



HAL
open science

Plant-TFClass: a structural classification for plant transcription factors

Romain Blanc-Mathieu, Renaud Dumas, Laura Turchi, Jérémy Lucas,
François Parcy

► To cite this version:

Romain Blanc-Mathieu, Renaud Dumas, Laura Turchi, Jérémy Lucas, François Parcy. Plant-TFClass: a structural classification for plant transcription factors. *Trends in Plant Science*, 2024, 29 (1), pp.40-51. 10.1016/j.tplants.2023.06.023 . hal-04212079

HAL Id: hal-04212079

<https://hal.science/hal-04212079>

Submitted on 20 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Highlights

Transcription factors (TF) are proteins that bind DNA specifically thanks to a DNA-binding domain (DBD).

TF DBDs fold into various 3D-structures that contact DNA through different elements (alpha helices, beta-strands and loops).

The vast diversity of DBD folds in mammalian TF families has been organized hierarchically, into superclasses and classes, based on their 3D-structure and their mode of interaction with DNA.

Such a classification has been missing for plants whose 56 recognized TF families harbor diverse types of DBD, many of which are absent in mammals.

The accumulated 3D-structures and recent availability of accurate models for plant TF DBDs open up the opportunity to organize plants TF families into a higher rank classification.

[Click here to view linked References](#)

1 **Plant-TFClass: a structural classification for plant transcription factors**

2

3 Romain Blanc-Mathieu¹, Renaud Dumas¹, Laura Turchi¹, Jérémy Lucas¹, François Parcy^{1,*}

4

5 ¹Laboratoire Physiologie Cellulaire et Végétale, Université Grenoble Alpes, CNRS, CEA, INRAE,
6 IRIG-DBSCI-LPCV, 17 avenue des martyrs, F-38054, Grenoble, France

7

8 * Correspondence: francois.parcy@cnrs.fr (F. Parcy)

9

10 ORCID / twitter handles:

11 Romain BLANC-MATHIEU: 0000-0002-9485-6330 / @romain_bm

12 Renaud DUMAS: 0000-0003-1404-3274 / NA

13 Laura TURCHI: 0000-0002-2791-5505 / @turchi_l

14 Jérémy LUCAS: 0000-0002-2252-4732 / NA

15 François PARCY: 0000-0003-2191-500X / @Francois_Parcy

16

17 **Keywords**

18 Transcription factors, classification, plants, 3D-structures, DNA-binding domain, TFClass

19 **Abstract**

20 Transcription factors (TFs) bind DNA at specific sequences to regulate gene expression. This
21 universal process is achieved thanks to their DNA-binding domain (DBD). In mammals, the vast
22 diversity of DBD structural conformations and the way they contact DNA has been used to
23 organize TFs in the TFClass hierarchical classification. However, the numerous DBD types
24 present in plants and absent from mammal genomes were missing from this classification. We
25 reviewed DBD 3D-structures and models available for plant TFs to classify most of the 56
26 recognized plant TF types within the TFClass framework. This extended classification adds
27 **eight** new classes and **37** new families corresponding to DBD structures absent in mammals.
28 Plant-TFClass provides a unique resource for TF comparison across families and organisms.

29 **Towards a universal classification for transcription factors**

30 Transcription factors (TFs) play essential roles in most processes occurring in living organisms.
31 Thanks to a dedicated DNA binding domain (DBD), these proteins bind DNA at specific

32 sequences (called TF binding sites or TFBS) and regulate the expression of associated genes.
33 Plant (Streptophyta: grouping the land plants – Embryophyta - and their closest algal relatives)
34 genomes contain a high diversity of TF-encoding genes including a collection of plant-specific
35 TFs [1–3] that evolved during their history. The initial analysis of *arabidopsis* (*Arabidopsis*
36 *thaliana*) genome identified 1500 TFs corresponding to 28 gene families of which 45% (16
37 families) were not found in mammals [4]. Since then, comparative genomic studies identified
38 additional TFs and putative TF families in plant genomes [2] and various plant TFs
39 classifications have been proposed, based on the nature of the DBD and other conserved
40 protein domains [2,5–7]. While some of these classifications mention existing homology
41 between different families, none of them proposes a hierarchical, higher-rank, organization.
42 With the increasing number of 3D DBD structures being solved or reliably predicted, it now
43 becomes possible to propose a hierarchical classification for plant TFs based on the 3D
44 structure and the DNA binding mode of their DBD. Traditionally, TFs have been classified into
45 superclasses based on broad structural similarities of their DBDs [8]. These superclasses
46 should not be regarded as equivalent to phylogenetic classification since it remains uncertain
47 whether all the diverse DBDs have independently emerged or if some have a shared ancestry.
48 However, they are useful for grouping TFs based on distinct structural features. In mammals,
49 the reference structural classification is TFClass [9,10]. This hierarchical classification starts
50 with nine structurally defined superclasses, further decomposed into classes, families and
51 subfamilies. It provides a framework for meta-analyses of TF properties and evolution [11], it
52 facilitates comparisons between organisms and it is the reference classification used by
53 JASPAR, the open access database for TFBS models [12].
54 Here we built upon the TFClass framework to classify all plant TF types described to date. The
55 new framework is referred to as “Plant-TFClass” (Figure 1, online supplemental information
56 Table S1), illustrated with the DBD folds with or without DNA for each TF family (Figure 2). For
57 this, we reviewed experimental and predicted data on DBD structures. For each TF type, we
58 proposed a classification when sufficient evidence exist and we left the remaining ones in a
59 tenth ‘Yet undefined DNA-binding domains’ superclass as in TFClass. In addition to this manual
60 construction, we also used an automatic approach based on the pairwise structural alignment
61 of TF DBDs: this returned a very similar grouping, but with some inconsistencies that highlight
62 the merits of the manual approach (online supplemental information Figure S1). Compared to
63 TFClass, Plant-TFClass has additional TFs of two natures: 1) TFs with a DBD type absent from

64 mammals. This is the case for **eight** new classes (LEAFY, **RWP-RK**, AP2/EREBP, EIL, **TCP**, **ALOG**,
65 SBP and B3) and we explain on which basis these new classes were organized under a given
66 superclass. Such TFs may have evolved (i) de novo, (ii) from existing DBD by rapid sequence
67 evolution or (iii) from transposon proteins [13,14]. 2) Another type of plant-specific TFs
68 corresponds to those possessing a DBD fold that exists in mammals but has diverged
69 extensively, in particular with the acquisition of auxiliary domains. This usually defines new
70 families within existing classes (e.g. HD-ZIP, MADS Type II, AHL).
71 Among the 56 recognized plants TF types, **50** were classified in one of the nine superclasses
72 from TFClass (Helix-turn-helix domains, Beta-hairpin exposed by an alpha/beta-scaffold, Other
73 all-alpha-helical DNA-binding structures, Beta-sheet binding to DNA, Zinc-coordinating DNA-
74 binding domains, Beta-barrel DNA-binding domains, Basic domains, Alpha-helices exposed by
75 beta-structures, Immunoglobulin-fold) and the remaining **six** were added to the “Yet
76 undefined DBD”.

77 **Helix-turn-helix domains**

78 TFs of the helix-turn-helix (HTH) domains superclass bind specific DNA sequences by inserting
79 a ‘recognition’ alpha-helix into the major groove. This recognition helix is followed by a turn
80 and, in most cases, two additional alpha-helices packed nearly perpendicularly against it. HTH
81 DBDs also often have an additional structure (usually a loop) that projects basic residues in
82 the DNA minor groove, also contributing to the sequence specificity. TFClass divides the **HTH**
83 **domains** superclass into seven classes, five of which are present in plants: **Heat shock factors**,
84 **Fork head / winged helix factors**, **AT-rich interaction domain (ARID)**, **Homeo domain factors**
85 **(HD)** and **Tryptophan cluster factors**.

86 The three first classes are represented by a single family in plants (**HSF**, **E2F** and **ARID**
87 respectively). The ARID is sometimes associated with a HMG-box domain but models suggest
88 that ARID-HMG proteins interact with DNA mainly thanks to their ARID domain [15]. Thus, we
89 classified them within the ARID family.

90 As for the HD TF, their variety in plants comes from the diversity of adjacent domains including
91 a leucine zipper in HD-Zip that affect DNA binding by allowing dimerization [16]. 3D structures
92 of plant HD DBDs are available for WUSCHEL (WOX family) (6RYI) and revealed a canonical HD
93 fold with loop regions connecting the three alpha-helices that are expanded compared to the
94 drosophila Engrailed HD protein [17].

95 The **Tryptophan cluster factors class** contains **MYB** and **MYB-related** TFs families that are
96 represented by few members in **mammals** but greatly expanded in land plants (as R2R3-MYB:
97 MYB genes containing two repeats of the MYB-type HTH domain) to form one of the largest
98 transcription factor family [18]. A structure of plant MYB DBD in complex with DNA is available
99 for the R2R3-MYB WEREWOLF (6KKS) and revealed a similar fold as for **other eukaryotic** MYB
100 [19]. We have added to the **Tryptophan cluster factors class**, three plant families **absent in**
101 **mammals**: the **GARP**, the **Trihelix** and the **Storekeeper** (STK) TFs (also known as GeBP). **GARP**
102 TFs have a DBD **sharing weak sequence similarity with MYB due to common ancestry**. These
103 TFs come in two subfamilies, G2-like and ARR-B, with structures available in complex with DNA
104 for the G2-like TFs PHR1 (6J4R) [20] and LUX (5LXU) [21] and without DNA for the ARR-B TF
105 ARR10 (1IRZ) [22]. The Trihelix motif was initially proposed to be **evolutionary** related to the
106 C-MYB DBD based on **weak sequence similarity** [23] and this hypothesis was later confirmed
107 by a solution structure of arabidopsis GT-1 DBD (2JMW) [24]. **STK** TFs were included in this
108 class based on the AlphaFold2 model of their DBD: this model shows similarity to HTH domains
109 and 3D structure comparisons returns good alignments with members of the Trihelix family,
110 MYB family and **GARP/G2-like** subfamily (with root mean square deviation (rmsd) of 3 Å on 80
111 equivalent positions out of 90 residues). All these families harbor conserved tryptophan
112 residues (three for R2R3-MYB WEREWOLF and STK down to one for GARP) within the HTH
113 domain, justifying their inclusion in this class.

114 Finally, we added to this HTH domain superclass the plant specific **LEAFY** (LFY) class. Crystal
115 structure of LFY DBD in complex with DNA have been solved for **arabidopsis** (2VY1, 2VY2) and
116 *Physcomitrium patens* (4BHK) proteins, revealing the presence of an HTH motif [25]
117 embedded within a 7-alpha-helices fold, plus a loop that adds base contacts in the DNA minor
118 groove. LFY bind DNA as a dimer and possess a capacity to higher order assembly thanks to
119 the presence of a SAM oligomerization domain [26].

120 **We also added RWP-RK or NIN-like Protein (NLP) factors as a new class. These TFs** possess
121 RWP-RK DNA binding domain with a basic region [27] predicted to fold as three short and one
122 long alpha-helix (AlphaFold2 model from Uniprot). They have been shown to bind DNA as
123 dimers [28]. 3D-structures comparison revealed significant alignment with the helix-turn-helix
124 motif **of bacterial DNA repair protein Ada (1U8B) (45 equivalent positions out of 132 residues**
125 **with an rmsd of 2.1 Å).**

126 **Beta-hairpin exposed by an alpha/beta-scaffold**

127 TFs from this superclass have a DBD made of an alpha/beta-structured scaffold with a beta-
128 hairpin inserting into the DNA major groove and acting as the main DNA contacting element.
129 In TFClass, this superclass contains the **GCM** domain factors class represented by a single
130 mammal-specific family (the GCM family) to which we have added five families found in
131 plants: the **NAC**, **VOZ**, **CAMTA**, **WRKY** and **FRS/FRF**. The determination of DBD/DNA
132 crystallographic structures of the **NAC** TF AtANAC019 (3SWP) and the **WRKY** TF AtWRKY4
133 (2LEX) revealed that the DNA-interaction mode is analogous to that of GCM (1ODH) with good
134 superimposition of the beta-strand's amino acids contacting the DNA [29]. We deliberately
135 chose to include WRKY TFs here and no longer classify them as Zn-coordinating because the
136 domain coordinated by Zn ions lies outside their DBD. We included the Vascular Plant One-
137 Zinc-Finger (**VOZ**) and the calmodulin binding transcription activator (**CAMTA**) families
138 because the AlphaFold2 structural predictions of their DBDs align well with NAC DBD (3SWP)
139 (rmsd of 3 Å on 116 equivalent positions out of 171 residues for VOZ and 2.66 Å on 94
140 equivalent positions out of 112 residues for CAMTA). These similarity levels are also found
141 between WRKY (2LEX) and NAC (3SWP) (55 equivalent positions out of 63 residues with an
142 rmsd of 3.06 Å).

143 **FRS/FRF** factors have been linked to WRKY factors based on an iterative PSI-blast search [30].
144 AlphaFold2-predicted structure of the DBD of the FRS/FRF protein FAR1 revealed a fold similar
145 to that of WRKY and 3D structures comparisons return significant alignment against AtWRKY4
146 (2LEX) (60 equivalent positions out of 90 residues with an rmsd of 3.27 Å). FRS/FRF factors
147 were thus added to the **GCM** class.

148

149 We made **one** more addition to this superclass by including the **AP2/EREBP** class
150 corresponding to proteins with the AP2/ERF DBD, **absent in mammals** [14]. The **AP2/EREBP**
151 class was divided into two families: the **AP2** and the **ERF/DREB**. In the **AP2** family, DBD/DNA
152 structures are available for TEM1 (7ET4) and AtERF1 (1GCC). Both structures revealed a beta-
153 hairpin overhanging the major groove and topped by an alpha-helix [31,32]. In the **ERF/DREB**
154 family, the TF/DNA structure of AtERF96 (5WX9) additionally shows an N-terminal alpha-helix
155 contributing to the DNA interaction by entering the minor groove [33].

156 Other all-alpha-helical DNA-binding domains

157 This superclass gathers TFs possessing a DBD exclusively made of alpha-helices. **The structure**
158 **of their DBDs and their mode of DNA-binding are different from other DBD made of alpha-**
159 **helices (HTH domains and basic domains).** It contains the **heteromeric CCAAT-binding factors**
160 (also known as nuclear factor Y or NF-Y) and the **High-mobility group (HMG) domain factors.**

161 The former class is represented by a single family, present in many eukaryotes including
162 plants. **Heteromeric CCAAT-binding factors** bind DNA as trimers made of NF-YB and NF-YC
163 subunits plus either NF-YA or the **viridiplantae**-specific CONSTANS (CO) protein. The NF-YB/C
164 dimer binds to DNA in a non-specific manner acting as a scaffold for the binding of NF-YA or
165 CO which provide DNA sequence specificity. Despite possessing **no high sequence similarity,**
166 NF-YA6 and CO show similar 3D-structures in complex with NF-YB/C and DNA (6R2V, 7CVQ
167 respectively) with one long helix interacting with the NF-YB/C dimer and one short helix
168 entering the DNA minor groove [34,35]. We thus added CO to the **heteromeric CCAAT-binding**
169 **factors** despite that it does not recognize a CCAAT motif.

170 **High-mobility group (HMG) domain factors** represent a broad group of proteins present in all
171 eukaryotes that bind DNA thanks to an HMG-box to remodel chromatin and to regulate gene
172 transcription. **The HMG-box domain forms an L-like structure made of three alpha helices.**
173 **Unlike HTH domains factors, they bind DNA in the minor groove, widening the minor groove**
174 **and bending the DNA at 90 degrees.** In TFClass, this class contains seven families. Plants HMG-
175 box containing proteins are less diversified than **mammals'** and have a TF activity that is either
176 not demonstrated or achieved via another DBD than the HMG-box (AT-hook or ARID domain)
177 [36].

178 We classified the YABBY proteins, **absent in mammals but found in several other eukaryotes,**
179 as a new family within the "**High-mobility group (HMG) domain factors**". YABBY TFs possess
180 two highly conserved domains that have been proposed to contribute to DNA binding : the N-
181 terminal zinc finger domain and the C-terminal YABBY domain [37,38]. **The predicted YABBY**
182 **domain (39 amino acids long) has a helix-loop-helix domain that superimposes well (39**
183 **equivalent positions, rmsd = 2.01 Å) with the two N-terminal helices of the HMG domain of**
184 **the SRY protein (1HRY), with residues of similar nature potentially contacting DNA.** This
185 superimposition does not exclude a direct DNA binding by the zinc finger domain but

186 experiments on the YABBY CRABS CLAW protein showed a more prominent role for the YABBY
187 domain compared to the zinc finger domain for DNA binding [39].

188 To this superclass, we also added a new plant-specific class: **EIL** (standing for Ethylene-
189 insensitive3 (EIN3)-like factors). The NMR solution structure of arabidopsis EIL3 DBD without
190 DNA (1WIJ) revealed a novel fold made of five alpha-helices with candidate DNA-contacting
191 residues that remain to be confirmed [40].

192 **Beta-sheet binding to DNA**

193 This superclass contains the **AT-hook factors** and TATA-binding proteins classes which bind
194 DNA through single extended strands or beta-sheets, respectively, preferentially in its minor
195 groove [10]. AT-hook motifs exist in a wide range of eukaryotic nuclear proteins [41]. In
196 TFClass, the AT-hook factors class contains a single family, the **HMGA**, that bind DNA at AT-
197 rich stretches thanks to a central Arg-Gly-Arg core that enters deep in the minor groove
198 [42,43]. In plants, HMGA proteins contains four AT-hook DNA binding motifs with at least two
199 required for efficient DNA binding [44]. The overall structure of plant and **metazoan** HMGA
200 proteins is thus quite different.

201 To this class, we added the **land plant**-specific AT-hook motif nuclear-localized (AHL)
202 transcription factors, not orthologous to the HMGA and possessing one or two AT-hook
203 motifs. In addition AHL possess a plant and prokaryote conserved (PPC) domain, involved in
204 protein-protein interactions [45,46].

205 **Finally, we also added the land plant-specific TCP factors. These TFs have been successively**
206 **reported as similar to bHLH based on the presence of a putative bHLH motif [47] and to**
207 **Ribbon-Helix-Helix (RHH) based on the crystallographic structure of the DBD of the rice (*Oryza***
208 ***sativa*) PCF6 [48]. However, the recent determination of TCP-DNA structures revealed a new**
209 **DBD fold (7VP2) where homodimers bind to DNA mainly through anti-parallel beta-strands at**
210 **the dimer interface and two flexible loops at the N-terminal side of each monomer [49]. We**
211 **have thus added the TCP class / TCP family to the “beta-sheet binding to DNA” superclass.**

212 **Zinc-coordinating DNA binding domains**

213 DBDs of TFs from this superclass have a fold organized by one or more zinc ions. The nature
214 of the Zn coordination and the resulting DBD fold define nine classes in TFClass [10]. Within
215 this superclass, plants possess two TFs families shared with animals: **C2H2** (from the “**C2H2**

216 Zinc finger factors” class) and the C4-GATA-related (from the “Other C4 zinc finger-type
217 factors” class). In the C2H2 fold, the “zinc finger” is a loop formed between a beta-hairpin and
218 an alpha-helix and the Zn atom is coordinated by two cysteine and two histidine residues.
219 Several Zn fingers can wrap around the major groove of the DNA via interactions of the alpha-
220 helices with the major groove, as shown in figure 2. In plants, the IDD TFs bind DNA thanks to
221 two C2H2 and two C2CH zinc fingers, with the first C2H2 zinc finger being the most critical for
222 specific binding [50]. We thus classified IDD TFs as a family within the C2H2 class. The C4-
223 GATA-related fold is a variant of the C2H2 fold in which the coordination involves four cysteine
224 residues.

225 We added five types of plant TFs within this superclass: the Squamosa promoter Binding
226 Proteins (SBP), the DNA-binding One Zinc Finger (DOF), LATERAL ORGAN BOUNDARIES (LOB)
227 DOMAIN (LBD), the Arabidopsis LSH1 and Oryza G1 (ALOG) and the Growth-Regulating Factors
228 (GRF) proteins. SBP TFs possess two zinc-finger like structures where one zinc ion is
229 coordinated by three cysteine and one histidine residues (or four cysteine residues in some
230 cases) and the second one is coordinated by two cysteine and two histidine residues. The N-
231 terminal Zn-finger like structure is the best candidate to contact DNA according to NMR
232 experiments (performed in the presence or absence of DNA) and DNA docking on the NMR
233 structure (1UL4) [51]. As the SBP DBD is dissimilar to other known zinc-finger structures it
234 defines a new class “C3H(C),C2H2 zinc-fingers like factors” in Plant-TFClass. DOF TFs bind DNA
235 via their highly conserved DOF domain. The DOF domain forms a single zinc-finger motif of
236 C2C2-type and was thus added to the “Other C4 zinc finger-type factors” class. Their
237 sequence-specific binding to DNA is supported by several *in vitro* functional studies [52–54]
238 and DBD/DNA structure modelling studies [55,56]. The DOF domain is also able to form
239 protein-protein interactions with other TFs [52,57–59]. LBD TFs bind DNA thanks to their
240 dimeric and highly conserved LOB DBD. 3D structural analysis of the LOB domain from the
241 wheat Ramosa2 protein (TtRa2LD, 5LY0) strongly supports that a C4 zinc finger motif is
242 responsible for DNA binding while homodimerization is achieved via a C-terminal leucine
243 zipper-like motif [60]. We thus classified LBD TFs as a new family in the existing Other C4 zinc
244 finger-type factors class.

245 With functional roles established in several angiosperms and *Marchantia polymorpha*, the
246 ALOG family of proteins, proposed as a novel plant-specific TF family [61], was recently
247 characterized at the structural level (8P5Q) [62]. These proteins likely originated from a

248 recombinase of a retrotransposon in which a zinc ribbon was inserted [63] in the core
249 tetrahelical structure. The structure of the DBD in complex with DNA revealed a bundle of
250 alpha helices with a HC₃-motif zinc-ribbon contributing to positioning the alpha-helices and
251 the loop that directly contact DNA. As the zinc ribbon has a HC₃ motif, we created a new class
252 (HC₃ zinc ribbon factors) in the zinc-coordinating DNA-binding proteins superclass for the
253 ALOG family.

254 Plants also encode zinc finger proteins containing a three cysteine and one histidine residues
255 (C3H) motif, that defines the C3H zinc finger factors class in TFClass. Plant C3H are unrelated
256 to those found in mammals and the majority of plant C3H motifs bind RNA. The only plant C3H
257 TFs shown to bind DNA are the Growth-Regulating Factors (GRF family) : their putative DNA-
258 binding domain WRC (tryptophan, arginine and cysteine) contains a conserved C3H
259 (CX₉CX₁₀CX₂H) motif [64]. The WRC domain also plays a role in JM28, a negative regulator of
260 immunity in arabidopsis, by providing target specificity for DNA binding of the RBL/ATX1/2-
261 COMPASS complex [65]. Until more evidence is provided for the other groups of plant C3H
262 proteins, only GRFs are considered as a TF, classified within the C3H zinc finger factors class
263 of TFClass.

264 **Beta-barrel DNA-binding domains**

265 TF from this superclass bind DNA thanks to a beta-barrel of variable number of beta-strands.
266 It is represented by a single family in mammals (The Dbp family of the Cold-shock domain
267 factors class). Distant homologs to this family in arabidopsis (CSP1 and CSP3) bind RNA to act
268 as chaperone but have no known DNA binding activities.

269 On the other hand and as noted by Wingender et al. [10], the plant-specific B3 domain TFs
270 have sequence specific DNA binding and belong to this superclass. The B3 domain contains
271 approximately 110 residues with seven beta-strands and two alpha-helices folded into a
272 pseudo-barrel. The residues belonging to the loops between the beta-strands 1-2 and 4-5
273 contact bases in the DNA major groove allowing each B3 family to recognize a specific
274 sequence [66–68]. B3 domains TFs are present in four families identified as LAV (LEAFY
275 COTYLEDON2 (LEC2)-ABSCISIC ACID INSENSITIVE-3 (ABI3)-VAL), ABI/VP1-related protein
276 (RAV), REPRODUCTIVE MERISTEM (REM), and AUXIN RESPONSE FACTOR (ARF) [69]. ARFs and
277 LAVs have a single B3 domain, whereas REM family proteins have from one to eleven B3
278 domains [70]. RAVs have an AP2/ERF DNA binding domain in addition to the B3 domain. Most

279 ARF TFs possess a PB1 oligomerization domain [71–73] and often bind DNA as dimers with
280 preference regarding the orientation and the spacing of ARF binding sites [54,74–76]. In
281 contrast to the other families, studies performed on VRN1 [77] and REM16 [78] belonging to
282 the REM family suggest that these TFs do not show sequence-specific DNA binding.

283 **Basic domains**

284 This superclass includes TFs that bind DNA with a basic alpha-helix usually inserted in the DNA
285 major groove. It contains two classes: the basic leucine zipper (**bZIP**) and basic helix-loop-helix
286 (**bHLH**) TFs that are present in most eukaryotes. Both types are dimeric TFs that pinch DNA
287 between two basic helices. We have divided the **bHLH** class into two families: the classical
288 **bHLH** and the **BES/BZR** plant-specific family. **BES/BZR** TFs slightly differs from classical bHLH.
289 Indeed, instead of having only the long basic helix interacting with DNA as in classical bHLH,
290 in **BES/BZR** DBDs, the loop between the two helices adds contacts with the minor groove and
291 the second helix involved in dimerization is much shorter.

292 **Alpha-helices exposed by beta-structures**

293 Members of this superclass have an all-alpha-helically folded DBD but contrary to the HTH
294 domains, the basic domains and the other all-alpha-helical DNA-binding domains
295 superclasses, the DNA-binding helices are exposed by a scaffold of beta-strands and do not
296 insert in either DNA groove but are packed against the DNA double helix. This superclass
297 contains two classes: the SAND domain factors and the MADS box factors. SAND domain
298 factors are represented by ULTRAPETALA genes in plants and by VARL genes in green algae
299 but a sequence-specific DNA binding activity has never been demonstrated [79,80]. The **MADS**
300 **box factors class** contains two families in plants (**Type I** and **Type II**) with a DBD/DNA complex
301 structure only described in animals. The DBD structure of arabidopsis floral organ identity
302 MADS TF SEPALLATA3 was recently obtained without DNA (7NB0) and did not reveal any
303 major differences with metazoan proteins [81].

304 **Immunoglobulin-fold**

305 In TFClass, this superclass contains 16 families grouped into six classes that have no homologs
306 with demonstrated TF activity in plants.

307 **Yet undefined DNA-binding domains**

308 This superclass gathers TFs with functionally well-characterized DBD but awaiting structural
309 data for a definitive classification. In TFClass, it includes four classes, plus an “uncharacterized”
310 class gathering six families. Among those families, plants possess NFX1-like proteins whose TF
311 activity remains to be firmly established [82].

312 To this superclass, we added nine classes corresponding to plant factor families with well
313 characterized TF activity but lacking crucial structural clues on the DBD/DNA interaction. For
314 three classes (the cysteine-rich polycomb-like protein (**CPP**) [83], the DNA-binding protein
315 phosphatase (**DBP**) [84] and the **S1Fa-like** factors [85]), there is no structural data and
316 AlphaFold2 predictions of the 3D structure for their DBD were uninformative (low confidence
317 model or no hit in 3D-structure comparison). Putative classification is discussed below for **3**
318 classes (**GRAS**, **PLATZ** and **BCP**) based on experimental or modeled structures and was
319 mentioned before for GRF TFs.

320

321 The **GRAS** domain of SCARECROW-LIKE7 from rice was crystallized as a dimer (5HYZ) with each
322 monomer consisting of a core region made of alpha and beta-structures, topped with an
323 alpha-helical cap structure. The two cap regions form a candidate DNA-binding groove
324 containing positives residues and DNA docking in this groove predicted tight protein/DNA
325 interactions that were validated experimentally [86]. Once validated at the structural level,
326 such DNA binding mode would likely define a new superclass.

327

328 The **Archaeplastida**-specific **PLATZ** TFs are zinc-dependent DNA binding proteins. The zinc
329 finger coordination motif has, in its N-terminal part, a consensus signature C-x2-H-x11-C-x2-
330 C-x(4–5)-C-x2-C-x(3–7)-H-x2-H which is different from other characterized zinc-binding motifs.
331 It also has four conserved cysteine residues in its central region [87]. Structural models show
332 various folds (alpha-helices, a beta-hairpin and a beta-sheet). In the absence of experimental
333 structure in complex with DNA or extensive modeling, it is not possible to firmly place this
334 family within the Zinc-coordinating DNA-binding domains superclass.

335

336 The **BBR/BPC** DBD has a conserved WAR/KHGTV motif required for DNA binding, reminiscent
337 of the WRKYCGK consensus of WRKY proteins suggesting that it may bind DNA in similar way
338 [88].

339 **Concluding remarks and future perspectives**

340 TFClass organizes mammalian transcription factors in a hierarchical classification
341 (Superclasses / Classes / Families) based on the 3D-structure of their DNA-binding domain and
342 the way it interacts with DNA. Historically, the determination of plant TF DBD structures have
343 lagged behind that of mammals preventing such classification for plants. The relatively recent
344 resolution of plant DBD-DNA structures along with breakthrough and massive prediction of
345 protein structures now enable classifying plant TF families based on the 3D-structures of their
346 DBD. In this opinion paper, we proposed a classification for plant TFs using the TFClass
347 framework. Most plant TFs were classified: either assigned to existing mammalian classes or
348 families or by defining new families or new classes under existing classes and superclasses,
349 respectively. **Until experimental data exist for all types of DBD in complex with DNA, Plant-
350 TFClass can be subject to updates.**

351 **To our knowledge, this is the first comprehensive hierarchical classification of plant TF**
352 **families. Other plant TF databases have listed TF families and grouped only those with high**
353 **sequence similarity of their DBD. The classification of biological entities is an important step**
354 **in understanding how the diversity of life is organized and, the merit of a particular**
355 **classification scheme depends on the perspective of the user (see “Outstanding questions”).**
356 **Our hierarchical classification is an ideal basis for evolutionary studies. Based on existing**
357 **sequence homology for portions of DBD containing the residues contacting the DNA, it is**
358 **indeed likely that TF families within a given class derive from a common ancestor. The only**
359 **exceptions are families within the “Other C4 zinc finger-type factors” and within the “bHLH”**
360 **classes. This work offers the necessary framework to study in details other evolutionary**
361 **relationships, for example between the different classes of a given superclass. Plant-TFClass**
362 **will also structure meta-analyses of TFs aiming for instance at evaluating whether TF**
363 **properties depend on their structural features.** Finally, it is the first classification that brings
364 mammals and plants TFs together under a same umbrella. As new TFs DBD types are reported
365 in other kingdoms of life, our work is another step towards a universal classification of
366 transcription factors.

367

368 **Supplemental information**

369

370 Supplemental information associated with this article can be found at doi:XXXXXXX'

371 **Acknowledgments**

372 We thank Edgar Wingender, Chloe Zubieta, Xuelei Lai and Gabrielle Tichtinsky for their advices
373 on this work. We are grateful to the Rensing Lab for integrating plant-TFClass within the
374 TAPscan database. We would like to thanks two anonymous referees for their constructive
375 comments on the initial version of this manuscript. This work was supported by the GRAL
376 Labex financed within the University Grenoble Alpes graduate school (Ecoles Universitaires de
377 Recherche) CBH-EUR-GS (ANR-17-EURE-0003), by the French National Research Agency in the
378 framework of the "Investissements d'avenir" program (ANR-15-IDEX-02) to RBM, by the ANR-
379 18-CE12-0014 ChromAuxi project to RD and by the ANR-17-CE20-0014-01 Ubiflor and ANR-
380 21-CE20-0024 Beflore projects to FP, and a PhD Fellowship from CNRS Prime80 to LT.

381 **Author contributions:** FP conceived the study. RBM and RD defined the plant structural
382 classes. All authors performed the classification. RBM, FP and RD wrote the manuscript.

383 **Competing interests**

384 The authors declare no competing interests.

385 **References**

- 386 1. Shiu, S.-H. *et al.* (2005) Transcription Factor Families Have Much Higher Expansion Rates
387 in Plants than in Animals. *Plant Physiol* 139, 18–26
388 2. Wilhelmsson, P.K.I. *et al.* (2017) Comprehensive Genome-Wide Classification Reveals That
389 Many Plant-Specific Transcription Factors Evolved in Streptophyte Algae. *Genome Biology*
390 *and Evolution* 9, 3384–3397
391 3. de Mendoza, A. *et al.* (2013) Transcription factor evolution in eukaryotes and the
392 assembly of the regulatory toolkit in multicellular lineages. *Proceedings of the National*
393 *Academy of Sciences* 110, E4858–E4866
394 4. Riechmann, J.L. *et al.* (2000) Arabidopsis transcription factors: genome-wide comparative
395 analysis among eukaryotes. *Science* 290, 2105–2110
396 5. Mukherjee, K. *et al.* (2009) A Comprehensive Classification and Evolutionary Analysis of
397 Plant Homeobox Genes. *Mol Biol Evol* 26, 2775–2794
398 6. Jin, J. *et al.* (2017) PlantTFDB 4.0: toward a central hub for transcription factors and
399 regulatory interactions in plants. *Nucleic Acids Research* 45, D1040–D1045
400 7. Yilmaz, A. *et al.* (2011) AGRIS: the Arabidopsis Gene Regulatory Information Server, an
401 update. *Nucleic Acids Research* 39, D1118–D1122

- 402 8. Luscombe, N.M. *et al.* (2000) An overview of the structures of protein-DNA complexes.
403 *Genome Biology* 1, reviews001.1
- 404 9. Wingender, E. *et al.* (2018) TFClass: expanding the classification of human transcription
405 factors to their mammalian orthologs. *Nucleic Acids Res* 46, D343–D347
- 406 10. Wingender, E. (2013) Criteria for an updated classification of human transcription
407 factor DNA-binding domains. *J Bioinform Comput Biol* 11, 1340007
- 408 11. Ambrosini, G. *et al.* (2020) Insights gained from a comprehensive all-against-all
409 transcription factor binding motif benchmarking study. *Genome Biology* 21, 114
- 410 12. Castro-Mondragon, J.A. *et al.* (2022) JASPAR 2022: the 9th release of the open-access
411 database of transcription factor binding profiles. *Nucleic Acids Research* 50, D165–D173
- 412 13. de Mendoza, A. and Seb e-Pedr os, A. (2019) Origin and evolution of eukaryotic
413 transcription factors. *Current Opinion in Genetics & Development* 58–59, 25–32
- 414 14. Yamasaki, K. *et al.* (2013) DNA-binding domains of plant-specific transcription factors:
415 structure, function, and evolution. *Trends in Plant Science* 18, 267–276
- 416 15. Roy, A. *et al.* (2016) Deciphering the role of the AT-rich interaction domain and the
417 HMG-box domain of ARID-HMG proteins of *Arabidopsis thaliana*. *Plant Mol Biol* 92, 371–
418 388
- 419 16. B urglin, T.R. and Affolter, M. (2016) Homeodomain proteins: an update. *Chromosoma*
420 125, 497–521
- 421 17. Sloan, J. *et al.* (2020) Structural basis for the complex DNA binding behavior of the
422 plant stem cell regulator WUSCHEL. *Nat Commun* 11, 2223
- 423 18. Du, H. *et al.* (2015) The Evolutionary History of R2R3-MYB Proteins Across 50
424 Eukaryotes: New Insights Into Subfamily Classification and Expansion. *Sci Rep* 5, 11037
- 425 19. Wang, B. *et al.* (2020) Structural insights into target DNA recognition by R2R3-MYB
426 transcription factors. *Nucleic Acids Research* 48, 460–471
- 427 20. Jiang, M. *et al.* (2019) Structural basis for the Target DNA recognition and binding by
428 the MYB domain of phosphate starvation response 1. *FEBS J* 286, 2809–2821
- 429 21. Silva, C.S. *et al.* (2016) The Myb domain of LUX ARRHYTHMO in complex with DNA:
430 expression, purification and crystallization. *Acta Crystallogr F Struct Biol Commun* 72,
431 356–361
- 432 22. Hosoda, K. *et al.* (2002) Molecular Structure of the GARP Family of Plant Myb-Related
433 DNA Binding Motifs of the *Arabidopsis* Response Regulators. *Plant Cell* 14, 2015–2029
- 434 23. Nagano, Y. (2000) Several features of the GT-factor trihelix domain resemble those of
435 the Myb DNA-binding domain. *Plant Physiol* 124, 491–494
- 436 24. Nagata, T. *et al.* (2010) Solution structures of the trihelix DNA-binding domains of the
437 wild-type and a phosphomimetic mutant of *Arabidopsis* GT-1: Mechanism for an increase
438 in DNA-binding affinity through phosphorylation. *Proteins: Structure, Function, and*
439 *Bioinformatics* 78, 3033–3047
- 440 25. Ham es, C. *et al.* (2008) Structural basis for LEAFY floral switch function and similarity
441 with helix-turn-helix proteins. *The EMBO Journal* 27, 2628–2637
- 442 26. Sayou, C. *et al.* (2016) A SAM oligomerization domain shapes the genomic binding
443 landscape of the LEAFY transcription factor. *Nature Communications* 7, 11222
- 444 27. Schauser, L. *et al.* (1999) A plant regulator controlling development of symbiotic root
445 nodules. *Nature* 402, 191–195
- 446 28. Nishida, H. *et al.* (2021) Different DNA-binding specificities of NLP and NIN
447 transcription factors underlie nitrate-induced control of root nodulation. *Plant Cell* 33,
448 2340–2359

- 449 29. Welner, D.H. *et al.* (2012) DNA binding by the plant-specific NAC transcription factors
450 in crystal and solution: a firm link to WRKY and GCM transcription factors. *Biochem J* 444,
451 395–404
- 452 30. Babu, M.M. *et al.* (2006) The natural history of the WRKY–GCM1 zinc fingers and the
453 relationship between transcription factors and transposons. *Nucleic Acids Research* 34,
454 6505–6520
- 455 31. Allen, M.D. *et al.* (1998) A novel mode of DNA recognition by a beta-sheet revealed
456 by the solution structure of the GCC-box binding domain in complex with DNA. *EMBO J*
457 17, 5484–5496
- 458 32. Hu, H. *et al.* (2021) TEM1 combinatorially binds to FLOWERING LOCUS T and recruits
459 a Polycomb factor to repress the floral transition in Arabidopsis. *Proc Natl Acad Sci U S A*
460 118, e2103895118
- 461 33. Chen, C.-Y. *et al.* (2020) Structural insights into Arabidopsis ethylene response factor
462 96 with an extended N-terminal binding to GCC box. *Plant Mol Biol* 104, 483–498
- 463 34. Chaves-Sanjuan, A. *et al.* (2021) Structural determinants for NF-Y subunit
464 organization and NF-Y/DNA association in plants. *The Plant Journal* 105, 49–61
- 465 35. Lv, X. *et al.* (2021) Structural insights into the multivalent binding of the Arabidopsis
466 FLOWERING LOCUS T promoter by the CO–NF–Y master transcription factor complex. *The*
467 *Plant Cell* 33, 1182–1195
- 468 36. Antosch, M. *et al.* (2012) Plant Proteins Containing High Mobility Group Box DNA-
469 Binding Domains Modulate Different Nuclear Processes1[W]. *Plant Physiol* 159, 875–883
- 470 37. Bowman, J.L. and Smyth, D.R. (1999) CRABS CLAW, a gene that regulates carpel and
471 nectary development in Arabidopsis, encodes a novel protein with zinc finger and helix-
472 loop-helix domains. *Development* 126, 2387–2396
- 473 38. Sawa, S. *et al.* (1999) FILAMENTOUS FLOWER, a meristem and organ identity gene of
474 Arabidopsis, encodes a protein with a zinc finger and HMG-related domains. *Genes Dev*
475 13, 1079–1088
- 476 39. Gross, T. *et al.* (2018) CRABS CLAW Acts as a Bifunctional Transcription Factor in
477 Flower Development. *Frontiers in Plant Science* 9
- 478 40. Yamasaki, K. *et al.* (2005) Solution Structure of the Major DNA-binding Domain of
479 Arabidopsis thaliana Ethylene-insensitive3-like3. *Journal of Molecular Biology* 348, 253–
480 264
- 481 41. Aravind, L. and Landsman, D. (1998) AT-hook motifs identified in a wide variety of
482 DNA-binding proteins. *Nucleic Acids Res* 26, 4413–4421
- 483 42. Fonfría-Subirós, E. *et al.* (2012) Crystal structure of a complex of DNA with one AT-
484 hook of HMGA1. *PLoS One* 7, e37120
- 485 43. Huth, J.R. *et al.* (1997) The solution structure of an HMG-I(Y)–DNA complex defines a
486 new architectural minor groove binding motif. *Nat Struct Mol Biol* 4, 657–665
- 487 44. Grasser, K.D. (2003) Chromatin-associated HMGA and HMGB proteins: versatile co-
488 regulators of DNA-dependent processes. *Plant Mol Biol* 53, 281–295
- 489 45. Yun, J. *et al.* (2012) The AT-hook Motif-containing Protein AHL22 Regulates Flowering
490 Initiation by Modifying FLOWERING LOCUS T Chromatin in Arabidopsis. *J Biol Chem* 287,
491 15307–15316
- 492 46. Meijer, A.H. *et al.* (1996) Novel members of a family of AT hook-containing DNA-
493 binding proteins from rice are identified through their in vitro interaction with consensus
494 target sites of plant and animal homeodomain proteins. *Plant Mol Biol* 31, 607–618

- 495 47. Kosugi, S. and Ohashi, Y. (1997) PCF1 and PCF2 specifically bind to cis elements in the
496 rice proliferating cell nuclear antigen gene. *Plant Cell* 9, 1607–1619
- 497 48. Sun, L. *et al.* (2020) The crystal structure of the TCP domain of PCF6 in *Oryza sativa* L.
498 reveals an RHH-like fold. *FEBS Letters* 594, 1296–1306
- 499 49. Zhang, Y. *et al.* (2023) DNA–TCP complex structures reveal a unique recognition
500 mechanism for TCP transcription factor families. *Nucleic Acids Research* 51, 434–448
- 501 50. Hirano, Y. *et al.* (2017) Structure of the SHR-SCR heterodimer bound to the BIRD/IDD
502 transcriptional factor JKD. *Nat Plants* 3, 17010
- 503 51. Yamasaki, K. *et al.* (2004) A novel zinc-binding motif revealed by solution structures
504 of DNA-binding domains of Arabidopsis SBP-family transcription factors. *J Mol Biol* 337,
505 49–63
- 506 52. Yanagisawa, S. (2004) Dof Domain Proteins: Plant-Specific Transcription Factors
507 Associated with Diverse Phenomena Unique to Plants. *Plant and Cell Physiology* 45, 386–
508 391
- 509 53. Umemura, Y. *et al.* (2004) The Dof domain, a zinc finger DNA-binding domain
510 conserved only in higher plants, truly functions as a Cys2/Cys2 Zn finger domain. *The Plant*
511 *Journal* 37, 741–749
- 512 54. O'Malley, R.C. *et al.* (2016) Cistrome and Epicistrome Features Shape the Regulatory
513 DNA Landscape. *Cell* 165, 1280–1292
- 514 55. Hamzeh-Mivehroud, M. *et al.* (2015) Identifying key interactions stabilizing DOF zinc
515 finger–DNA complexes using in silico approaches. *Journal of Theoretical Biology* 382, 150–
516 159
- 517 56. Pandey, B. *et al.* (2018) Dynamics of Dof domain-DNA interaction in wheat: Insights
518 from atomistic simulations and free energy landscape. *Journal of Cellular Biochemistry*
519 119, 8818–8829
- 520 57. Zhang, B. *et al.* (1995) Interactions between distinct types of DNA binding proteins
521 enhance binding to ocs element promoter sequences. *Plant Cell* 7, 2241–2252
- 522 58. Ruta, V. *et al.* (2020) The DOF Transcription Factors in Seed and Seedling
523 Development. *Plants (Basel)* 9, 218
- 524 59. Gao, H. *et al.* (2021) PIF4 enhances DNA binding of CDF2 to co-regulate target gene
525 expression and promote Arabidopsis hypocotyl cell elongation. *Nat. Plants* 8, 1082–1093
- 526 60. Chen, W.-F. *et al.* (2019) Structural analysis reveals a “molecular calipers” mechanism
527 for a LATERAL ORGAN BOUNDARIES DOMAIN transcription factor protein from wheat. *J*
528 *Biol Chem* 294, 142–156
- 529 61. Naramoto, S. *et al.* (2020) The origin and evolution of the ALOG proteins, members of
530 a plant-specific transcription factor family, in land plants. *J Plant Res* 133, 323–329
- 531 62. Rieu, P. *et al.* (2023) The ALOG domain defines a new family of plant-specific
532 Transcription Factors acting during Arabidopsis flower development bioRxiv,
533 2023.06.21.545689
- 534 63. Iyer, L.M. and Aravind, L. (2012) ALOG domains: provenance of plant homeotic and
535 developmental regulators from the DNA-binding domain of a novel class of DIRS1-type
536 retroposons. *Biology Direct* 7, 39
- 537 64. Omidbakhshfard, M.A. *et al.* (2015) Growth-Regulating Factors (GRFs): A Small
538 Transcription Factor Family with Important Functions in Plant Biology. *Molecular Plant* 8,
539 998–1010
- 540 65. Xie, S.-S. *et al.* (2023) JMJ28 guides sequence-specific targeting of ATX1/2-containing
541 COMPASS-like complex in Arabidopsis. *Cell Reports* 42, 112163

- 542 66. Ulmasov, T. *et al.* (1997) ARF1, a transcription factor that binds to auxin response
543 elements. *Science* 276, 1865–1868
- 544 67. Kagaya, Y. *et al.* (1999) RAV1, a novel DNA-binding protein, binds to bipartite
545 recognition sequence through two distinct DNA-binding domains uniquely found in higher
546 plants. *Nucleic Acids Research* 27, 470–478
- 547 68. Mönke, G. *et al.* (2004) Seed-specific transcription factors ABI3 and FUS3: molecular
548 interaction with DNA. *Planta* 219, 158–166
- 549 69. Swaminathan, K. *et al.* (2008) The plant B3 superfamily. *Trends in Plant Science* 13,
550 647–655
- 551 70. Romanel, E.A.C. *et al.* (2009) Evolution of the B3 DNA Binding Superfamily: New
552 Insights into REM Family Gene Diversification. *PLOS ONE* 4, e5791
- 553 71. Korasick, D.A. *et al.* (2014) Molecular basis for AUXIN RESPONSE FACTOR protein
554 interaction and the control of auxin response repression. *Proceedings of the National*
555 *Academy of Sciences* 111, 5427–5432
- 556 72. Nanao, M.H. *et al.* (2014) Structural basis for oligomerization of auxin transcriptional
557 regulators. *Nat Commun* 5, 3617
- 558 73. Han, M. *et al.* (2014) Structural basis for the auxin-induced transcriptional regulation
559 by Aux/IAA17. *Proc Natl Acad Sci U S A* 111, 18613–18618
- 560 74. Boer, D.R. *et al.* (2014) Structural basis for DNA binding specificity by the auxin-
561 dependent ARF transcription factors. *Cell* 156, 577–589
- 562 75. Stigliani, A. *et al.* (2019) Capturing Auxin Response Factors Syntax Using DNA Binding
563 Models. *Molecular Plant* 12, 822–832
- 564 76. Cancé, C. *et al.* (2022) Auxin response factors are keys to the many auxin doors. *New*
565 *Phytologist* 235, 402–419
- 566 77. Levy, Y.Y. *et al.* (2002) Multiple Roles of Arabidopsis VRN1 in Vernalization and
567 Flowering Time Control. *Science* 297, 243–246
- 568 78. Yu, Y. *et al.* (2020) Arabidopsis REM16 acts as a B3 domain transcription factor to
569 promote flowering time via directly binding to the promoters of SOC1 and FT. *The Plant*
570 *Journal* 103, 1386–1398
- 571 79. Carles, C.C. and Fletcher, J.C. (2009) The SAND domain protein ULTRAPETALA1 acts as
572 a trithorax group factor to regulate cell fate in plants. *Genes Dev.* 23, 2723–2728
- 573 80. Duncan, L. *et al.* (2006) Orthologs and paralogs of *regA*, a master cell-type regulatory
574 gene in *Volvox carteri*. *Curr Genet* 50, 61–72
- 575 81. Lai, X. *et al.* (2021) The intervening domain is required for DNA-binding and
576 functional identity of plant MADS transcription factors. *Nature Communications* 12
- 577 82. Lisso, J. *et al.* (2012) NFXL2 modifies cuticle properties in Arabidopsis. *Plant Signal*
578 *Behav* 7, 551–555
- 579 83. Cvitanich, C. *et al.* (2000) CPP1, a DNA-binding protein involved in the expression of a
580 soybean leghemoglobin *c3* gene. *Proc Natl Acad Sci U S A* 97, 8163–8168
- 581 84. Carrasco, J.L. *et al.* (2005) A novel DNA-binding motif, hallmark of a new family of
582 plant transcription factors. *Plant Physiol* 137, 602–606
- 583 85. Santi, L. *et al.* (2003) The GA octodinucleotide repeat binding factor BBR participates
584 in the transcriptional regulation of the homeobox gene *Bkn3*. *Plant J* 34, 813–826
- 585 86. Li, S. *et al.* (2016) Crystal Structure of the GRAS Domain of SCARECROW-LIKE7 in
586 *Oryza sativa*. *The Plant Cell* 28, 1025–1034
- 587 87. Nagano, Y. *et al.* (2001) A novel class of plant-specific zinc-dependent DNA-binding
588 protein that binds to A/T-rich DNA sequences. *Nucleic Acids Res* 29, 4097–4105

589 88. Theune, M.L. *et al.* (2019) Phylogenetic Analyses and GAGA-Motif Binding Studies of
590 BBR/BPC Proteins Lend to Clues in GAGA-Motif Recognition and a Regulatory Role in
591 Brassinosteroid Signaling. *Frontiers in Plant Science* 10
592

593 **Figure legend**

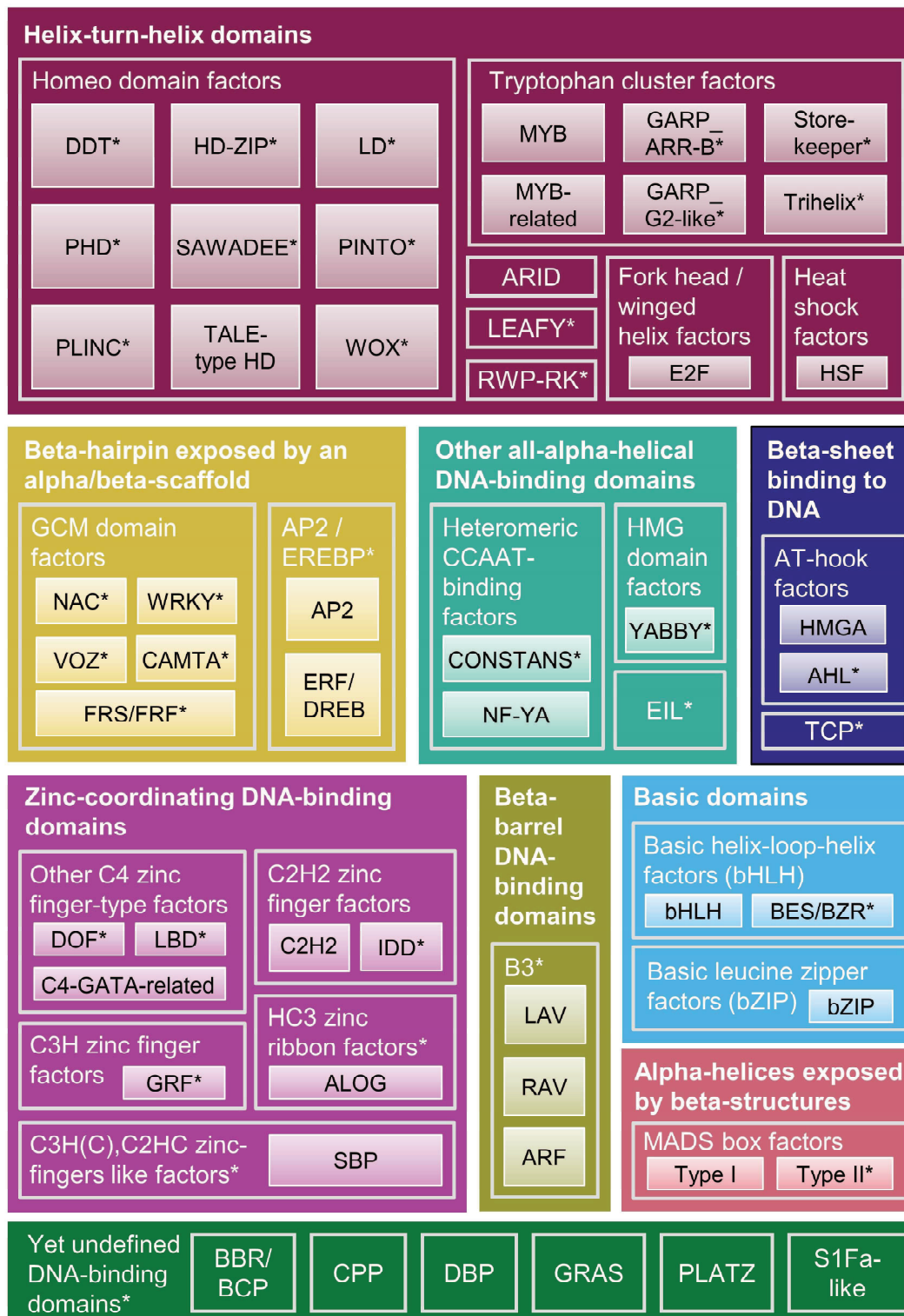
594

595 **Figure 1:** Color map illustrating the Plant-TFClass structural classification. Colors separate the
596 **eight** structurally defined superclasses of transcription factors DNA binding domain **with**
597 **homologs in plants**, plus the “Yet undefined DNA-binding domains”. Gray frame delimits
598 classes, while the families are framed with a gradient-colored background. Asterisks indicate
599 TF families and classes absent from TFClass. Asterisks in classes correspond to DBD fold absent
600 in mammals. TF families within a given class are evolutionary related, at least based on remote
601 sequence homology for a portion of DBD containing the residues contacting the DNA. The only
602 exception are families within the “Other C4 zinc finger-type factors” and within the “bHLH”
603 classes. No evolutionary links are established outside classes.

604

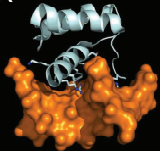
605 **Figure 2:** Illustration of plant TFs DNA binding domains with or without DNA. TFs are classified
606 in their respective superclasses as in figure 1 (frames follow the superclasses color codes from
607 Figure 1). Experimentally determined structures are colored in pale cyan (first monomer) and
608 green cyan (second monomer). The side chains of DNA interacting residues are shown as
609 sticks. DNA is colored orange when it was crystalized with the TF DBD and grey when it comes
610 from another experimentally derived TF/DBD structure used in superimposition (example: the
611 Trihelix superimposed on a MYB crystalized with DNA). When no experimental structure
612 exists, the superimposed AlphaFold2 computed structure is shown in yellow. In
613 superimposition the structure that was used as a template is hidden. For the two heteromeric
614 CCAAT-binding factors, NF-YB and C are colored in deep teal. PyMOL session files are available
615 at <https://github.com/Bioinfo-LPCV-RDF/Plant-TFClass>. For source of structures see online
616 supplemental information Method.

Figure1



Helix-turn-helix domains

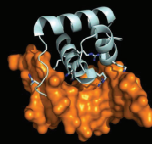
WOX



MYB



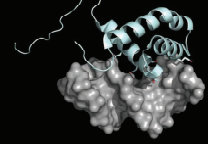
GARP



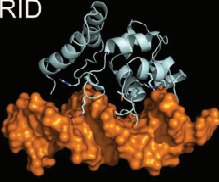
Storekeeper



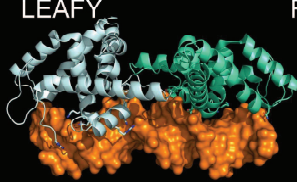
Trihelix



ARID



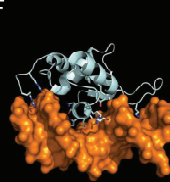
LEAFY



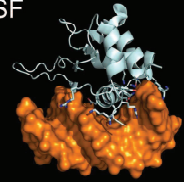
RWP-RK



E2F

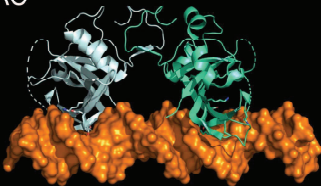


HSF



Beta-hairpin exposed by an alpha/beta-scaffold

NAC



AP2



VOZ



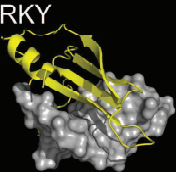
CAMTA



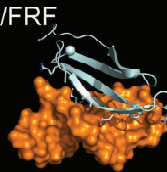
ERF/DREB



WRKY

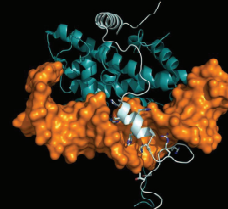


FRS/FRF

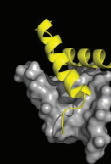


Other all-alpha-helical DNA-binding domains

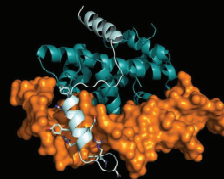
CONSTANS



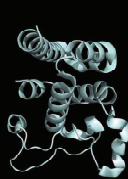
YABBY



NF-YA

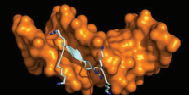


EIL

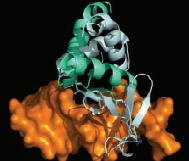


Beta-sheet binding to DNA

AT-hook factors

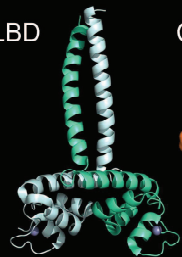


TCP

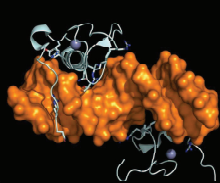


Zinc-coordinating DNA-binding domains

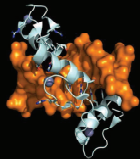
LBD



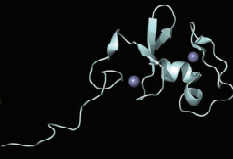
C4-GATA-related



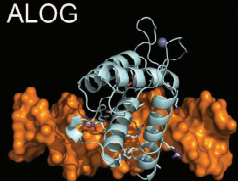
C2H2



SBP



ALOG

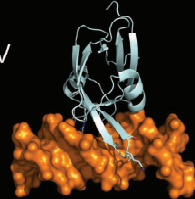


C3H

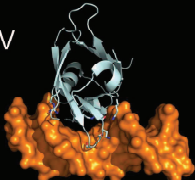


Beta-barrel DNA-binding domains

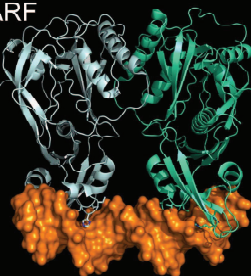
LAV



RAV

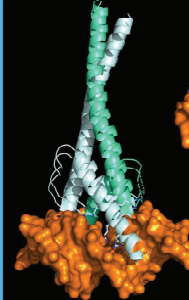


ARF

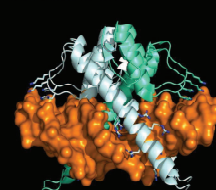


Basic domains

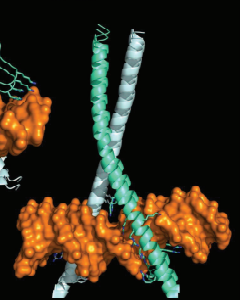
bHLH



BES/BZR



bZIP

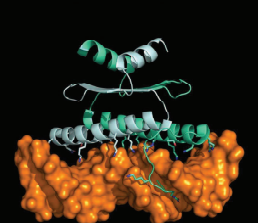


Alpha helices exposed by beta-structures

MADS (Type I)



MADS (Type II)



Supplemental information

Plant-TFClass: a structural classification for plant transcription factors

Romain Blanc-Mathieu¹, Renaud Dumas¹, Laura Turchi¹, Jérémy Lucas, François Parcy^{1,*}

¹Laboratoire Physiologie Cellulaire et Végétale, Université Grenoble Alpes, CNRS, CEA, INRAE, IRIG-DBSCI-LPCV, 17 avenue des martyrs, F-38054, Grenoble, France

* Correspondence: francois.parcy@cnrs.fr (F. Parcy)

Method

Experimentally-determined 3D structures were downloaded from the Protein Data Bank. AlphaFold2 [86] computed structure models were downloaded from UniProt. For each model, position of the DBD was obtained directly from the UniProt “Family and Domains” annotation or identified using CD-search against the CDD database [87]. Structure predictions at the DBD positions were extracted from the PDB files using the Gemmi library (<https://gemmi.readthedocs.io/en/latest/>). Protein 3D structure comparisons and representation were performed using DALI [88], FATCAT [89] and PyMOL. In the main article and for the sake of clarity, we colored names of TF classes and families in blue and orange, respectively.

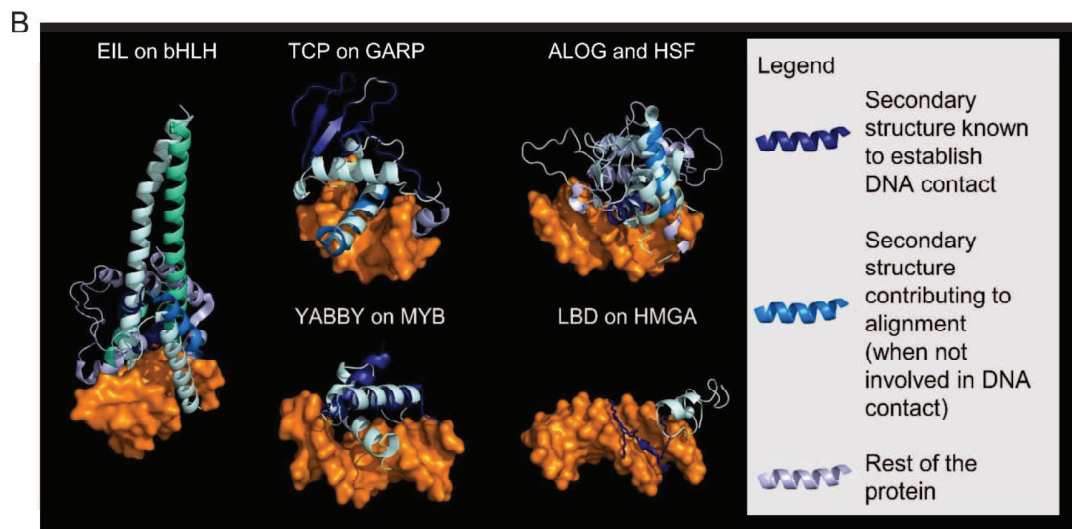
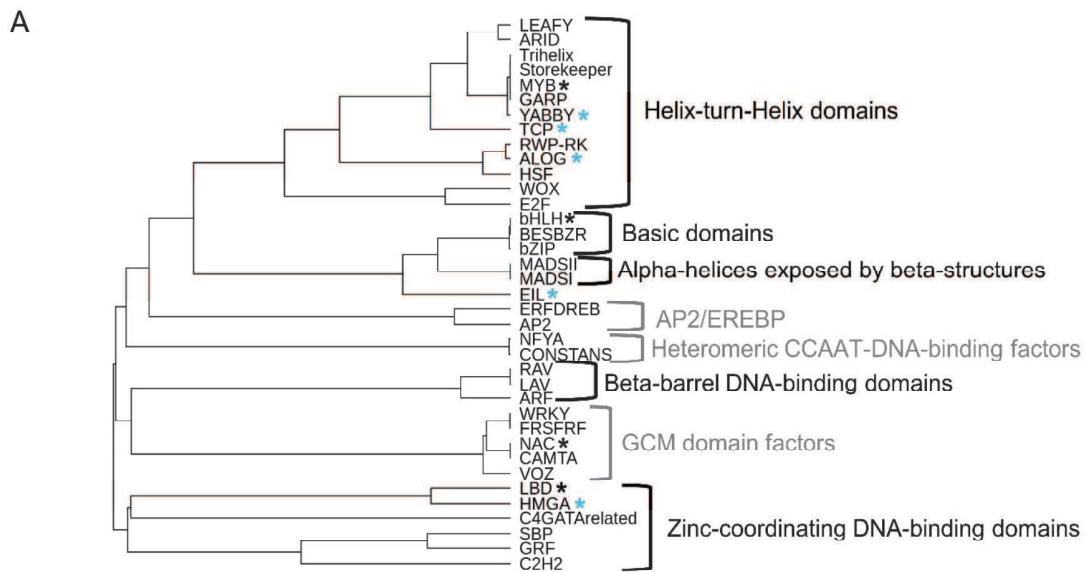


Figure S1: Automatic grouping of transcription factor (TF) families based on similarity of their DNA binding domains (DBDs). A: Dendrogram illustrating the structural distance of DBDs. The dendrogram is generated through hierarchical clustering using P-values from FATCAT structural alignments. Superclasses are labeled in black if they are correctly identified, while split superclass classes are in grey. Blue asterisks indicate placements that are incompatible with our classification in Figure 1. Black asterisks represent PDBs (Protein Data Bank entries) where protein segments not directly involved in DNA binding have been removed. B: Superimposition of DBD structural alignments between families that do not cluster as per our classification (colored as indicated in the legend) and one of the families with the best alignment in A (using the same color code as in Figure 2).

Table S1. List of plant TFs as organized in Plant-TFClass framework

Common names	SuperClass	Class	Family	PDB ID (used in fig 2)
ARID/HMG, ARID/Bright	Helix-turn-helix domains	ARID		1KQQ
E2F, E2F-DP, E2F/DP	Helix-turn-helix domains	Fork head / winged helix factors	E2F	2C6Y
HSF	Helix-turn-helix domains	Heat shock factors	HSF	5D8K
HD-DTT	Helix-turn-helix domains	Homeo domain factors	DDT	NA
HD-Zip	Helix-turn-helix domains	Homeo domain factors	HD-ZIP	NA
LD	Helix-turn-helix domains	Homeo domain factors	LD	NA
HD_PHD	Helix-turn-helix domains	Homeo domain factors	PHD	NA
HD_PINTOX	Helix-turn-helix domains	Homeo domain factors	PINTOX	NA
HD_PLINC	Helix-turn-helix domains	Homeo domain factors	PLINC	NA
HD SAWADEE	Helix-turn-helix domains	Homeo domain factors	SAWADEE	NA
HD TALE, HD_BEL, HD_KNOX	Helix-turn-helix domains	Homeo domain factors	TALE-type HD	NA
HD_WOX	Helix-turn-helix domains	Homeo domain factors	WOX	6RYI
LEAFY, LFY	Helix-turn-helix domains	LEAFY		2VY1
RWP-RK, NIN like, AtRKD, RWPRK	Helix-turn-helix domains	RWP-RK		AF2 model
GARP, ARR-related, ARR-B, G2like, G2-like	Helix-turn-helix domains	Tryptophan cluster factors	GARP_ARR-B	Not shown
GARP, ARR-related, ARR-B, G2like, G2-like	Helix-turn-helix domains	Tryptophan cluster factors	GARP_G2-like	5LXU
MYB	Helix-turn-helix domains	Tryptophan cluster factors	MYB	6KKS
MYB, MYB-related	Helix-turn-helix domains	Tryptophan cluster factors	MYB-related	NA
GeBP, Storekeeper	Helix-turn-helix domains	Tryptophan cluster factors	Storekeeper	AF2 model
Trihelix	Helix-turn-helix domains	Tryptophan cluster factors	Trihelix	2JMW
AP2	Beta-hairpin exposed by an alpha/beta-scaffold	AP2/EREBP	AP2	7ET4
AP2, EREBP, ERF	Beta-hairpin exposed by an alpha/beta-scaffold	AP2/EREBP	ERF/DREB	5WX9
CAMTA, CAMTA-like	Beta-hairpin exposed by an alpha/beta-scaffold	GCM domain factors	CAMTA	AF2 model
FAR1, FHY3	Beta-hairpin exposed by an alpha/beta-scaffold	GCM domain factors	FRS/FRF	AF2 model
NAC, NAC/NAM	Beta-hairpin exposed by an alpha/beta-scaffold	GCM domain factors	NAC	3SWP
VOZ	Beta-hairpin exposed by an alpha/beta-scaffold	GCM domain factors	VOZ	AF2 model
WRKY	Beta-hairpin exposed by an alpha/beta-scaffold	GCM domain factors	WRKY	2LEX
EIL	Other all-alpha-helical DNA-binding domains	EIL		4ZDS
CONSTANS,CO	Other all-alpha-helical DNA-binding domains	Heteromeric CCAAT-binding factors	CONSTANS	7CVQ
CCAAT, HAP3, NF-YB	Other all-alpha-helical DNA-binding domains	Heteromeric CCAAT-binding factors	NF-YA	6R2V
YABBY, C2C2-YABBY, C2C2_YABBY	Other all-alpha-helical DNA-binding domains	High-mobility group (HMG) domain factors	YABBY	AF2 model
A.T hook, AT hook	Beta-sheet binding to DNA	A.T hook factors	AHL	NA

HMGA	Beta-sheet binding to DNA	A.T hook factors	HMGA	3UXW
TCP, bHLH_TCP	Beta-sheet binding to DNA	TCP	TCP	7VP2
C2H2, C2H2 ZF	Zinc-coordinating DNA-binding domains	C2H2 zinc finger factors	C2H2	1A1F
IDD	Zinc-coordinating DNA-binding domains	C2H2 zinc finger factors	IDD	NA
GRF	Zinc-coordinating DNA-binding domains	C3H zinc finger factors	GRF	AF2 model
SBP, SBP-type zinc finger	Zinc-coordinating DNA-binding domains	C3H(C),C2HC zinc fingers-like factors	SBP	1UL5
ALOG	Zinc-coordinating DNA-binding domains	HC3 zinc ribbon factors	ALOG	NA
GATA, Tify	Zinc-coordinating DNA-binding domains	Other C4 zinc finger-type factors	C4-GATA-related	4HC7
DOF, C2C2-Dof, C2C2_Dof	Zinc-coordinating DNA-binding domains	Other C4 zinc finger-type factors	DOF	NA
AS2/LOB, LOBAS2	Zinc-coordinating DNA-binding domains	Other C4 zinc finger-type factors	LBD	5LY0
ARF	Beta-barrel DNA-binding domains	B3	ARF	6YCQ
ABI3VP1, ABI3/VP1, B3	Beta-barrel DNA-binding domains	B3	LAV	6J9C
RAV	Beta-barrel DNA-binding domains	B3	RAV	7RT6
BES/BZR, BES1	Basic domains	Basic helix-loop-helix factors (bHLH)	BES/BZR	5ZD4
bHLH	Basic domains	Basic helix-loop-helix factors (bHLH)	bHLH	5GNJ
bZIP	Basic domains	Basic leucine zipper factors (bZIP)	bZIP	1NWQ
MADS type I, MADS-Box	Alpha-helices exposed by beta-structures	MADS box factors	Type I	1SRS
MADS type II, MADS-Box	Alpha-helices exposed by beta-structures	MADS box factors	Type II	1EGW
BBR/BPC, BBR-BPC	Yet undefined DNA-binding domains	BBR/BPC		NA
CPP Cysteine-rich polycomb-like protein	Yet undefined DNA-binding domains	CPP		NA
DBP	Yet undefined DNA-binding domains	DBP		NA
GRAS	Yet undefined DNA-binding domains	GRAS		NA
PLATZ	Yet undefined DNA-binding domains	PLATZ		NA
S1Fa1, S1Fa-like	Yet undefined DNA-binding domains	S1Fa-like		NA