



Article

Computational Disorder Analysis in Ethylene Response Factors Uncovers Binding Motifs Critical to Their Diverse Functions

Xiaolin Sun ¹, Nawar Malhis ², Bi Zhao ³ , Bin Xue ³, Joerg Gsponer ² and Erik H. A. Rikkerink ^{1,*}

¹ The New Zealand Institute for Plant & Food Research Ltd., 120 Mt. Albert Rd, Private Bag 92169, 1025 Auckland, New Zealand; xiaolin.sun@plantandfood.co.nz

² Michael Smith Laboratories—Centre for High-Throughput Biology, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada; nmalhis@chibi.ubc.ca (N.M.); gsponer@msl.ubc.ca (J.G.)

³ Department of Cell Biology, Microbiology and Molecular Biology, School of Natural Sciences and Mathematics, College of Arts and Sciences, University of South Florida, 4202 East Fowler Avenue, ISA 2015, Tampa, FL 33620-5150, USA; bizhao@mail.usf.edu (B.Z.); binxue@usf.edu (B.X.)

* Correspondence: erik.rikkerink@plantandfood.co.nz; Tel.: +64-9-925-7157

Received: 8 November 2019; Accepted: 18 December 2019; Published: 20 December 2019



Abstract: APETALA2/ETHYLENE RESPONSE FACTOR transcription factors (AP2/ERFs) play crucial roles in adaptation to stresses such as those caused by pathogens, wounding and cold. Although their name suggests a specific role in ethylene signalling, some ERF members also co-ordinate signals regulated by other key plant stress hormones such as jasmonate, abscisic acid and salicylate. We analysed a set of ERF proteins from three divergent plant species for intrinsically disorder regions containing conserved segments involved in protein–protein interaction known as Molecular Recognition Features (MoRFs). Then we correlated the MoRFs identified with a number of known functional features where these could be identified. Our analyses suggest that MoRFs, with plasticity in their disordered surroundings, are highly functional and may have been shuffled between related protein families driven by selection. A particularly important role may be played by the alpha helical component of the structured DNA binding domain to permit specificity. We also present examples of computationally identified MoRFs that have no known function and provide a valuable conceptual framework to link both disordered and ordered structural features within this family to diverse function.

Keywords: abiotic and biotic stress; AP2 domain; DNA binding; ERF transcription factors; intrinsically disordered protein; molecular recognition features (MoRFs); transcription activation; transcription repression

1. Introduction

Plants are constantly challenged to survive under environmental stresses. Abiotic stresses such as drought, salinity, and extreme temperatures influence plant growth and development by altering phytohormonal balance and redox processes, potentially leading to damage [1]. On the other hand, biotic stress stimulates the synthesis of phytohormones like salicylic acid (SA), jasmonic acid (JA), and ethylene (ET), which regulate specific immune responses [2]. Plants respond and adapt to these environmental stresses by triggering molecular signal transduction cascades to rapidly fine-tune their metabolic status in order to maintain homeostasis. Plant hormones function in these cascades as central integrators, reprogramming complex stress-adaptive signalling cascades [3]. Among the range of plant response events, these phytohormones are responsible for transcriptional changes that enable

physiological adaptation and trigger plant disease resistance. Transcription factors (TFs) are essential components of the integration of hormone signalling pathways in transcriptional relays.

APETALA2/ETHYLENE RESPONSE FACTOR (AP2/ERF) transcription factors (ERFs), one of the largest families of the plant-specific TFs, play a pivotal role in adaptation to biotic and abiotic stresses such as those caused by pathogens, wounding, and salinity by acting as key integrators of the SA, JA, and ET stress signalling pathways [4]. Plants contain a large array of ERFs, consisting of 122 proteins in *Arabidopsis* (*At*) and 139 in rice (*Os*) [5]. Each ERF protein contains a highly conserved DNA-binding domain (DBD or AP2 domain), which consists of about 57–59 amino acid residues. Many ERFs have been shown to bind specifically to the GCC box (AGCCGCC), a core sequence essential for ethylene-responsive transcription of genes, and are associated with biotic stress responses. While other ERFs bind to the C-repeat/dehydration-responsive-element (CRT/DRE; A/GCCGAC) cis-acting sequence in the promoters of target genes and are usually associated with responses to abiotic stress [6]. TFs recognize specific DNA sequences through their DBD, but typically recruit and assemble key transcription machinery partners through their transcription regulatory domains (TRDs). These two components integrate the protein-DNA and protein-protein recognition events that are central to function. Intrinsically disordered proteins (IDPs) or regions (IDR) are over-represented in proteins located in the nucleus and prevalent in cell signalling and transcriptional regulation [7,8]. IDRs are often characterized by the unique combination of high specificity and low affinity in their interactions with functional partners, which is important for transient protein-protein and protein-nucleic acid interactions occurring during signal transduction, recognition, and regulation events. Disorder also leads to binding promiscuity that enables one IDR to interact with multiple structurally diverse partners and binding plasticity through Molecular Recognition Features (MoRFs, an interaction prone segment within a long-disordered region). These MoRFs initiate recognition upon binding to various specific partners. Together these factors facilitate specific recognition by IDPs of their biological targets, allowing the formation of a flexible interaction network and conferring functional advantages on IDPs operating in stress response, signalling and regulation [9–11].

Bioinformatic studies revealed that 94% to 82% of the TFs possess IDRs as well as numerous MoRFs mainly found within their TRDs, relative to 54% to 18% of the proteins in two control data sets [8]. Experimental studies of TF families such as NAC and bZIP have demonstrated that IDRs dominate in their regulatory domains, providing the necessary structural flexibility for the TFs to interact with other TFs and/or signalling proteins in interaction networks and sense a wide variety of intracellular and extracellular stimuli [12,13]. MoRFs contained in these IDRs are conserved in each sub-group of the TFs families as similar binding motifs, enabling the functionalisation of TF subgroups by interacting with specific proteins of transcriptional complexes using either one or more of these specific motifs. Such mechanisms can exert fine-tuned transcriptional control over signalling pathways and sometimes involve specific binding-induced folding of the MoRFs [13]. Recent studies across a number of plant species have identified interesting correlations (positive or negative) between cellular complexity and an overall increase in disorder in some transcription factor families (e.g., MYB and bZIP) but not in other families such as ERFs [14,15]. ERFs are involved in a wide variety of biological functions in response to both biotic and abiotic stresses, and are required for cross-talk between signalling pathways crucial for adapting to environmental challenges. We present an intrinsic disorder analysis of ERFs from three divergent plant species, *Arabidopsis*, rice and moss. A number of computational approaches determined the degree of disorder or structure along the length of the ERF protein sequences and detected conserved patterns of disorder, MoRFs and ordered regions that could be correlated with ERF proteins in general or specific ERF subfamilies in particular. Functional specificities for some of the subgroups are correlated with their respective conserved MoRFs in TRDs. Local structural flexibility in DBDs of ERFs which, this analysis indicates, may be another key factor contributing to the functional specificity of individual members, could also help explain the plethora of stress phenotypes influenced by this large plant-specific disordered family. We illustrate the biological significance of these computational analyses with examples from phenotypes as diverse as cold stress,

cytokinin control, hypoxia, and pathogen response, where they correlate with protein motifs identified as critical components of the control exerted by these versatile proteins.

2. Results and Discussion

2.1. Phylogenetic Analysis and Classification of ERF Families

To generate a family tree and classify proteins into ERF families, a total of 375 ERF sequences from three different plant species were included in a phylogenetic analysis based on multiple alignment of the AP2/ERF domains and re-alignment of the regions outside of the AP2/ERF domains (Figures 1 and 2). Based on this family tree, these ERFs can be classified into 10 groups and 25 subgroups, not including four individual Moss ERFs which may represent other new groups. The resultant classification is mostly consistent with earlier reports of Arabidopsis and rice ERFs [5] with some subgroups refined. As shown in Figure 2, many of the ERFs classified previously still cluster in the same group, e.g., most of IIb, IIIb, IIIc, VIIa, VIIIa and IXc group members remain in the same groups. Some minor re-classifications are reported in Figure 2 (e.g., some members previously in IIb have moved to IIa, previous subgroups IIIId and IIIe are combined into subgroup IIIId, and IVa and IVb members are combined into subgroup IV). These refinements are largely based on the conserved motifs in regions outside of the ERF domain aligned under the subgroup classification.

Phylogenetic classification of the big ERF family is complicated by the highly variable sequences in the regions outside of the AP2/ERF domains while the high degree of similarity in the rather short conserved AP2/ERF domains reduces the statistical confidence in critical basal branch points. We suggest that this combination reduces the power of normal phylogenetic approaches to illustrate the relationship between groups or subgroups of genes. Hence the fact that some of the individual ERFs sequences have been reclassified into different subgroups compared to previous classifications [5,6] is not surprising. In order to provide more confidence in these groupings we therefore tested the support for each subgroup in sub-trees with the same outgroup (Pp1s107_32V6.1) to verify the final groupings. Isolated proteins that were positioned basal to the other members of the group identified by this approach were re-tested for membership with other families. As a result, a few ERFs were moved between groups or subgroups after this analysis, while no new grouping could be identified for others and these were retained as distant members of the group (identified in Figure 2). Four representative examples of these subgroupings complete with their branch support values are shown in Supplemental Figure S1. There are limitations to the degree to which we can predict functional and protein interaction characteristics for ERFs by depending on sequence alignment alone, therefore we subjected the 375 ERFs sequences to disorder and MoRF predictions to investigate the protein interaction characteristics in their 10 groups (25 subgroups). The members of each group/subgroup together with their alternative names and functions where available are listed in Supplementary Table S1.

2.2. Multiple Analyses Show that the TRDs Domain of ERF Proteins are Intrinsically Disordered

Previously, it was demonstrated that most TFs are intrinsically disordered in their TRDs [8,13]. We carried out multiple bioinformatics analyses using our ERF protein dataset. This revealed similar results to previous findings on other TFs in that all of the TRDs of the ERFs are intrinsically disordered while the DBDs largely consist of folded structures (Figure 3A,B). This had already been confirmed experimentally by NMR spectrometry in one example [16]. Within the DBDs of AtERF1 the first three β -strands are highly ordered while the following α -helical regions are relatively flexible.

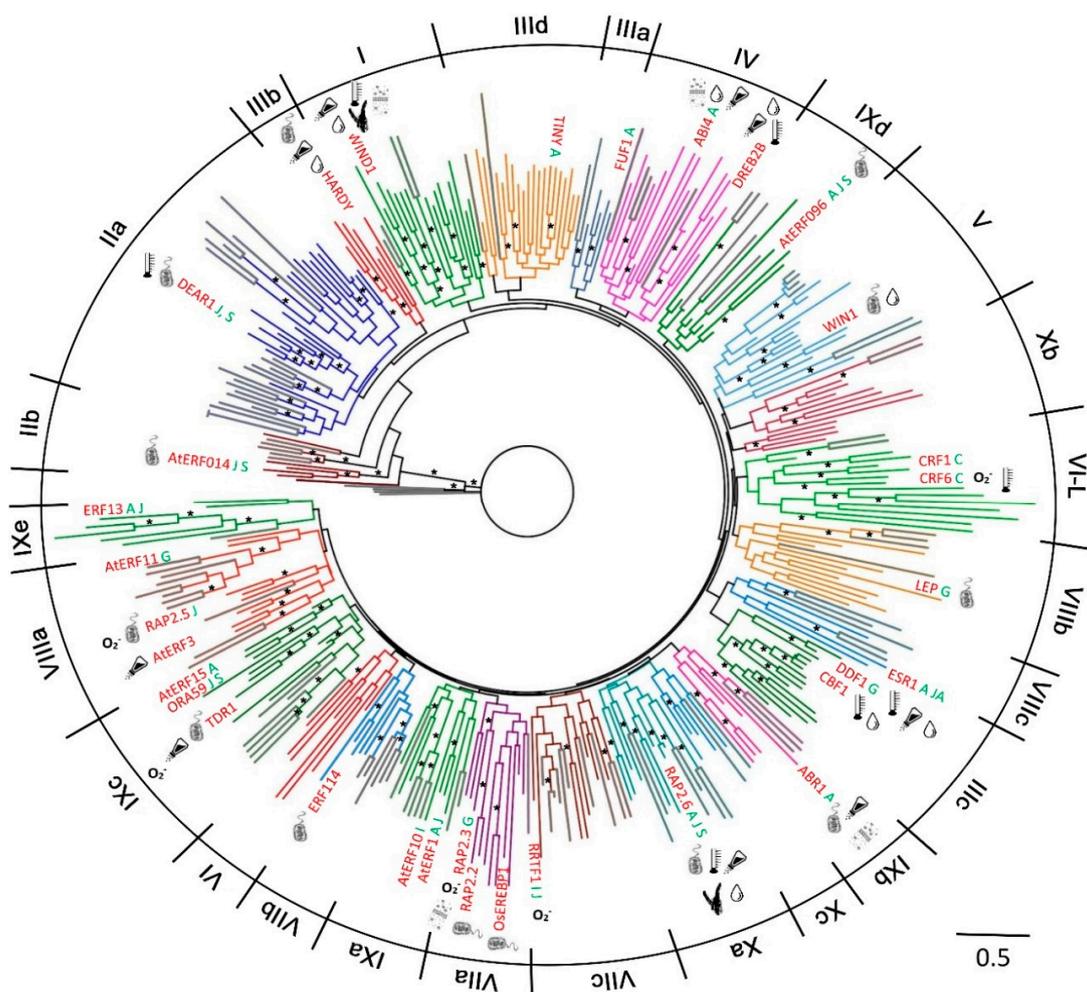


Figure 1. Phylogenetic tree of 375 ETHYLENE RESPONSE FACTOR transcription factors (ERFs) from Arabidopsis, rice and moss. The groups/subgroups are indicated in Roman numbers. Some key proteins with functional data are given for each of the subgroups where we could identify these and the types of stress they are associated with is also given to help orient the reader with respect to groups of related genes with functional data. Groups and sub-groups were initially identified from a tree including all members from all species. Given that the gene family is so diverse the support for the group and subgroup branches at the base of this tree was sometimes insufficient (less than 50%) to be confident that the grouping reflects the most likely phylogeny. Similar trees presented in the previous published analyses have often not shown any values for the support of groupings, branches with greater than 70% support are indicated by an asterisk. Angiosperm branches are coloured by clade with moss branches indicated in grey. Keys: hormones (single letter text after gene name)—A, Abscisic acid; C, Cytokinin; G, gibberellic acid; I, IAA/auxin; J, Jasmonic acid; S, salicylic acid. Stress symbols associated with individual identified genes (inside the circle) or un-identified genes in the subgroup (outside of the circle)— Biotic ; Cold/Heat ; Osmotic ; Oxidative ; Salt ; Water/Drought ; Wounding .

Group I	Group by Nakano and (Sakuma) †	Group IIa	Group by Nakano and (Sakuma)	Group IV	Group by Nakano and (Sakuma)
AT1G22190.1	Ib (A-6)	AT1G01250.1	IIIa (A-4)	AT1G75490.1	IVb (A-2)
AT1G36060.1	Ib (A-6)	AT1G19210.1	Iib	AT2G38340.1	IVa (A-2)
AT1G64380.1	Ib (A-6)	AT1G22810.1	Iic	AT2G40220.1	IVb (A-3)
AT1G78080.1	Ib (A-6)	AT1G33760.1	IIIa (A-4)	AT2G40340.1	IVa (A-2)
AT2G22200.1	Ib (A-6)	AT1G46768.1	IIa	AT2G40350.1	IVa (A-2)
AT4G39780.1	Ib (A-6)	AT1G71520.1	Iic	AT3G11020.1	IVa (A-2)
AT5G65130.1	Ib (A-6)	AT1G74930.1	Iib	AT3G23220.1	IXc (B-3c)
Os01g66270.1	Iic	AT2G23340.1	IIa	AT3G57600.1	IVb (A-2)
Os02g51670.1	Ib	AT3G50260.1	IIa	AT5G05410.1	IVa (A-2)
Os03g09170.1	Ib	AT4G06746.1	IIa	AT5G18450.1	IVb (A-2)
Os04g44670.1 ^①	Ia	AT4G36900.1	IIa	Os01g07120.1	IVa
Os05g49700.1	Ib	AT5G21960.1	Iib	Os03g64260.1	IXc
Os05g34730.1	Iic	AT5G67190.1	IIa	Os05g27930.1	IVb
Os06g11860.1	Ib	Os02g54050.1	Iib	Os05g39590.1	IVb
Os08g31580.1	Ib	Os04g5520.1	IIa	Os08g35240.1 ^①	Iib
Os09g20350.1	Ib	Os06g07030.1	IIa	Os08g45110.1	IVb
Os10g22600.1	Ib	Os06g09717.1		Pp1s23_27V6.1	
Pp1s21_214V6.1		Os06g09760.1	Iib	Pp1s81_44V6.1	
Pp1s79_25V6.1		Os06g09790.1	Iib	Pp1s162_92V6.1	
Pp1s187_47V6.1		Os06g09810.1	Iib	Pp1s292_5V6.1	
Pp1s335_8V6.1		Os06g11940.1	Iib		
		Pp1s8_127V6.1		Group V	Group by Nakano and (Sakuma)
Group Iib	Group by Nakano and (Sakuma)	Pp1s15_186V6.1		AT1G15360.1	Va (B-6)
AT1G44830.1	Iib (A-5)	Pp1s15_189V6.1		AT4G13620.1	Ib (A-6)
AT1G21910.1	Iib (A-5)	Pp1s21_320V6.1		AT4G17490.1	IXb (B-3b)
AT1G77640.1	Iib (A-5)	Pp1s29_45V6.1		AT5G11190.1	Va (B-6)
AT4G31060.1	Iib (A-5)	Pp1s55_253V6.1		AT5G25190.1	Va (B-6)
Os03g15660.1	Iib	Pp1s55_256V6.1		AT5G25390.1	Va (B-6)
Os06g10780.1	Iib	Pp1s60_269V6.1		AT5G47230.1	IXb (B-3b)
Pp1s11_280v6.1		Pp1s117_23V6.1		Os02g10760.1	Va
Pp1s87_98v6.1		Pp1s133_11V6.1		Os02g55380.1	Va
Pp1s119_43v6.1		Pp1s155_21V6.1		Os04g56150.1	Va
Pp1s86_33v6.1		Pp1s183_112V6.1		Os06g08340.1	Va
Pp1s373_18v6.1		Pp1s215_64V6.1		Os06g40150.1	Va
		Pp1s259_94V6.1		Os08g42550.1	Xc
Group IIIa	Group by Nakano and (Sakuma)	Pp1s259_96V6.1		Pp1s17_332V6.1	
AT1G71450.1	IIIa (A-4)	Pp1s259_97V6.1		Pp1s107_45V6.1	
AT5G07580.1	IXb (B-3b)	Pp1s263_66V6.1		Pp1s174_79V6.1	
AT5G61590.1	IXb (B-3b)	Pp1s263_67V6.1		Pp1s179_56V6.1	
Os02g35240.1	IIIb	Pp1s276_3V6.1		Pp1s225_102V6.1	
Os11g13840.1	IIIb	Pp1s480_12V6.1		Pp1s296_69V6.1	
Pp1s115_154V6.1		Pp1s480_16V6.1			
Pp1s357_12V6.1				Group VI	Group by Nakano and (Sakuma)
Group IIIb	Group by Nakano and (Sakuma)	Group IIIc	Group by Nakano and (Sakuma)	AT2G33710.1 ^①	Xc (B-4)
AT1G12630.1	IIIb (A-4)	AT1G71130.1	VI (B-5)	AT5G50080.1	Xc (B-4)
AT1G63040.1	IIIb (A-4)	AT1G77200.1	IIIe (A-4)	Os01g12440.1	VI
AT2G36450.1	IIIb (A-4)	AT2G25820.1	IIIe (A-4)	Os01g46870.1	VI
AT5G52020.1	IIIb (A-4)	AT2G44940.1	IIIc (A-1)	Os03g60120.1	VI
Os02g45420.1	IIIb	AT3G16280.1	IIIe (A-4)	Os05g25260.1	VI
Os10g38000.1	IIIc	AT3G60490.1	IIIc (A-4)	Os06g06540.1	VI
Group IIIc	Group by Nakano and (Sakuma)	AT4G32800.1	IIIc (A-4)	Os07g12510.1	VI
AT1G12610.1	IIIc (A-1)	AT5G11590.1	IIIe (A-4)		
AT1G63030.1	IIIc (A-1)	AT5G25810.1	IIIe (A-4)	Group VIIa	Group by Nakano and (Sakuma)
AT4G25470.1	IIIc (A-1)	Os01g10370.1	IIIc	AT1G53910.1	VIIa (B-2)
AT4G25480.1	IIIc (A-1)	Os02g13710.1	IIIe	AT1G72360.1	VIIa (B-2)
AT4G25490.1	IIIc (A-1)	Os02g43940.1	IIIc	AT2G47520.1	VIIa (B-2)
AT5G51990.1	IIIc (A-1)	Os02g43970.1	IIIc	AT3G14230.1	VIIa (B-2)
Os01g73770.1	IIIe	Os03g07830.1	IVb	AT3G16770.1	VIIa (B-2)
Os02g45450.1	IIIc	Os04g46400.1	IIIc	Os02g54160.1	VIIa
Os04g48350.1	IIIc	Os04g46410.1	IIIc	Os03g08470.1	VIIa
Os06g03670.1	IIIc	Os04g46440.1	IIIc	Os05g29810.1	VIIa
Os08g43200.1	IIIc	Os05g28350.1	IVb	Os06g09390.1	VIIa
Os08g43210.1	IIIc	Os06g36000.1	IIIe	Os07g42510.1	VIIa
Os09g35010.1	IIIc	Os10g41130.1	IIIc	Os09g26420.1	VIIa
Os09g35020.1	IIIc	Pp1s14_196V6.1 ^①		Pp1s140_167V6.1 ^①	
Os09g35030.1	IIIc	Pp1s28_144V6.1			

Figure 2. Cont.

Group VI-L	Group by Nakano and (Sakuma)	Group VIIIb	Group by Nakano and (Sakuma)	Group VIIc	Group by Nakano and (Sakuma)
AT1G22985.1 AT1G25470.1 AT1G49120.1 AT1G68550.1 AT2G46310.1 AT3G25890.1 AT3G61630.1 AT4G11140.1 AT4G23750.1 AT4G27950.1 AT5G53290.1 Os01g04020.1 Os08g27220.1 Os09g13940.1 Pp1s130_106V6.1 Pp1s138_190V6.1	VI (B-5) VI-L (B-6) VI-L (B-6) VI-L (B-6) VI (B-5) VI-L (B-6) VI (B-5) VI (B-5) VI (B-5) VI (B-5) VI (B-5) VI-L VI-L VI-L	AT5G07310.1 AT5G61890.1 Os02g38090.1 Os09g11460.1 Os09g11480.1 Pp1s90_184V6.1 Pp1s105_212V6.1 Pp1s124_33V6.1 Pp1s127_75V6.1 Pp1s149_253V6.1 Pp1s487_24V6.1	Xa (B-4) Xa (B-4) VIIIb VIIIb VIIIa	AT4G34410.1 Os02g06330.1 Os03g08460.1 Os03g08490.1 Os03g08500.1 Os04g34970.1 Os04g52090.1 Os06g47590.1 Pp1s1_34V6.1 Pp1s13_457V6.1 Pp1s25_13V6.1 Pp1s26_306V6.1 Pp1s37_293V6.1 Pp1s45_159V6.1 Pp1s61_230V6.1 Pp1s83_10V6.1 Pp1s199_50V6.1 Pp1s296_32V6.1	Xb (B-3) VIIIa VIIa VIIa VIIa Xa VIIIa VIIIa
Group VIIla	Group by Nakano and (Sakuma)	AT1G12890.1 AT1G28160.1 Os04g32790.1 Os07g47330.1 Os08g07700.1 Os12g41030.1 Os12g41060.1 Pp1s5_432V6.1 Pp1s6_75V6.1 Pp1s74_86V6.1 Pp1s221_67V6.1 [Ⓢ]	VIIIb (B-1) VIIIb (B-1) VIIIb VIIIb VIIIb XI XI	Group IXa	Group by Nakano and (Sakuma)
AT1G28360.1 AT1G28370.1 AT1G50640.1 AT1G53170.1 AT3G15210.1 AT3G20310.1 AT5G44210.1 Os01g58420.1 Os05g41760.1 Os05g41780.1 Pp1s7_472V6.1 Pp1s7_481V6.1 Pp1s21_239V6.1 Pp1s21_247V6.1 Pp1s23_8V6.1 Pp1s30_284V6.1 Pp1s38_262V6.1 Pp1s126_30V6.1 Pp1s293_53V6.1 Pp1s293_60V6.1 Pp1s340_16V6.1	VIIIa (B-1) VIIIa (B-1) VIIIa (B-1) VIIIa (B-1) VIIIa (B-1) VIIIa (B-1) VIIIa (B-1) VIIIa VIIIa VIIIa VIIIa	AT1G12980.1 AT1G24590.1 Os06g06970.1 Pp1s15_480V6.1 Pp1s145_10V6.1 Pp1s204_42V6.1 Pp1s233_111V6.1 Pp1s276_72V6.1	VIIIb (B-1) VIIIb (B-1)	AT1G03800.1 AT4G17500.1 AT5G47220.1 AT5G51190.1 AT5G61600.1 Os01g54890.1 Pp1s14_458V6.1 Pp1s15_234V6.1 Pp1s41_212V6.1 Pp1s112_23V6.1 Pp1s132_34V6.1 Pp1s181_69V6.1 Pp1s276_50V6.1	VIIIa (B-1) IXa (B-3a) IXa (B-3a) IXb (B-3b) IXb (B-3b) IXa
Group IXb	Group by Nakano and (Sakuma)	Group IXd	Group by Nakano and (Sakuma)	Group IXc	Group by Nakano and (Sakuma)
AT5G64750.1 Os01g21120.1 Os02g09650.1 Os02g43820.1 Os04g46240.1 Os04g46250.1 Pp1s58_260V6.1 Pp1s79_169V6.1 Pp1s243_70V6.1	Xa (B-4) VIIa VIIIa IXb IXb IXb	AT1G04370.1 AT2G20880.1 AT4G28140.1 AT5G43410.1 Os02g42585.1 Os03g05590.1 Os04g18650.1 Os05g49010.1 Os08g44960.1 Pp1s53_80V6.1 Pp1s60_148V6.1 Pp1s72_230V6.1 Pp1s120_13V6.1 Pp1s240_84V6.1 [Ⓢ]	IXc (B-3c) Ia (A-6) Ia (A-6) IXc (B-3c) IXc IXc IXc IXc IXc	AT1G06160.1 AT2G31230.1 AT3G23230.1 AT3G23240.1 AT4G18450.1 Os05g37640.1 Os07g22730.1 Os07g22770.1 Os09g39810.1 Os09g39850.1 Os10g30840.1 Pp1s2_404V6.1 Pp1s19_272V6.1 Pp1s20_184V6.1 Pp1s38_257V6.1 Pp1s121_50V6.1 Pp1s265_50V6.1 Pp1s345_35V6.1	IXc (B-3c) IXc (B-3c) IXc (B-3c) IXc (B-3c) IXc (B-3c) IXc IXc IXc IXc IXc IXc
AT2G44840.1 Os01g59780.1 Os03g07940.1 Os06g44750.1 Os07g03250.1 Os08g34360.1 Os10g41330.1 Pp1s2_410V6.1	IXa (B-3a) IXb	Group Xa	Group by Nakano and (Sakuma)	Group Xc	Group by Nakano and (Sakuma)
Group Xb	Group by Nakano and (Sakuma)	AT1G43160.1 AT5G13330.1 Os01g64790.1 Os02g34260.1 Os02g34270.1 Os02g43790.1 Os03g22170.1 Os04g46220.1 Os05g36100.1 Pp1s31_112V6.1 Pp1s67_144V6.1 [Ⓢ] Pp1s117_32V6.1 Pp1s136_219V6.1 Pp1s160_12V6.1 Pp1s259_104V6.1	Xa (B-4) Xa (B-4) Xa Xa Xa IXa VIIa IXa Xc	AT2G35700.1 AT4G16750.1 Os02g32140.1 Os04g32620.1 Os11g06770.1 Pp1s6_391V6.1 Pp1s289_49V6.1	III d (A-4) III d (A-4) Xc Xa Xa Xa
AT2G20350.1 AT5G19790.1 AT5G67000.1 AT5G67010.1 Os02g52670.1 Os04g57340.1 Os07g10410.1 Os07g38750.1 Os08g36920.1 Os09g28440.1 Os12g39330.1 Pp1s64_47V6.1 Pp1s195_38V6.1	Xb-L (B-6) Vb (B-6) Xb-L (B-6) Xb-L (B-6) Xb VIIIa Vb Vb Xb Xb Vb	AT1G43160.1 AT5G13330.1 Os01g64790.1 Os02g34260.1 Os02g34270.1 Os02g43790.1 Os03g22170.1 Os04g46220.1 Os05g36100.1 Pp1s31_112V6.1 Pp1s67_144V6.1 [Ⓢ] Pp1s117_32V6.1 Pp1s136_219V6.1 Pp1s160_12V6.1 Pp1s259_104V6.1	Xa (B-4) Xa (B-4) Xa Xa Xa IXa VIIa IXa Xc	Individuals	Group by Nakano and (Sakuma)
				Pp1s1_860V6.1 Pp1s107_32V6.1 Pp1s60_228V6.1 Pp1s6_134V6.1	N/A N/A N/A N/A

Figure 2. Groups of ERFs. More distantly related genes that have uncertain group membership status are labelled [Ⓢ]. On the phylogenetic trees they occupy a position that is basal to the rest of the members of the group. They have also been tested for membership of other groups but appear not to fit with any other group. [†] Grouping of this ERF by Nakano et al. (2006) where it exists while its grouping by Sakuma et al. (2002) where it exists is given in brackets.

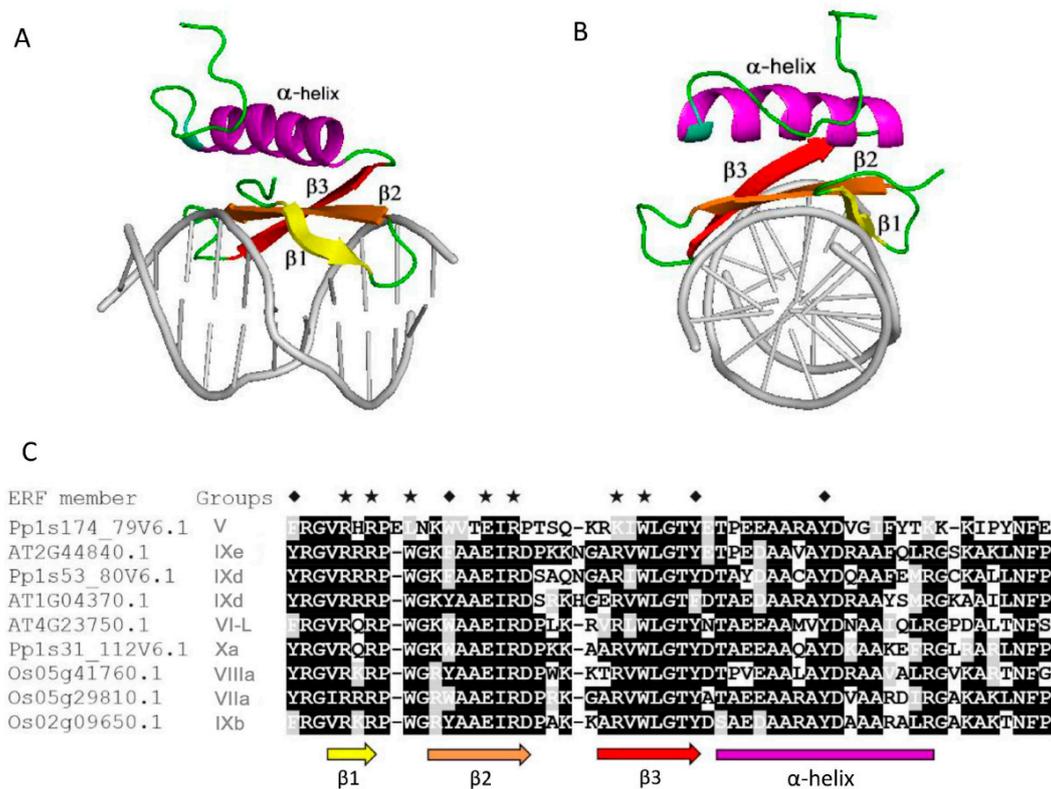


Figure 3. Structure of the DNA binding AP2 domain of AtERF1 complexed with DNA and sequence alignments. (A) Horizontal perspective. (B) Perpendicular perspective. (C) Sequence alignment of AP2 domain sequences randomly selected from a set of groups and different plant species (Arabidopsis, rice and moss). Black shadings indicate identical residues, grey shadings indicate conservatively substituted residues. The coloured arrows and bar represent β -strands and α -helix regions, respectively. The asterisks represent residues directly contacting with DNA. The diamonds represent highly conserved tyrosine or conservatively substituted tyrosine residues discussed in the text. The structural figures were prepared using PyMOL software with PDB code 1GCC.

The propensity for disorder is encoded in the composition of amino acid sequences. Low overall hydrophobicity and high net charge are widely used as one of the criteria for IDPs predictions [17,18]. Amino acid compositional profiles of ERF families from the three species analysed are similar to those typical of intrinsically disordered proteins. This is the case for full length sequences, but is even more pronounced when the mainly structured AP2 domain (DBD) is omitted (Figure 4A,C). The ERF sequences generally show a lack of order-promoting residues and an enrichment in disorder-promoting residues. The exception to the trend for over-representation of disorder promoting residues is an overall lower content of lysine residues in the disordered TRD regions, a unique compositional feature to be investigated. In contrast, the AP2 domains within the ERFs display significant differences to the compositional profile of IDPs and several residues that vary in amino acids composition compared to fully structured proteins (Figure 4B). The AP2 domains utilize an unusually high portion of W, A and R compared to both structured and disordered proteins. We deduce from sequences of the AP2 domains (Figure 3C) that these three residues, especially R, are frequently used for direct contact of DNA bases or for structural framing throughout the ordered AP2 domains.

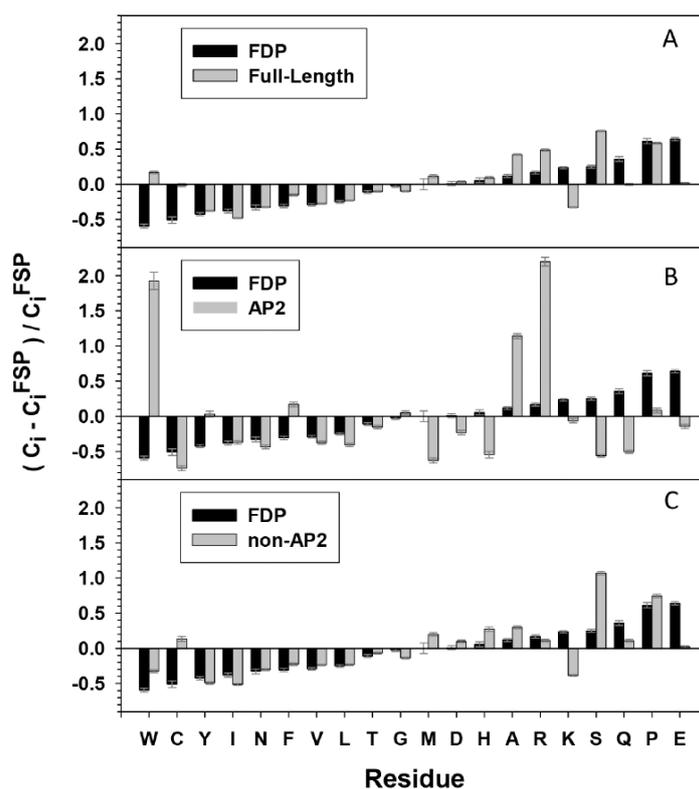


Figure 4. Relative compositional profile of amino acids of the 375 ERF sequences compared to Fully Structured Proteins (FSP). For (A) full-length ERF proteins, (B) AP2 domains only, and (C) AP2 domain-deleted ERF proteins. FDP shows the amino acids composition of Fully Disordered Proteins (FDP) compared to FSP. Probability distributions were estimated by computing means and confidence intervals of the relative frequencies of residues observed over a set of pseudo-replicate datasets obtained by bootstrap sampling of whole proteins from the original samples, the bootstrapping was repeated 10,000 times. The amino acid residues are placed along a continuous scale from the greatest contribution to ordered structure (W, left) to the greatest contribution to disorder (E, right).

Charge and hydropathy plot (CH-plot) of a protein can be used as a linear protein disorder classifier that differentiates proteins with substantial amounts of disorder from proteins with globular conformations [18,19]. Another binary disorder classifier, cumulative distribution function (CDF) analysis, differentiates all proteins containing any IDRs from those that contain mainly folded domains [20]. Simultaneous CDF-CH plotting provides an even more accurate prediction for a wider range of proteins [21]. The quadrants of CDF-CH phase space correspond to the following expectations: Q1 (upper-right), proteins predicted to be disordered by CH-plots, but ordered by CDFs; Q2 (lower-right), ordered proteins in terms of both methods; Q3 (lower-left), proteins predicted to be disordered by CDFs, but compact by CH-plots; Q4 (upper-left), proteins predicted to be disordered by both methods.

Full-length ERF proteins all cluster in the quadrants Q4 and Q3 of the CH-CDF phase space (Figure 5A). The AP2 domains as a set move to the right of this plot to occupy space between Q1 and Q2 (Figure 5B), while the AP2 domain-deleted regions as a set move further to the left within Q4 and Q3 (Figure 5C). This demonstrates that the TRDs of the ERFs proteins (located outside of the AP2 domain) are intrinsically disordered while most of the AP2 domains (DBD) are predicted to be ordered in terms of the CDF criterion. On the other hand, more than half of full-length ERFs and the AP2 domain-deleted regions have CH distances located in Q3 while the rest of the AP2 domain regions have CH distances falling in quadrant Q4. CH-plots by themselves appear not to be a suitable disorder predictor in the case of ERF proteins, indicating that charge and hydropathy are not the dominant

factors driving their structural disorder. This is also reflected in the fact that the folded AP2 domains (DBD) contain a high portion of R, E and D polar residues interspersed within a frame formed by conserved Y, W and F that facilitate DNA binding.

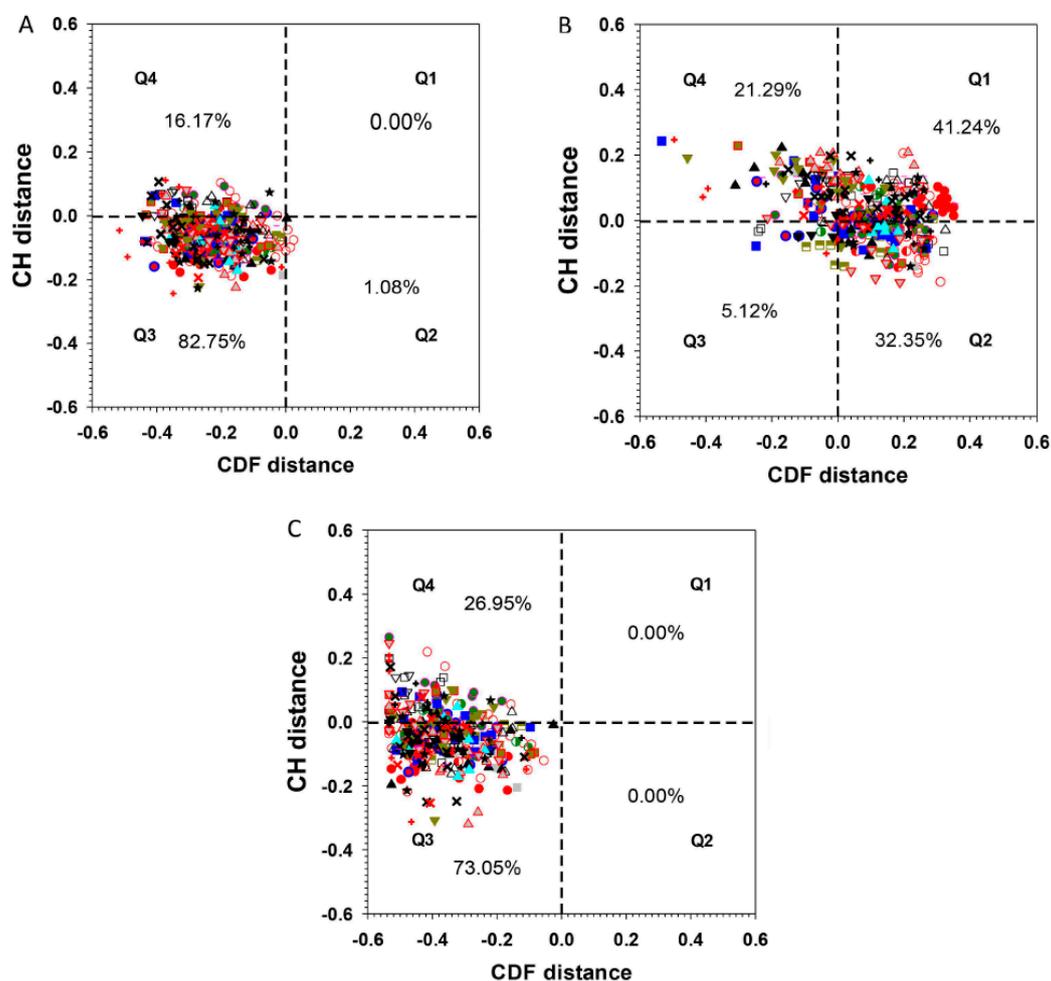


Figure 5. Charge-hydropathy and cumulative distribution function (CH-CDF) plot. For (A) full-length ERF proteins, (B) ERF-AP2 domains only, and (C) AP2-deleted ERF proteins. The percentages indicate total number of ERFs in each of the quadrant. The symbols representing the members of each group/subgroup: I ●, IIa ○, IIb ★, IIIa ●, IIIb ●, IIIc ■, IIId ●, IV ●, V △, VI-L ■, VI ▼, VIIa ▼, VIIb ▽, VIIc +, VIIIa △, VIIIb ■, VIIIc ■, IXa +, IXb -, IXc ▼, IXd □, IXe ▲, Xa ×, Xb ▲, Xc ×.

Statistics on the fraction of intrinsically disorder (promoting) amino acids (IDAA%) for both full-length, AP2 domains alone, and the AP2 domains-deleted ERFs are given in Figure 6. These figures reveal that the non-AP2 regions of ERFs host the majority of disorder promoting residues, with IDAA% spanning from 30% to 80% in the case of full length sequences and 40% to 100% in the case of AP2 domains-deleted sequences only (Figure 6A,C), in comparison with 0% to 40% in the case of AP2 domains alone (Figure 6B). In addition, the differential correlation of IDAA% between the AP2 domains alone (Figure 6D), AP2 domains-deleted (Figure 6E) and full-length sequences of ERFs also show that the non-AP2 regions are the predominant determinant of intrinsic disorder propensity for ERFs.

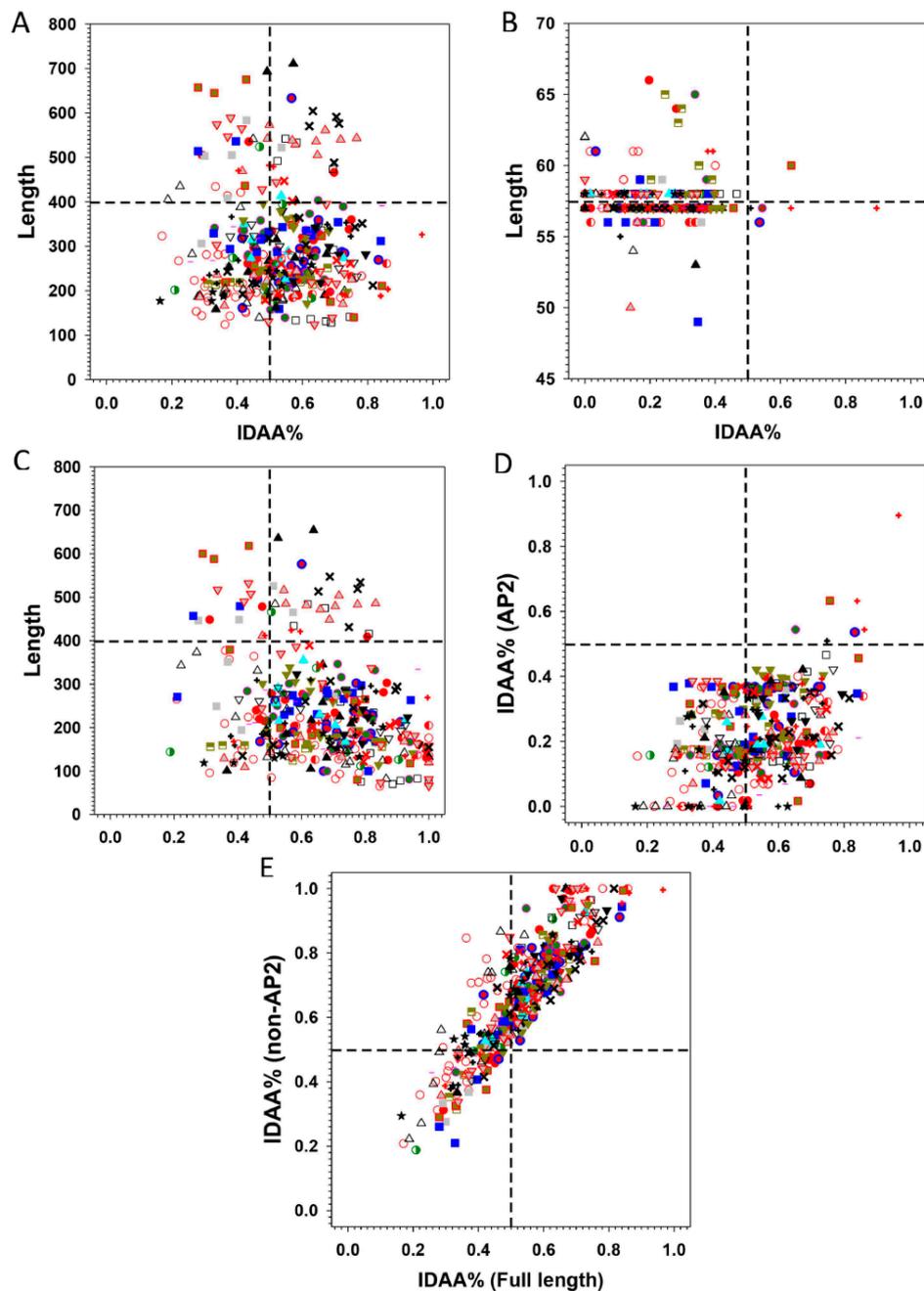


Figure 6. Comparison of different sequence features with the fraction of intrinsically disordered amino acids (IDAA%). (A) Length distribution of full-length ERFs, (B) Length distribution of AP2 domains, (C) Length distribution of AP2-deleted ERFs, (D) Comparison of the IDAA% between full-length and AP2 domains, (E) Comparison of the IDAA% between full-length proteins and AP2-deleted ERFs. The symbols representing the members of each group/subgroup: I ●, IIa ○, IIb ★, IIIa ●, IIIb ●, IIIc ■, IIId ●, IV ●, V △, VI-L ■, VI ▼, VIIa ▼, VIIb ▼, VIIc +, VIIId △, VIIIb ■, VIIIc ■, IXa +, IXb −, IXc ▽, IXd □, IXe ▲, Xa ×, Xb ▲, Xc ×.

Distribution of the intrinsic disorder propensities along the length of sequences for each subgroup are shown in PONDR-FIT plots (Figure 7 and Supplemental Figure S2). Six members in each subgroup are presented to illustrate the overall distribution pattern of disordered residues and MoRFs identified by using MoRFchibi. The DNA binding AP2 domains marked by grey thick bars are universally predicted to consist of mainly ordered segments (Figure 7), i.e., structurally folded zone. The regions outside of the AP2 domains, harbouring TRDs, are generally located in the disordered zone (disorder score >0.5

threshold). Frequent downward spikes representing local short segments with order propensity within the long-disordered region and generally coincide with MoRFs (marked by various short coloured bars) that typically act as binding sites in the interactions with partners during transcriptional regulation. These results reinforce the findings that ERFs have folded and highly conserved AP2 domains, flanked on both sides by intrinsically disordered regions containing TRDs. Solution structures solved by 2D NMR spectrometry [16] reveal that these AP2 domains fold in three β -strands followed by an α -helix (Figure 3A,B). Experimental evidence from plant NAC and bZIP TFs show a similar general arrangement with the ordered DBDs flanked by the disordered TRDs [12,22].

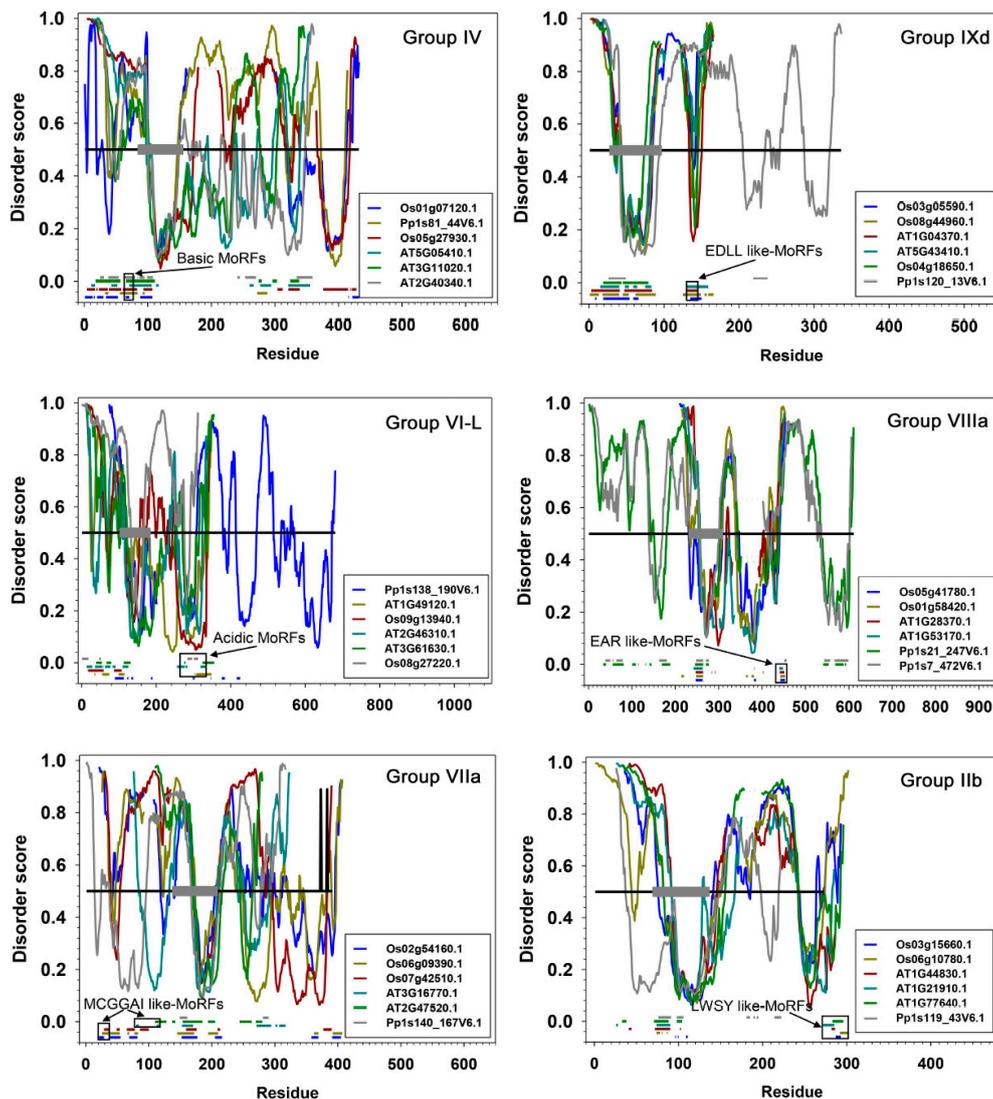


Figure 7. Disorder prediction of ERFs. Six groups with six representative members selected in each group are presented while all groups are presented in Supplemental Figure S2. Disorder score was predicted using PONDR-FIT. The disorder curves were aligned using IDAlign. The grey thick bars indicate the span of the AP2 (DBD) domains including the three highly structured beta sheets and a less ordered alpha helix at the C-terminal end of this domain. The black horizontal line represents the disorder threshold score (≥ 0.5). The short horizontal colour bars at the bottom of each panel are Molecular Recognition Features (MoRFs) predicted using MoRFchibi. In each group/subgroup presented, arrows point at a specific type of MoRF shown in the black boxes.

IDRs and low complexity sequences have similar compositional bias—more disorder-promoting residues and less order-promoting residues [17,23]. Simultaneous use of sequence complexity analysis

2.3. Phosphorylation/Dephosphorylation is Involved in ERF Protein Interactions

IDPs manipulate molecular recognition and signalling cascades by utilizing posttranslational modifications (PTMs) such as phosphorylation. For example, phosphorylation has been postulated to (1) fine-tune the electrostatic interactions of disordered regions in protein complexes [25]; to drive folding of an IDP as a regulatory switch [26]; (2) decide between two tiers of the plant immune system as it copes with attacks by type III effectors through a serine/threonine phosphoswitch of the disordered RIN4 protein [27]; and (3) activate the GA signalling pathway by dephosphorylation of serine/threonine residues and phosphorylation of tyrosine residues of the disordered DELLA proteins prior to triggering GA-induced DELLA protein degradation [28]. Compared to structured regions, IDRs possess a much higher probability of phosphorylation, partly attributable to the greater steric access for kinases/phosphatases in open disordered regions and the high portion of disorder-promoting serine residues in these regions [29]. Some of these principles apply equally to other types of PTMs and so disorder facilitates a plethora of alternative protein states to regulate complicated interaction networks and explains why these proteins can occupy pivotal hub positions.

Phosphorylation is already known to play a role in the function of ERF proteins. A DREB2A ortholog in the grain crop *Pennisetum glaucum* differentially interacts with DNA when an uncharacterised threonine residue is modified and allows it to play a role in the abiotic response in this stress tolerant plant [30]. Two Ser-Pro sites in the C-terminus of AtERF104 are specifically phosphorylated by the MAP kinase MPK6 and increase its stability [31]. Tyrosine sites in the N- and C-terminus of AtERF13 (now deduced to be disordered) were identified to be phosphorylated by cysteine-rich receptor like kinases AtCRK2 and AtCRK3 [32], and AtERF13 is linked with the ABA response [33]. Disorder-based phosphorylation analyses using MuSite for the ERF proteins show that the disordered non-AP2 domain regions display a higher fraction of predicted phosphorylation sites than the full-length sequences and this is true for serine and threonine residues (Figure 9A and Table 1). This result is similar to most other plant IDP families studied [34]. Conversely the ordered AP2 domains of ERFs showed an unexpectedly high proportion of tyrosine residues prone to phosphorylation when compared to the non-AP2 domain regions (Figure 9B). These ordered regions showed a lower proportion of serine/threonine residues prone to phosphorylation. This indicates it is likely that tyrosine phosphorylation plays a special role in AP2 domains in altering DNA binding specificities. As a result of this unusual high portion of tyrosine phosphorylation, the sum of all three types of potential phosphorylation sites (S, T and Y) offsets the S/T phosphorylation preferences in the non-AP2 regions (Figure 9C) and uplift the total potential phosphorylation sites in AP2 domain regions versus full length sequences (Figure 9D). Four notable tyrosine residues in the AP2 domain regions are marked with diamonds in Figure 3C, three occur within the beta sheets and one in the middle of the alpha helix. Interestingly the tyrosines occurring in the alpha helical region are highly conserved whereas those in the more structured core of the beta sheets can be conservatively substituted by other aromatic residues. In the beta sheet cases we postulate that aromatic side chain of these residues play a role in DNA binding through aromatic ring interaction with spatially nearby sidechains from other residues, which has been exemplified in the structure of AtERF1 [16]. These interactions either do not feature phosphorylation or phosphorylation of tyrosine residues introduces further binding or functional diversity between the different families or subfamilies. In contrast, phosphorylation of the highly conserved tyrosines in the less ordered alpha helical region could either encourage or discourage a more extensive DNA interaction with the alpha helix that could be used to fine tune regulation by increasing or decreasing the strength of binding or altering its equilibrium state. The alpha helical fragment in the AP2 domain structure of AtERF1 (Figure 3) is capped on the beta sheet through a hydrogen bond between the hydroxyl group of tyrosine 186 in the α helix and arginine 147 in the β 1 strand, implying the potential impact of phosphorylation of tyrosine 186 on the helical structure and interaction between the helix and the beta strand and (we suggest) the resultant DNA binding specificities. Our analyses indicate that future experimental investigation should focus on these tyrosine residues and how they utilize phosphorylation to affect DNA binding.

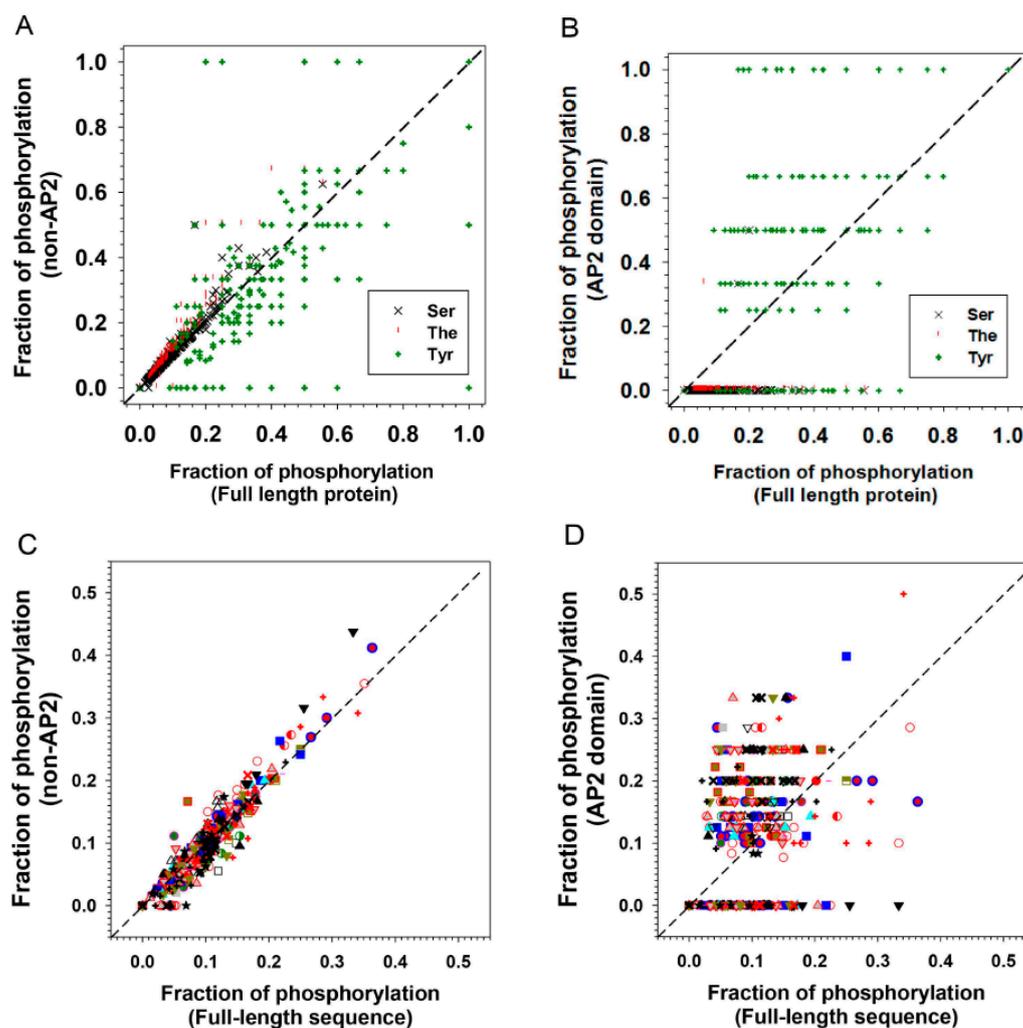


Figure 9. Distribution of phosphorylation sites of 375 ERF proteins. The phosphorylation sites were predicted using MuSite. (A) Comparison of fractions of predicted phosphorylation site for each ERF in all groups/subgroups between full-length sequence and AP2-deleted ERF and (B) between full-length sequence and AP2 domains. (C) Comparison of fractions of predicted phosphorylation site for all three types (Ser, The, Tyr) in all ERF sequences between full-length sequence and AP2-deleted ERFs and (D) between full-length sequence and AP2 domains. The symbols representing the members of each group/subgroup: I ●, IIa ○, IIb ★, IIIa ●, IIIb ●, IIIc ■, IIId ●, IV ●, V △, VI-L ■, VI ▼, VIIa ▼, VIIb ▼, VIIc +, VIIId △, VIIIf ■, VIIIfc ■, IXa +, IXb -, IXc ▼, IXd □, IXe ▲, Xa ×, Xb ▲, Xc ×.

Table 1. Fraction of predicted phosphorylation of 375 ERFs by residue types and domains.

	S	T	Y	All (S + T + Y)
AP2 domain only	0.28%	0.10%	34.29%	11.15%
Non-AP2 domain	8.59%	5.93%	24.85%	9.67%
Full-length	8.09%	4.92%	30.32%	9.97%

2.4. Analyses of MoRFs in ERF Groups and Their Functional Implications

In addition to the DNA-binding AP2 domains, ERFs contain functionally important regions outside of AP2 domains, regulating transcriptional activation and protein–protein interactions, defined as the TRD. Generally, many short conserved amino acids motifs are often identified in this region for plant transcription factors such as MYB, WRKY, NAC and GRAS [11,35,36]. ERFs are no exception in this regard. In particular the *Arabidopsis thaliana* ERF family has been found to possess 53 conserved

motifs distributed in the regions outside of the AP2 domains across all of the ERF groups/subgroups as determined by sequence alignments [5]. However, the functions of most of these conserved motifs are unknown and there are relatively few clues about which of these motifs are important and what kind of functions they may be important for. Given that ERFs are IDPs with long IDRs in the TRDs regions, it is plausible that analyses for MoRFs can provide a unique way to specifically identify the conserved motifs in these regions that act as key interacting sites in transcriptional regulation. As revealed in our previous studies, MoRFs are often conserved within a subgroup, indicating that members in the same subgroup sharing these MoRFs have similar functions [34]. We carried out MoRFs analyses for all 375 selected ERFs and identified conserved MoRFs (potential interacting sites for partners) in each of the subgroups.

The DNA binding AP2 domains in most subgroups of the ERFs family have been found to contain MoRFs (Figure 10A). The MoRFs identified in the AP2 domains are found within nearly all of the $\beta 1$ and $\beta 2$ strands, meaning that the $\beta 1$ and $\beta 2$ strands are the major binding region in response to interaction between the ERFs and DNA moiety. This is supported by the fact that more than 70% of the direct DNA contact residues (R, E and W) are in the $\beta 1$ and $\beta 2$ strand regions and the rest of the direct DNA contact residues are in the $\beta 3$ strand. This highlights the reliability of the new algorithm MoRFchibi employed in this analysis. Despite being in the DNA binding AP2 domain, the latter parts of the AP2 domains which harbour the α -helix generally lack MoRFs. On the other hand, these α -helical regions have been shown to be more structurally flexible with significant low complexity and more variable sequences. These results are consistent with the observation from the NMR solution structure (Figure 3) that the α -helical region of AtERF1 impacts on ERF DNA binding indirectly rather than by direct binding. We note some exceptions that in three subgroups (V, IXc and IXd) MoRFs can be identified across the whole AP2 domain, indicating that the α -helix may also be involved in direct DNA binding in these individual cases. In contrast, the five rice ERFs in subgroup IXe have no MoRFs identified within their AP2 domains and possess unusual sequence arrangements in which stretches of multiple histidine residues are inserted in between the $\beta 1$ and $\beta 2$ strands while a highly conserved glutamate residue (typically involved in direct contact with DNA) is missing (Figure 10B). To contribute to the divergent DNA-Binding specificities of ERFs, it would be worth investigating experimentally how this clade of ERFs bind DNA. Alternatively, if they do not bind DNA, whether they are involved in other ways such as competing for limited protein complex partners with ERFs that do in fact bind DNA. This same group contains the Arabidopsis AtERF13 (AT2G44840) protein referred to above that is subject to phosphorylation and involved in ABA response.

It has been shown that the DNA binding basic regions in the bZIP transcription factor family, in the absence of DNA, populate an ensemble of highly dynamic transient structures, either completely structured, containing a certain amount of helical structure, or being completely disordered [13]. Experimental evidence revealed that the basic regions of bZIP uniformly form α -helical conformations once in complex with DNA [37]. Our MoRFs and low complexity analyses suggest that the latter part of AP2 domains of ERFs could also be flexible or even disordered in the unbound state, although this region has been shown to be folded as an α -helix when bound with DNA in the case of AtERF1 [16].

All of the MoRFs identified outside of the AP2 domains are shown in Supplemental Table S2. Many of these are exclusive to specific subgroups in the phylogenetic tree, and likely contribute to similarity in protein interactions within the same subgroup. Although the functions of most of these subgroup-conserved MoRFs are yet to be investigated, the functional roles of some examples that correlate with sites important for transcriptional regulation are discussed below. Many of the conserved MoRFs identified in this study have unique features in their amino acid compositions (Supplemental Table S2). The most common pattern is that either hydrophobic or aromatic residues repeat to form the framework of the MoRFs with acidic residues flanking the repeated hydrophobic/aromatic residues (Figure 11A, also see IIIId C2, V C2, VI-L C1*, VI-L C4, VIIa N3, VIIIa N2*, VIIIa C3*, IXb N1*, IXb C2*, and Xb C2* in Supplemental Table S2). Being associated with transcriptional activation [38], such patterns of acidic MoRFs have also been found in other IDPs from the GRAS family. In GRAS

proteins the repeated hydrophobic/aromatic residues interspersed between acidic residues have been experimentally demonstrated to directly bind to partners (as in the DELLA subfamily; [34,39,40]), or are responsible for strong transcriptional activation as in the LISCL subfamily; [41]. We postulate that the repeated hydrophobic/aromatic residues together with the acidic residues play similar functional roles in ERFs. Some positively charged residue repeats are identified in the conserved MoRFs of ERFs as well, mostly located in the N-termini of the ERFs (Figure 11B, also see IIIc N1*, VIIIa N4* and IXc C1 in Supplemental Table S2), which are also likely to be involved in specific functions for these groups. Further experimental investigation of these unique patterns of MoRFs is now warranted as this analysis has provided a short-cut to identify likely sites for elucidating the structure-function paradigms of these various ERF families.

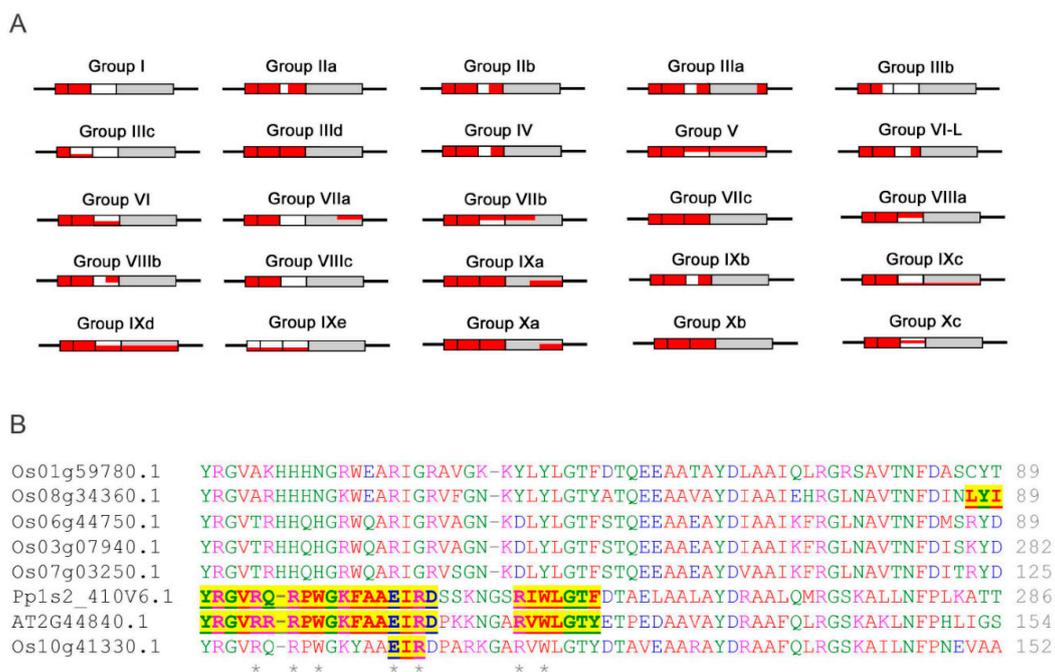


Figure 10. (A) Schematic representation of conserved MoRFs in the DNA binding AP2 domains of the ERF groups/subgroups. The $\beta 1$, $\beta 2$ and $\beta 3$ regions are indicated as a set of three boxes followed by the α -helix regions (grey box). The red colour represents conserved MoRFs distributed in each of the groups/subgroups and their location with respect to the β sheets and α helices. Partial red colouring versus white (in β sheets), or grey (in α -helices), reflects incomplete prevalence of MoRFs among members of the groups/subgroups. (B) More detailed sequence alignment of the AP2 domains of subgroup IXe. The yellow highlighted regions are MoRFs identified for the three members with DNA contacting residues marked by asterisks. Additional alignments of AP2/ERF regions in other subgroups are shown in Supplemental Figure S3.

Among the subgroup-conserved MoRFs, some overlap directly with motifs that have been experimentally studied. For example three C-terminal ERF-associated amphiphilic repression (EAR) motif-like MoRFs (Figure 11C, also see VIIb C3* and VIIa C1 in Supplemental Table S2) have been identified and the EAR motifs have been shown to repress GCC box-mediated transcription via transient assays [42,43]. Three EDLL motif-like MoRFs (Figure 11D, also see IXd C1 and IXc C2 in Supplemental Table S2) have been identified where acidic glutamic and aspartic residues are interspersed with hydrophobic leucine residues, constituting acidic transcriptional activation domains (TAD) often found at the C-terminus of transcription factors [44,45]. The EDLL motif has the ability to activate transcription and can confer activation domain function on heterologous DNA-binding proteins [46]. Six LWSY motif-like MoRFs (Figure 11E,F, also see IIa C5, IIb C2*, IIIb C1, and VIIa C2* in Supplemental Table S2) are found to be conserved at the end of the C-terminus of many ERF proteins,

they are mostly DREB genes encoding transcription activators that recognise the C-repeat elements of cold-responsive genes (CORs), function in regulation of drought, cold, and salinity responsive gene expression and promote plant resistance to cold stress [47,48]. It has been noted that the cold responsive CBF ERFs (assigned to group IIIc here) contain a series of motifs characterised by clusters of hydrophobic residues delineated by short stretches of acidic residues and prolines within the trans-activating C-terminus [49]. The most highly conserved of these hydrophobic clusters (called HC6 by Wang and colleagues) coincides with the MoRF IIIc C1. Although much of the functional research on the ERFs comes from Arabidopsis, there is also evidence that orthologues from other plants perform functions in similar pathways. For example, overexpression of CBF1 from tomato could also boost cold tolerance in Arabidopsis and mutation of the highly conserved tryptophan residue shown in MoRF IIIc C1 destroyed the trans-activating ability of the tomato orthologue of AtCBF1 [50], highlighting that this MoRF corresponds to a critical motif in these important regulators of plant cold tolerance.

A	VI-L C3*		
AT1G49120.1	L LDfdpli DdF kle DfP mdDlgl LgD		
AT2G46310.1	L PDlfrsg DyF wds E .ItpDpl. F lD		
AT5G53290.1	P LDtpfl. DnY fn. EsPp . E ms. I fd		
AT4G27950.1	P LDitfl. DnY fn. EsLp . D is. I fd		
AT3G61630.1	L PDlslag E c F wdt E .IapD l . F lD		
B	IV N2*		
Os05g27930.1	KKGcM KGKGGPEN		
Pp1s162_92V6.1	KKGaM KGKGGPEN		
AT1G75490.1	RKGC M RKGGPDN		
Pp1s81_44V6.1	KKGC M KGKGGPEN		
AT3G11020.1	KKGC M KGKGGPDN		
AT2G40340.1	RKGC M KGKGGPEN		
AT2G40350.1	RKGC M KGKGGPEN		
AT5G05410.1	KKGC M KGKGGPEN		
Pp1s23_27V6.1	KRGcM KGKGGPEN		
Os08g45110.1	RKGC M KGKGGPEN		
Os01g07120.1	KKGC M MAGKGGPEN		
AT2G38340.1	KKGC M RKGGPEN		
AT5G18450.1	RKGC M KGKGGPEN		
Pp1s292_5V6.1	KKGaM KGKGGPEN		
C	IIa C2*		
AT4G06746.1	L DMN k l P ...Eay.. G		
AT4G36900.1	V DLN k l P pd.Etsdd D		
Os04g55520.1	P DLN q a P ps.Ensda E		
AT3G50260.1	V DLN k l P pd.Enl l dd D		
AT2G23340.1	V DLN k l P pd.Ens d g D		
AT5G67190.1	V DLN k l P pdPEs s de E		
Os06g07030.1	P DLN r e P pt.Dts s de E		
AT1G46768.1	P DLN q i P ...Esg.. D		
D	III d C1*		
AT5G25810.1	E e L g E .I V E L P s L		
AT4G32800.1	D d L d E .I V K L P s L		
Os02g13710.1	D e L . E aI V E L P r L		
AT1G77200.1	E e L s E .I V E L P s L		
AT5G11590.1	E e L g E .I V E L P s L		
AT2G25820.1	E . L g D .I V E L P s L		
Os06g36000.1	E e L e E .I V E L P p I		
AT3G16280.1	E r L . E kI V E L P n I		
E	IXa C3*		
Pp1s181_69V6.1	DGDeattyn WEPrLWSF		
Pp1s15_234V6.1	DGEegsst. WEPhLWSY		
Pp1s41_212V6.1	DGEegsnt. WEPhLWSY		
Pp1s14_458V6.1	DGEegsti. WEPhLWSY		
Pp1s276_50V6.1	DGEegsti. WEPhLWSY		
Pp1s132_34V6.1	SDEgtsays WEPrLWSF		
F	IIIc C1		
Os09g35010.1	DM pLWS		
Os04g48350.1	EV nLWS		
AT1G63030.1	D LsLWS		
AT5G51990.1	D VsLWS		
AT1G12610.1	DM sLWS		
AT4G25470.1	D VsLWS		
Os08g43200.1	EM eLWS		
AT4G25490.1	D VsLWS		
Os09g35020.1	DI aLWS		
Os06g03670.1	D VtLWS		
Os09g35030.1	D VpLWS		
AT4G25480.1	D VsLWS		
Os02g45450.1	EV qLWS		
Os08g43210.1	D VsLWS		
G	VIIa N1		
Os02g54160.1	M CG G A I IhhLkgh P		
AT3G16770.1	M CG G A I IsdYa.. P		
Os03g08470.1	M CG G A I IaeFipa P		
Os05g29810.1	M CG G A I IadFvp. P		
Os06g09390.1	M CG G A I IsdLipp P		
AT2G47520.1	M CG G A I IsdFi.. W		
Os09g26420.1	M CG G A I IsgFip. P		
Os07g42510.1	M CG S S I L gdLhl. P		
AT3G14230.1	M CG G A I IsdFipp P		
AT1G53910.1	M CG G A I IsdFipp P		

Figure 11. Examples of conserved MoRFs predicted outside of AP2 domains. (A) Acidic MoRFs, (B) Basic MoRFs, (C) ERF-associated amphiphilic repression (EAR) motif-like MoRFs, (D) EDLL motif-like MoRFs, (E) and (F) LWSY motif-like MoRFs, (G) MCGGAI(I/L) motif-like MoRFs. The consensus residues of each patterns are in bold. The name of each MoRFs starts with group/subgroup number followed by N (prior to AP2 domain) or C (after AP2 domain), asterisks indicate that this MoRF exclusively belongs to this group/subgroup.

There are two highly conserved MCGGAI(I/L) motif-like MoRFs (Figure 11G, also see VIIc N1* in Supplemental Table S2) that exist in the N-terminal part of the ERFs. In tomato LeERF2 and its homologues from monocot and dicot species are characterized by this unique N-terminal signature. Deletion studies revealed that this motif is required neither for nuclear localization nor for binding to the GCC box, indicating specific functions for the ERFs harbouring the MCGGAI(I/L) signature [51]. This motif is found at the N-terminus of the group VIIa where it appears to be a defining feature that results in these ERFs becoming oxygen and nitric oxide-dependent substrates for the N-end rule targeted degradation [52]. This enables these proteins to act as homeostatic sensors of hypoxia and links them to an ancient and conserved branch of the ubiquitin-mediated proteosomal degradation system. The Met and Cys residues serve as a degron, are specifically cleaved, arginylated (but only in the presence of O or NO) and then recognised by a specific E3 ligase and poly-ubiquitinated. Therefore, it is likely this MoRF is specifically involved in this process.

The Cytokinin Response Factor (CRF) group form yet another distinct subset of ERFs that belong to group VI-L and have been shown to regulate leaf development as part of the cytokinin signal transduction pathway in Arabidopsis. The conserved motifs in the N-terminus of this group of proteins, including one motif that appears to be characteristic of this group that is described as the CRF domain [53]. This corresponds to MoRF VI-L N1 (Supplemental Table S2) that we identified in the disordered N-terminus of group VI-L. These ERF proteins affect a set of cytokinin responsive genes that largely overlap with the type-B ARR positive cytokinin regulators and are dependent on the two-component histidine kinase system for cytokinin signal transduction [54]. The CRF domain has been demonstrated to play a key role in interactions between the CRF proteins and with members of the histidine-phosphotransferase component of the cytokinin signalling system [55] and illustrate yet again the importance of MoRFs in ERFs for driving key interactions within the hormone signalling networks that they operate in.

Compared to the 53 conserved motifs identified by sequence alignments in the regions outside of AP2 domains for the 122 Arabidopsis thaliana ERFs [5], our disorder-based potential protein binding sites analyses have resulted in a total 72 conserved MoRFs among the 375 ERFs from three plant species. Among these are 41 conserved MoRFs found in the Arabidopsis thaliana ERFs. This suggests that some conserved motifs generated from sequence alignment may function not by way of modifying direct protein–protein interactions but instead by alternative properties such as maintenance of structural integrity.

2.5. Conserved MoRFs are Potentially Mobile Through Evolution

Given the much greater evolutionary distance between mosses and Angiosperms, the moss ERFs are generally distantly related to Arabidopsis and rice ERFs. It can be observed from the conserved MoRFs (Supplemental Table S2) that the moss ERFs frequently have some independently conserved MoRFs themselves. The phylogenetic tree (Figures 1 and 2) also support this tendency with moss ERFs quite often clustering into local clades.

Our analyses suggest the possibility that conserved MoRFs could be shuffled between different protein families during evolution. We have identified a number of cases where conserved MoRFs common in one family or subfamily also occur in a distantly related family or subfamily. These include the C-terminal LWSY-like MoRFs in groups IIa, IIb, IIIb, IIIc, VIIa, VIIIc, and IXa, the C-terminal EDLL-like MoRFs in groups VI, VIIb, IXc and IXd, and the EAR-like MoRFs in groups VI, VIIb, VIIIa, IXc, and IXe (see Supplemental Table S2). Some conserved MoRFs are even found in totally unrelated protein families. For instance, EAR-like MoRFs are also identified in the zinc-finger transcriptional repressors ZAT7, ZAT10, and ZAT11 and play a role in modulating responses to multiple stress factors including drought, cold, pathogens, as well as hormone signalling in general [56]. More recently they were also found in the pathogen effector XopD [57]. The disordered nature of sequences surrounding MoRFs implies that they need not necessarily be highly conserved to retain their disordered structure. This means there is considerable potential for these MoRFs to act as small modules that can be moved

around between different proteins by non-homologous recombination events between genes within their respective disordered regions. Where such an event creates a novel functional combination that is beneficial, evolution can then operate to select for the new combination of these modules. This could increase its frequency in a population and sometimes it could then become fixed if the selective advantage is high enough. Given that these types of events provide a method for coordinating protein–protein interaction networks and thereby can initiate cross-talk and extra robustness in networks, the potential for significant evolutionary advantages gained by such a route is substantial. We surmise that this may well be one of the forces operating on re-distribution of conserved MoRFs between related (and occasionally even unrelated) protein families that display disorder. An extreme case of this may be the presence of two EAR-like MoRFs in the pathogen effector XopD which are required to target the tomato ERF SIERF4 for desumoylation and suppress the ethylene response [57]. These may be acting as a mimic of the host EAR-like MoRFs present in ERF groups VI, VIIb, VIIIa, IXc, and IXe. While a horizontal gene transfer event between the host and pathogen is a possible origin for this novel effector construct, given that the length of some of these MoRFs is short some could alternatively evolve by convergent evolution pathways.

3. Materials and Methods

3.1. Sequence and Phylogenetic Analysis

All ERF sequences of three diverse species were selected for this study. This includes 122 *Arabidopsis* (*Arabidopsis thaliana*) ERFs from TAIR, 133 rice ERFs (*Oryza sativa* L. subsp. *japonica*) from PlantTFDB (<http://planttfdb.cbi.pku.edu.cn/index.php>) and 120 Moss ERFs (*Physcomitrella patens*) from iTAK (<http://bioinfo.bti.cornell.edu/cgi-bin/itak/index.cgi>). Some individual ERFs with unusual features such as double or incomplete AP2 domains are excluded from this analysis. All sequence alignments were carried out using Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) with manual adjustment in some alignments by placing a greater emphasis on motifs harbouring MoRFs. A phylogenetic tree was constructed using the neighbour-joining method with bootstrapping (1000) in Geneious.

3.2. Amino Acids Compositional Profile Analysis

The relative compositional profile is defined by $(C_i - C_i^{\text{FSP}})/C_i^{\text{FSP}}$, where C_i is the absolute compositional percentage of *i*th amino acid in the query dataset and C_i^{FSP} is the absolute compositional percentage of *i*th amino acid in the fully-structured protein dataset (FSP) [58]. The FSP is composed of sequences which are monomeric, non-membrane protein crystal structures extracted from Protein Data Bank. The FSP candidate sequences with 25% or higher sequence identity were filtered to keep the longest one, and the final FSP dataset has 554 protein sequences and 113,895 residues.

3.3. Charge-Hydropathy (CH) and Cumulative Distribution Function (CDF) Plot

ERF sequences were subjected to a combined analysis of Charge-Hydropathy (CH) plot [18] and Cumulative Distribution Function (CDF) plot [59]. CH plot consists of plotting the averaged Kyte-Doolittle hydrophobicity of a protein on the X-axis against the averaged net charges of the same protein on the Y-axis. There is a boundary line that divides the CH plot into the upper-left region where disordered proteins are normally located and the lower-right region where structured proteins are normally located. CDF plot is a cumulated histogram of disordered residues with various disordered score. Similar to CH plots there is also a boundary line identified in the CDF plot to separate disordered proteins from structured proteins. The disordered scores and CDF distance used for CDF plot are from the PONDR-VSL2 disorder predictor [20]. The distances to the boundary lines in both CDF and CH plots from a specific protein were further used as coordinates on x- and y-axes to develop a combined CH-CDF plot.

3.4. Disorder Predictions

Protein disorder predictions were carried out using predictor PONDR-FIT which is one of the predictors with high prediction accuracy [60]. Disorder curves were aligned using IDAlign (IDAA; [61]). Molecular Recognition Features (MoRFs) were identified using MoRFChibi predictors (<http://morf.chibi.ubc.ca:8080/mcw/index.html>; [62]) and were also plotted onto the PONDR-FIT profiles. Sequence complexity analyses were carried out by using an iterative algorithm (CAST; <http://athina.biol.uoa.gr/CAST/>; [24]). The fraction of intrinsically disordered amino acids (IDAA%) is the percentage of predicted disorder (promoting) amino acids in a certain length of sequence.

3.5. Predictions of Phosphorylation Sites

The predictions of phosphorylated sites are carried out by using the Musite predictor which integrates both protein disorder and amino acid specificity. Musite is a bioinformatics tool for predicting both general and kinase-specific protein phosphorylation sites. (<http://musite.net/>; [63]).

4. Conclusions

MoRFs are short interaction-prone fragments located within extended disorder regions and can undergo disorder-to-order transitions upon binding to their interacting partners. We suggest the MoRFs we have identified in ERFs use these properties to act in complex networks. Several of the conserved MoRFs identified here are verified by overlapping with experimentally confirmed binding sites including the AP2 DNA binding sites and other motifs outside of AP2 domains referred to above. Given the disordered nature of the TRDs, it is likely that many of the protein interactions between ERFs and their partners follow the disorder-to-order transition mode in response to signal stimuli. In this way they can regulate complicated interaction networks that can be turned on, off, or maintained in a dynamic state of equilibrium. One example of this is the EDLL-like activation MoRF that can overcome the repression of the EAR-like MoRF under some circumstances [46]. Our analysis suggests conserved MoRFs should become focal points for functional studies of each ERF subgroup. The AP2 domain-associated components can lead to divergent DNA binding specificities and, combined with the likely disorder/flexibility-based interactions of the MoRFs in the N- and C- terminal disordered domains, can reasonably explain the versatility of regulatory roles played by this important plant protein family. We suggest therefore that disorder in ERFs has played a crucial role, preparing sessile plants for the multitude of stresses that they will inevitably face in their lifetime and allowing flowering plants to adapt to a rich variety of biological niches.

Supplementary Materials: Supplementary materials can be found at <http://www.mdpi.com/1422-0067/21/1/74/s1>. Table S1. New and previous grouping of ERFs (where applicable) along with their alternative names. Table S2. Summary of conserved MoRFs in carboxy and amino terminal regions outside of the AP2 domains. Figure S1. Phylogenetic trees of a representative set of groups/subgroups. Figure S2. Disorder prediction of ERFs. Figure S3. Alignments of some ERF subgroups. File S1. Table S1 References.

Author Contributions: X.S.: E.H.A.R. and B.X. conceived and designed the research, N.M. and J.G. contributed MoRF-Chibi prediction software and analysis, X.S., B.Z. and B.X. contributed to computational and phylogenetic data analysis, X.S. and E.H.A.R. wrote the original draft of the manuscript with revisions from B.X., N.M. and J.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Discovery Science grant DS2002 from The New Zealand Institute for Plant & Food Research Ltd. and research grant C11X1804 from the New Zealand Ministry of Business, Innovation and Employment.

Acknowledgments: We thank Marion Wood and Niels Nieuwenhuizen for critical review of the manuscript prior to submission.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Muller, M.; Munne-Bosch, S. Ethylene response factors: A key regulatory hub in hormone and stress signaling. *Plant Physiol.* **2015**, *169*, 32–41. [[CrossRef](#)] [[PubMed](#)]
2. Pre, M.; Atallah, M.; Champion, A.; De Vos, M.; Pieterse, C.M.J.; Memelink, J. The AP2/ERF domain transcription factor ORA59 integrates jasmonic acid and ethylene signals in plant defense. *Plant Physiol.* **2008**, *147*, 1347–1357. [[CrossRef](#)] [[PubMed](#)]
3. Gollack, D.; Li, C.; Mohan, H.; Probst, N. Tolerance to drought and salt stress in plants: Unraveling the signaling networks. *Front. Plant Sci.* **2014**, *5*, 151. [[CrossRef](#)] [[PubMed](#)]
4. Dey, S.; Vlot, A.C. Ethylene responsive factors in the orchestration of stress responses in monocotyledonous plants. *Front. Plant Sci.* **2015**, *6*, 640. [[CrossRef](#)] [[PubMed](#)]
5. Nakano, T.; Suzuki, K.; Fujimura, T.; Shinshi, H. Genome-wide analysis of the ERF gene family in Arabidopsis and Rice. *Plant Physiol.* **2006**, *140*, 411–432. [[CrossRef](#)]
6. Sakuma, Y.; Liu, Q.; Dubouzet, J.G.; Abe, H.; Shinozaki, K.; Yamaguchi-Shinozaki, K. DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBS, transcription factors involved in dehydration- and cold-inducible gene expression. *Biochem. Biophys. Res. Commun.* **2002**, *290*, 998–1009. [[CrossRef](#)]
7. Uversky, V.N. The mysterious unfoldome: Structureless, underappreciated, yet vital part of any given proteome. *J. Biomed. Biotechnol.* **2010**, *2010*, 568068. [[CrossRef](#)]
8. Liu, J.G.; Perumal, N.B.; Oldfield, C.J.; Su, E.W.; Uversky, V.N.; Dunker, A.K. Intrinsic disorder in transcription factors. *Biochemistry* **2006**, *45*, 6873–6888. [[CrossRef](#)]
9. Patil, A.; Nakamura, H. Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks. *FEBS Lett.* **2006**, *580*, 2041–2045. [[CrossRef](#)]
10. Oldfield, C.J.; Meng, J.; Yang, J.Y.; Yang, M.Q.; Uversky, V.N.; Dunker, A.K. Flexible nets: Disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics* **2008**, *9* (Suppl. S1). [[CrossRef](#)]
11. Sun, X.; Jones, W.T.; Rikkerink, E.H.A. GRAS proteins: The versatile roles of intrinsically disordered proteins in plant signalling. *Biochem. J.* **2012**, *442*, 1–12. [[CrossRef](#)] [[PubMed](#)]
12. Kjaersgaard, T.; Jensen, M.K.; Christiansen, M.W.; Gregersen, P.; Kragelund, B.B.; Skriver, K. Senescence-associated barley NAC (NAM, ATAF1,2, CUC) transcription factor interacts with radical-induced cell death 1 through a disordered regulatory domain. *J. Biol. Chem.* **2011**, *286*, 35418–35429. [[CrossRef](#)] [[PubMed](#)]
13. Sun, X.; Rikkerink, E.H.A.; Jones, W.T.; Uversky, V.N. Multifarious roles of intrinsic disorder in proteins illustrate its broad impact on plant biology. *Plant Cell* **2013**, *25*, 38–55. [[CrossRef](#)] [[PubMed](#)]
14. Yruela, I.; Oldfield, C.J.; Niklas, K.J.; Dunker, A.K. Evidence for a strong correlation between transcription factor protein disorder and organismic complexity. *Genome Biol. Evol.* **2017**, *9*, 1248–1265. [[CrossRef](#)]
15. Niklas, K.J.; Dunker, A.K.; Yruela, I. The evolutionary origins of cell type diversification and the role of intrinsically disordered proteins. *J. Exp. Bot.* **2018**, *69*, 1437–1446. [[CrossRef](#)] [[PubMed](#)]
16. Allen, M.D.; Yamasaki, K.; Ohme-Takagi, M.; Tateno, M.; Suzuki, M. A novel mode of DNA recognition by a Beta-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. *EMBO J.* **1998**, *17*, 5484–5496. [[CrossRef](#)]
17. Dunker, A.K.; Lawson, J.D.; Brown, C.J.; Williams, R.M.; Romero, P.; Oh, J.S.; Oldfield, C.J.; Campen, A.M.; Ratliff, C.R.; Hipps, K.W.; et al. Intrinsically disordered protein. *J. Mol. Gr. Model.* **2001**, *19*, 26–59. [[CrossRef](#)]
18. Uversky, V.N.; Gillespie, J.R.; Fink, A.L. Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins-Struct. Funct. Genet.* **2000**, *41*, 415–427. [[CrossRef](#)]
19. Sun, X.; Jones, W.T.; Uversky, V.N. Applications of bioinformatics and experimental methods to intrinsic disorder-based protein-protein interactions. In *Protein Engineering*; Kaumaya, P., Ed.; InTech: Rijeka, Croatia, 2012; pp. 181–206.
20. Xue, B.; Oldfield, C.J.; Dunker, A.K.; Uversky, V.N. Cdf it all: Consensus prediction of intrinsically disordered proteins based on various cumulative distribution functions. *FEBS Lett.* **2009**, *583*, 1469–1474. [[CrossRef](#)]
21. Mohan, A.; Sullivan, W.J.; Radivojac, P.; Dunker, A.K.; Uversky, V.N. Intrinsic disorder in pathogenic and non-pathogenic microbes: Discovering and analyzing the unfoldomes of early-branching eukaryotes. *Mol. Biosyst.* **2008**, *4*, 328–340. [[CrossRef](#)]

22. Yoon, M.K.; Shin, J.; Choi, G.; Choi, B.S. Intrinsically unstructured n-terminal domain of BZIP transcription factor HY5. *Proteins-Struct. Funct. Bioinforma.* **2006**, *65*, 856–866. [[CrossRef](#)] [[PubMed](#)]
23. Romero, P.; Obradovic, Z.; Li, X.H.; Garner, E.C.; Brown, C.J.; Dunker, A.K. Sequence complexity of disordered protein. *Proteins-Struct. Funct. Genet.* **2001**, *42*, 38–48. [[CrossRef](#)]
24. Promponas, V.J.; Enright, A.J.; Tsoka, S.; Kreil, D.P.; Leroy, C.; Hamodrakas, S.; Sander, C.; Ouzounis, C.A. Cast: An iterative algorithm for the complexity analysis of sequence tracts. *Bioinformatics* **2000**, *16*, 915–922. [[CrossRef](#)]
25. Mittag, T.; Kay, L.E.; Forman-Kay, J.D. Protein dynamics and conformational disorder in molecular recognition. *J. Mol. Recognit.* **2010**, *23*, 105–116. [[CrossRef](#)] [[PubMed](#)]
26. Bah, A.; Vernon, R.M.; Siddiqui, Z.; Krzeminski, M.; Muhandiram, R.; Zhao, C.; Sonenberg, N.; Kay, L.E.; Forman-Kay, J.D. Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature* **2015**, *519*, 106–109. [[CrossRef](#)]
27. Chung, E.-H.; El-Kasmi, F.; He, Y.; Loehr, A.; Dangl, J.L. A plant phosphoswitch platform repeatedly targeted by type III effector proteins regulates the output of both tiers of plant immune receptors. *Cell Host Microbe* **2014**, *16*, 484–494. [[CrossRef](#)]
28. Hussain, A.; Cao, D.N.; Peng, J.R. Identification of conserved tyrosine residues important for gibberellin sensitivity of Arabidopsis RGL2 protein. *Planta* **2007**, *226*, 475–483. [[CrossRef](#)]
29. Iakoucheva, L.M.; Radivojac, P.; Brown, C.J.; O'Connor, T.R.; Sikes, J.G.; Obradovic, Z.; Dunker, A.K. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* **2004**, *32*, 1037–1049. [[CrossRef](#)]
30. Agarwal, P.; Agarwal, P.K.; Nair, S.; Sopory, S.K.; Reddy, M.K. Stress-inducible DREB2A transcription factor from *Pennisetum glaucum* is a phosphoprotein and its phosphorylation negatively regulates its DNA-binding activity. *Mol. Genet. Genomics* **2007**, *277*, 189–198. [[CrossRef](#)]
31. Bethke, G.; Unthan, T.; Uhrig, J.; Poschl, Y.; Gust, A.; Scheel, D.; Lee, J. Flg22 regulates the release of an ethylene response factor substrate from MAP kinase 6 in *Arabidopsis thaliana* via ethylene signaling. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 8067–8072. [[CrossRef](#)]
32. Nemoto, K.; Takemori, N.; Seki, M.; Shinozaki, K.; Sawasaki, T. Members of the plant CRK superfamily are capable of trans- and autophosphorylation of tyrosine residues. *J. Biol. Chem.* **2015**, *290*, 16665–16677. [[CrossRef](#)] [[PubMed](#)]
33. Lee, S.J.; Park, J.H.; Lee, M.H.; Yu, J.H.; Kim, S.Y. Isolation and functional characterization of CE1 binding proteins. *BMC Plant Biol.* **2010**, *10*, 277. [[CrossRef](#)] [[PubMed](#)]
34. Sun, X.; Xue, B.; Jones, W.T.; Rikkerink, E.; Dunker, A.K.; Uversky, V.N. A functionally required unfoldome from the plant kingdom: Intrinsically disordered N-terminal domains of GRAS proteins are involved in molecular recognition during plant development. *Plant Mol. Biol.* **2011**, *77*, 205–223. [[CrossRef](#)] [[PubMed](#)]
35. Kranz, H.D.; Denekamp, M.; Greco, R.; Jin, H.; Leyva, A.; Meissner, R.C.; Petroni, K.; Urzainqui, A.; Bevan, M.; Martin, C.; et al. Towards functional characterisation of the members of the R2R3-MYB gene family from *Arabidopsis thaliana*. *Plant J.* **1998**, *16*, 263–276. [[CrossRef](#)] [[PubMed](#)]
36. Eulgem, T.; Rushton, P.J.; Robatzek, S.; Somssich, I.E. The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* **2000**, *5*, 199–206. [[CrossRef](#)]
37. Hollenbeck, J.J.; McClain, D.L.; Oakley, M.G. The role of helix stabilizing residues in GCN4 basic region folding and DNA binding. *Protein Sci.* **2002**, *11*, 2740–2747. [[CrossRef](#)]
38. Triezenberg, S.J. Structure and function of transcriptional activation domains. *Curr. Opin. Genet. Dev.* **1995**, *5*, 190–196. [[CrossRef](#)]
39. Sun, X.L.; Jones, W.T.; Harvey, D.; Edwards, P.J.B.; Pascal, S.M.; Kirk, C.; Considine, T.; Sheerin, D.J.; Rakonjac, J.; Oldfield, C.J.; et al. N-terminal domains of DELLA proteins are intrinsically unstructured in the absence of interaction with GID1/gibberellin acid receptors. *J. Biol. Chem.* **2010**, *285*, 11557–11571. [[CrossRef](#)]
40. Murase, K.; Hirano, Y.; Sun, T.P.; Hakoshima, T. Gibberellin-induced DELLA recognition by the gibberellin receptor GID1. *Nature* **2008**, *456*, 459–464. [[CrossRef](#)]
41. Morohashi, K.; Minami, M.; Takase, H.; Hotta, Y.; Hiratsuka, K. Isolation and characterization of a novel *Gras* gene that regulates meiosis-associated gene expression. *J. Biol. Chem.* **2003**, *278*, 20865–20873. [[CrossRef](#)]

42. Fujimoto, S.Y.; Ohta, M.; Usui, A.; Shinshi, H.; Ohme-Takagi, M. Arabidopsis ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box-mediated gene expression. *Plant Cell* **2000**, *12*, 393–404.
43. Ohta, M.; Matsui, K.; Hiratsu, K.; Shinshi, H.; Ohme-Takagi, M. Repression domains of class II ERF transcriptional repressors share an essential motif for active repression. *Plant Cell* **2001**, *13*, 1959–1968. [[CrossRef](#)] [[PubMed](#)]
44. Blair, W.S.; Bogerd, H.; Cullen, B.R. Genetic-analysis indicates that the human foamy virus BEL-1 protein contains a transcription activation domain of the acidic class. *J. Virol.* **1994**, *68*, 3803–3808. [[PubMed](#)]
45. Tanaka, M.; Herr, W. Reconstitution of transcriptional activation domains by reiteration of short peptide segments reveals the modular organization of a glutamine-rich activation domain. *Mol. Cell. Biol.* **1994**, *14*, 6056–6067. [[CrossRef](#)] [[PubMed](#)]
46. Tiwari, S.B.; Belachew, A.; Ma, S.F.; Young, M.; Ade, J.; Shen, Y.; Marion, C.M.; Holtan, H.E.; Bailey, A.; Stone, J.K.; et al. The EDLL motif: A potent plant transcriptional activation domain from AP2/ERF transcription factors. *Plant J.* **2012**, *70*, 855–865. [[CrossRef](#)]
47. Mao, D.H.; Chen, C.Y. Colinearity and similar expression pattern of rice DREB1S reveal their functional conservation in the cold-responsive pathway. *PLoS ONE* **2012**, *7*, e47275. [[CrossRef](#)]
48. Dubouzet, J.G.; Sakuma, Y.; Ito, Y.; Kasuga, M.; Dubouzet, E.G.; Miura, S.; Seki, M.; Shinozaki, K.; Yamaguchi-Shinozaki, K. OsDREB genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. *Plant J.* **2003**, *33*, 751–763. [[CrossRef](#)] [[PubMed](#)]
49. Wang, Z.B.; Triezenberg, S.J.; Thomashow, M.F.; Stockinger, E.J. Multiple hydrophobic motifs in Arabidopsis CBF1 COOH-terminus provide functional redundancy in trans-activation. *Plant Mol. Biol.* **2005**, *58*, 543–559. [[CrossRef](#)]
50. Zhang, L.L.; Li, Z.J.; Li, J.F.; Wang, A.X. Ectopic overexpression of SSCBF1, a CRT/DRE-binding factor from the nightshade plant *Solanum lycopersicoides*, confers freezing and salt tolerance in transgenic Arabidopsis. *PLoS ONE* **2013**, *8*, e61810. [[CrossRef](#)]
51. Tournier, B.; Sanchez-Ballesta, M.T.; Jones, B.; Pesquet, E.; Regad, F.; Latche, A.; Pech, J.C.; Bouzayen, M. New members of the tomato ERF family show specific expression pattern and diverse DNA-binding capacity to the GCC box element. *FEBS Lett.* **2003**, *550*, 149–154. [[CrossRef](#)]
52. Gibbs, D.J.; Conde, J.V.; Berckhan, S.; Prasad, G.; Mendiondo, G.M.; Holdsworth, M.J. Group VII ethylene response factors coordinate oxygen and nitric oxide signal transduction and stress responses in plants. *Plant Physiol.* **2015**, *169*, 23–31. [[CrossRef](#)] [[PubMed](#)]
53. Rashotte, A.M.; Goertzen, L.R. The CRF domain defines cytokinin response factor proteins in plants. *BMC Plant Biol.* **2010**, *10*, 74. [[CrossRef](#)] [[PubMed](#)]
54. Rashotte, A.M.; Mason, M.G.; Hutchison, C.E.; Ferreira, F.J.; Schaller, G.E.; Kieber, J.J. A subset of Arabidopsis AP2 transcription factors mediates cytokinin responses in concert with a two-component pathway. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 11081–11085. [[CrossRef](#)] [[PubMed](#)]
55. Cutcliffe, J.W.; Hellmann, E.; Heyl, A.; Rashotte, A.M. CRFs form protein-protein interactions with each other and with members of the cytokinin signalling pathway in Arabidopsis via the CRF domain. *J. Exp. Bot.* **2011**, *62*, 4995–5002. [[CrossRef](#)] [[PubMed](#)]
56. Dong, C.J.; Liu, J.Y. The Arabidopsis EAR-motif-containing protein RAP2.1 functions as an active transcriptional repressor to keep stress responses under tight control. *BMC Plant Biol.* **2010**, *10*, 47. [[CrossRef](#)] [[PubMed](#)]
57. Kim, J.G.; Stork, W.; Mudgett, M.B. Xanthomonas type III effector XopD desumoylates tomato transcription factor SlERF4 to suppress ethylene responses and promote pathogen growth. *Cell Host Microbe* **2013**, *13*, 143–154. [[CrossRef](#)]
58. Vacic, V.; Uversky, V.N.; Dunker, A.K.; Lonardi, S. Composition profiler: A tool for discovery and visualization of amino acid composition differences. *BMC Bioinform.* **2007**, *8*, 211. [[CrossRef](#)]
59. Huang, F.; Oldfield, C.; Meng, J.; Hsu, W.-L.; Xue, B.; Uversky, V.N.; Romero, P.; Dunker, A.K. Subclassifying disordered proteins by the CH-CDF plot method. *Pac. Symp. Biocomput.* **2012**, *17*, 128–139.
60. Xue, B.; Dunbrack, R.L.; Williams, R.W.; Dunker, A.K.; Uversky, V.N. PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta* **2010**, *1804*, 996–1010. [[CrossRef](#)]

61. Petrovich, A.; Borne, A.; Uversky, V.N.; Xue, B. Identifying similar patterns of structural flexibility in proteins by disorder prediction and dynamic programming. *Int. J. Mol. Sci.* **2015**, *16*, 13829–13849. [[CrossRef](#)]
62. Malhis, N.; Jacobson, M.; Gsponer, J. MoRFchibi SYSTEM: Software tools for the identification of MoRFs in protein sequences. *Nucleic Acids Res.* **2016**, *44*, W488–W493. [[CrossRef](#)] [[PubMed](#)]
63. Gao, J.J.; Thelen, J.J.; Dunker, A.K.; Xu, D. Musite, a tool for global prediction of general and kinase-specific phosphorylation sites. *Mol. Cell. Proteomics* **2010**, *9*, 2586–2600. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).