molimm.2004.06.033)
- [18] Wang, J., Yan, H.Y. and Kang, Q.S. (2005) Construction of a plant transgenic expression vector of peanut. *Journal of Wuhan Botanical Research*, **23**, 514-518.
- [19] Day R.C., Herridge R.P., Ambrose B.A. and Macknight R.C. (2008) Transcriptome analysis of proliferating arabidopsis endosperm reveals biological implications for the control of syncytial division, cytokinin signaling, and gene expression regulation. *Plant Physiology*, **148**, 1964-1984. [doi:10.1104/pp.108.128108](https://doi.org/10.1104/pp.108.128108)

- [20] Guo, Y., Qin, G., Gu, H. and Qu, L. (2009) *Dof5.6/HCA2*, a *Dof* transcription factor gene, regulates interfascicular cambium formation and vascular tissue development in arabidopsis. *Plant Cell*, **21**, 3518-3534. doi:10.1105/tpc.108.064139
- [21] Ogawa, M., Hanada, A., Yamauchi, Y., Kuwahara, A., Kamiya, Y. and Yamaguchi, S. (2003) Gibberellin biosynthesis and response during arabidopsis seed germination. *Plant Cell*, **15**, 1591-1604. doi:10.1105/tpc.011650
- [22] Depuydt, S., Trenkamp, S., Fernie, A.R., Elftieh, S., Renou, J., Vuylsteke, M., Holsters, M. and Vereecke, D. (2009) An integrated genomics approach to define niche establishment by *Rhodococcus fascians*. *Plant Physiology*, **149**, 1366-1386. doi:10.1104/pp.108.131805
- [23] Wellmer, F., Alves-Ferreira, M., Dubois, A., Riechmann, J.L. and Meyerowitz, E.M. (2006) Genome-wide analysis of gene expression during early arabidopsis flower development. *PLoS Genetics*, **2**, 1012-1024. doi:10.1371/journal.pgen.0020117
- [24] Fulton, L., Batoux, M., Vaddepalli, P., Yadav, R.K., Busch, W., Andersen, S.U., Jeong, S., Lohmann, J.U. and Schneitz, K. (2009) DETORQUEO, QUIRKY, and ZERZAUST represent novel components involved in organ development mediated by the receptor-like kinase STRUBBELIG in *Arabidopsis thaliana*. *PLoS Genetics*, **5**, 1-22. doi:10.1371/journal.pgen.1000355
- [25] Zhao, C., Craig, J.C., Petzold, H.E., Dickerman, A.W. and Beers, E.P. (2005) The xylem and phloem transcriptomes from secondary tissues of the arabidopsis root-hypocotyl. *Plant Physiology*, **138**, 803-818. doi:10.1104/pp.105.060202
- [26] Ward, J.M., Cufir, C.A., Denzel, M.A. and Neff, M.M. (2005) The *Dof* transcription factor OBP3 modulates phytochrome and cryptochrome signaling in arabidopsis. *Plant Cell*, **17**, 475-485. doi:10.1105/tpc.104.027722
- [27] Mouhu, K., Hytönen, T., Folta, K., Rantanen, M., Paulin, L., Auvinen, P. and Elomaa, P. (2009) Identification of flowering genes in strawberry, a perennial SD plant. *BMC Plant Biology*, **9**, 122-137. doi:10.1186/1471-2229-9-122
- [28] Borges, F., Gomes, G., Gardner, R., Moreno, N., McCormick, S., Feijó, J.A. and Becker, J.D. (2008) Comparative transcriptomics of arabidopsis sperm cells. *Plant Physiology*, **148**, 1168-1181. doi:10.1104/pp.108.125229
- [29] Para, A., Farré, E.M., Imaizumi, T., Pruneda-Paz, J.L., Harmon, F.G. and Kay, S.A. (2007) PRR3 is a vascular regulator of TOC1 stability in the arabidopsis circadian clock. *Plant Cell*, **19**, 3462-3473. doi:10.1105/tpc.107.054775
- [30] Chawade, A., Bräutigam, M., Lindlöf, A., Olsson, O. and Olsson, B. (2007) Putative cold acclimation pathways in *Arabidopsis thaliana* identified by a combined analysis of mRNA co-expression patterns, promoter motifs and transcription factors. *BMC Genomics*, **8**, 304-318. doi:10.1186/1471-2164-8-304
- [31] Kreps, J.A., Wu, Y., Chang, H., Zhu, T., Wang, X. and Harper, J.F. (2002) Transcriptome changes for arabidopsis in response to salt, osmotic, and cold stress. *Plant Physiology*, **130**, 2129-2141. doi:10.1104/pp.008532
- [32] Budhiraja, R., Hermkes, R., Müller, S., Schmidt, J., Colby, T., Panigrahi, K., Coupland, G. and Bachmair, A. (2009) Substrates related to chromatin and to RNA-dependent processes are modified by arabidopsis SUMO isoforms that differ in a conserved residue with influence on deSUMOylation. *Plant Physiology*, **149**, 1529-1540. doi:10.1104/pp.108.135053
- [33] Rohde, A., Morreel, K., Ralph, J., Goeminne, G., Hostyn, V., Rycke, R.D., Kushnir, S., Doorsselaere, J.V., Joseleau, J., Vuylsteke, M., Van Driessche, G., Van Beeumen, J., Messens, E. and Boerjan, W. (2004) Molecular phenotyping of the *pal1* and *pal2* mutants of *Arabidopsis thaliana* reveals far-reaching consequences on phenylpropanoid, amino acid, and carbohydrate metabolism. *Plant Cell*, **16**, 2749-2771. doi:10.1105/tpc.104.023705
- [34] Khanna, R., Shen, Y., Toledo-Ortiz, G., Kikis, E.A., Johannesson, H., Hwang, Y. and Quail, P.H. (2006) Functional profiling reveals that only a small number of phytochrome-regulated early-response genes in arabidopsis are necessary for optimal deetiolation. *Plant Cell*, **18**, 2157-2171. doi:10.1105/tpc.106.042200
- [35] Cheong, Y.H., Chang, H., Gupta, R., Wang, X., Zhu, T., Luan, S. (2002) Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in arabidopsis. *Plant Physiology*, **129**, 661-677. doi:10.1104/pp.002857
- [36] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389-3402. doi:10.1093/nar/25.17.3389
- [37] Stanke, M., Diekhans, M., Baertsch, R. and Haussler, D. (2008) Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics*, **24**, 637-644. doi:10.1093/bioinformatics/btn013
- [38] Yan, H., Ding, Y. and Wu, Y. (2009) Genetic engineering and molecular biology lab tutorial. Wuhan University Publisher, Wuhan, 3-9.
- [39] Skiryecz, A., Radziejowski, A., Busch, W., Hannah, M.A., Czeszejko, J., Kwaśniewski, M., Zanol, M., Lohmann, J.U., De Veylder, L., Witt, I. and Mueller-Roeber, B. (2008) The DOF transcription factor OBP1 is involved in cell cycle regulation in *Arabidopsis thaliana*. *Plant Journal*, **56**, 779-792. doi:10.1111/j.1365-3113X.2008.03641.x
- [40] Hoffmann, R. and Valencia, A. (2004) A gene network for navigating the literature. *Nature Genetics*, **36**, 664. doi:10.1038/ng0704-664
- [41] Hoffmann, R. and Valencia, A. (2004) iHOP-Information Hyperlinked over Proteins. <http://www.ihop-net.org/>
- [42] Varagona, M.J., Schmidt, R.J. and Raikhel, N.V. (1992) Nuclear localization signals required for nuclear targeting of the maize regulatory protein Opaque-2. *Plant Cell*, **4**, 1213-1227.
- [43] Hicks, C.R., Smith, H.M.S., Shieh, M. and Raikhel, N.V. (1995) Three classes of nuclear import signals bind to plant nuclei. *Plant Physiology*, **107**, 1055-1058. doi:10.1104/pp.107.4.1055
- [44] Jung, S., Tate, P.L., Horn, R., Kochert, G., Moore, K. and Abbott, A.G. (2003) The phylogenetic relationship of

- possible progenitors of the cultivated peanut, *Journal of Heredity*, **94**, 334-340. [doi:10.1093/jhered/esg061](https://doi.org/10.1093/jhered/esg061)
- [45] Imaizumi, T., Schultz, T.F., Harmon, F.G., Ho, L.A. and Kay, S.A. (2005) FKF1 F-Box protein mediates cyclic degradation of a repressor of CONSTANS in arabidopsis. *Science*, **309**, 293-297. [doi:10.1126/science.1110586](https://doi.org/10.1126/science.1110586)
- [46] Sawa, M., Nusinow, D.A., Kay, S.A. and Imaizumi, T. (2007) FKF1 and GIGANTEA complex formation is required for day-length measurement in arabidopsis. *Science*, **318**, 261-265. [doi:10.1126/science.1146994](https://doi.org/10.1126/science.1146994)
- [47] Fornara, F., Panigrahi, K.C.S., Gissot, L., Sauerbrunn, N., Rühl, M., Jarillo, J.A. and Coupland, G. (2009) Arabidopsis DOF transcription factors act redundantly to reduce CONSTANS expression and are essential for a photoperiodic flowering response. *Development Cell*, **17**, 75-86. [doi:10.1016/j.devcel.2009.06.015](https://doi.org/10.1016/j.devcel.2009.06.015)