



Review

Diversity, expansion, and evolutionary novelty of plant DNA-binding transcription factor families☆

Melissa D. Lehti-Shiu ^{a,*}, Nicholas Panchy ^b, Peipei Wang ^a, Sahra Uygun ^b, Shin-Han Shiu ^{a,b,*}^a Department of Plant Biology, East Lansing, MI 48824, USA^b The Genetics Graduate Program, Michigan State University, East Lansing, MI 48824, USA

ARTICLE INFO

Article history:

Received 15 April 2016

Received in revised form 21 July 2016

Accepted 6 August 2016

Available online 10 August 2016

Keywords:

Transcription factor domain families

Evolutionary conservation

Family expansion

Retention mechanisms

Plant novelties

ABSTRACT

Plant transcription factors (TFs) that interact with specific sequences via DNA-binding domains are crucial for regulating transcriptional initiation and are fundamental to plant development and environmental response. In addition, expansion of TF families has allowed functional divergence of duplicate copies, which has contributed to novel, and in some cases adaptive, traits in plants. Thus, TFs are central to the generation of the diverse plant species that we see today. Major plant agronomic traits, including those relevant to domestication, have also frequently arisen through changes in TF coding sequence or expression patterns. Here our goal is to provide an overview of plant TF evolution by first comparing the diversity of DNA-binding domains and the sizes of these domain families in plants and other eukaryotes. Because TFs are among the most highly expanded gene families in plants, the birth and death process of TFs as well as the mechanisms contributing to their retention are discussed. We also provide recent examples of how TFs have contributed to novel traits that are important in plant evolution and in agriculture. This article is part of a Special Issue entitled: Plant Gene Regulatory Mechanisms and Networks, edited by Dr. Erich Grotewold and Dr. Nathan Springer.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

1.1. Overview of transcriptional regulation

Transcriptional regulation of gene expression, i.e. the control of the amount of RNA transcribed from DNA, is complex and requires multiple regulators acting at various stages of transcription [1,2]. The proteins involved in plant transcription have recently been reviewed [3], and here we provide a brief overview of the stages and regulators of transcription. During transcription initiation, binding of RNA polymerase (Pol) requires both general and specific transcription factors as well as chromatin remodeling complexes that modulate DNA accessibility (Fig. 1). In plants, there are multiple RNA polymerases. Protein coding and some noncoding RNA genes are transcribed by RNA Pol II, rRNA is transcribed by RNA Pol I, and tRNAs/5S rRNA is transcribed by Pol III. Pol IV and Pol V, which are derived from Pol II but evolved a specific role in gene silencing [4,5], small interfering and a subset of long non-coding RNAs, respectively [6]. In this review we focus specifically on specific DNA-binding transcription factors that regulate RNA Pol II transcription.

Whether or not a DNA sequence is transcribed first depends on the accessibility of the DNA to the RNA Pol II initiation complex [7]. DNA exists as a complex with histone proteins, called chromatin. The basic unit of chromatin is a nucleosome, which is ~150 bp of DNA sequence wrapped around a histone octamer. The modulation of the spacing of nucleosomes and higher order chromatin structure is an important regulatory mechanism that is required for developmental switches, and many transcriptional regulators exert their effects on transcription via recruitment and interactions with chromatin-associated proteins [8]. Accessibility can be controlled by post-translational modifications of both DNA and histone proteins. For example, transcriptional silencing, which is crucial for the correct progression of developmental transitions [9], is associated with the methylation of DNA and histone 3 at lysine 9 (H3K9) [10]. Additional modifications of histone proteins by histone-modifying enzymes and the composition of histone variants in each octamer also affect chromatin accessibility and the spacing of nucleosomes [11]. Chromatin structure can also be altered by the recruitment of ATP-dependent chromatin remodeling complexes to precise genomic regions by specific DNA-binding transcription factors [12], transcriptional coregulators [13], proteins that bind methylated DNA [14], and noncoding RNAs [15,16]. In addition to local changes in chromatin structure, higher order genome organization is being increasingly recognized as an important mechanism for regulating gene transcription [17].

☆ This article is part of a Special Issue entitled: Plant Gene Regulatory Mechanisms and Networks, edited by Dr. Erich Grotewold and Dr. Nathan Springer.

* Corresponding authors.

E-mail addresses: lehtishi@msu.edu (M.D. Lehti-Shiu), shius@msu.edu (S.-H. Shiu).

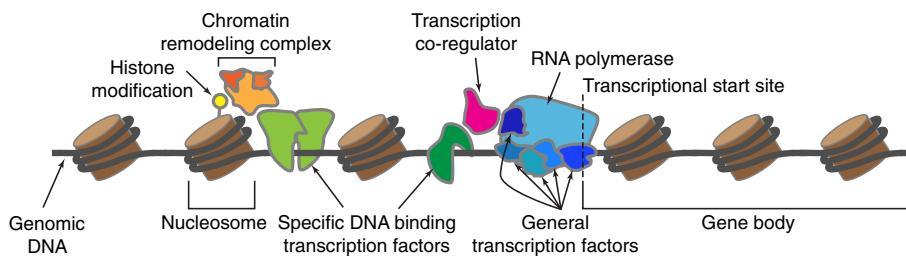


Fig. 1. Regulation of transcription initiation. Transcription initiation begins with the binding of RNA polymerase, a process that requires general transcription factors and, for most genes, specific DNA-binding transcription factors (referred to as TF in this review) that bind upstream of the transcriptional start site and interact with the RNA Polymerase complex either directly or indirectly via transcriptional co-regulators. TFs can also regulate DNA accessibility by recruiting chromatin remodeling complexes that modify chromatin structure by repositioning nucleosomes and changing histone composition. Chromatin structure is also influenced by post translational modifications of histones.

After transcriptional initiation, elongation of the transcript is also subject to regulation, and recent work suggests that in plants initiation and elongation rate may be coordinately regulated by chromatin state [18]. Several plant transcript elongation factors have been identified (for a recent review see [1]) and include proteins that modulate RNA Pol II activity, histone chaperones and ATP-dependent chromatin remodeling complexes that allow the procession of the polymerase through nucleosomes, and histone modification complexes that modulate DNA accessibility. The final step of transcription, transcriptional termination, is tightly coupled with mRNA 3' end processing in eukaryotes [19]. Not much is known about the factors involved in plant transcriptional termination, but there is evidence that the type of terminator impacts the level of transcript produced [20]. Furthermore, as seen in other eukaryotes, alternative polyadenylation signals are present in plants [21]. Use of alternative termination sites can lead to the production of multiple mRNAs with different functions as well as prevent the accumulation of aberrant transcripts [22].

1.2. Specific DNA-binding transcription factors

A large number of proteins and RNAs act as transcriptional regulators in one or more stages of transcription. In this review we focus on the characteristics of specific DNA-binding transcription factors (referred to as TFs from this point on), which are proteins with documented DNA-binding domains (DBDs) that bind to specific, short DNA sequences [23,24]. Note that this definition excludes transcriptional coregulators that do not directly bind DNA and general transcription factors that are associated with Pol II and regulate most genes non-specifically (Fig. 1). By binding to specific target sites, which may be adjacent to or far away from the target gene, TFs regulate gene expression, either positively or negatively, by influencing transcriptional initiation. TFs can enhance or inhibit the recruitment of the RNA polymerase II complex to the core promoter through direct interactions with the complex and/or through indirect interactions with coregulators [25]. TFs can also direct chromatin modifying complexes to specific locations to control the accessibility of DNA to the RNA polymerase II complex [26].

Because TFs regulate the transcription of genes involved in all cellular functions, they significantly impact developmental processes, physiological functions, and ultimately the evolutionary trajectories of plant species. For example, TFs can act as major regulatory network hubs; therefore, changes in TF expression can dramatically affect the expression of a large number of downstream genes [27]. These downstream genes in turn initiate developmental programs that influence morphological features [28]. The crucial roles of TFs in specifying plant morphology are best illustrated by the classical discovery that gains and losses of MADS (MCM1/AGAMOUS/DEFICIENS/SRF)-box TFs lead to homeotic transformations in plants [29]. Because plants are sessile organisms, a robust response system for perceiving and responding to environmental challenges is especially important. TFs play a crucial role in environmental response by regulating stress-responsive

transcription [30–33]. TFs also modulate the transcription of genes underlying other physiological processes in response to environmental cues such as circadian rhythms [34,35], flowering [36], and fruit maturation [37]. Finally, TFs control the transcription of metabolic enzyme-encoding genes, e.g. genes involved in xylem development and secondary cell wall biosynthesis [38,39], and impact plant cellular metabolism.

The influence of TFs on essentially all aspects of plant development and physiology make them central components of a genetic toolkit contributing to plant morphological and developmental evolution [40]. The importance of TFs in plant evolution and the complexity of TF regulatory networks is illustrated by the large number of TFs in sequenced plant genomes, with average numbers ranging from 1500 to 2000 [41–45]. Moreover, these TFs are diverse and can be classified into >50 families based on their characteristic DBDs [44,46–48]. The large number of TFs is due to a higher gene duplication rate in plants in general [49,50] and the higher retention rate of TF duplicates in flowering plants specifically [51–54]. This repeated, on-going process of TF duplication and retention has contributed to new, and in some cases adaptive, functions that have fueled innovation over the course of plant evolution.

An evolutionary understanding of TF diversity and functional divergence is important not only for improving our knowledge of plant TF functions but also for revealing their contributions to plant diversity and adaptation. This knowledge can also be applied to crop species to improve yield and stress tolerance. The evolution of specific TF families has been studied extensively, e.g. APETALA2/Ethylene Response Factor (AP2/ERF) [55,56], MADS-box [57,58], and MYB [59–61]. In addition, there are a number of excellent reviews discussing plant TF evolution (e.g. [62–68]). In this review, our goal is to provide an overview of the evolution of plant TF families with a focus on the following three areas. First, we outline which plant TF families (defined based on DBDs) are shared with other eukaryotic species and which ones are plant-specific. Second, we discuss the evolutionary features of plant TFs, including their expansion history and the factors that have contributed to TF duplication and retention. Finally, we survey examples illustrating how TFs have contributed to evolutionary novelties that impact plant development, environmental response, and domestication.

2. Diversity and conservation of TF families

2.1. Plant TF classification scheme

Plant TFs have previously been classified into families and compared across species [46–48,69–71]. In this review, a TF is defined as a protein with a DBD that binds to specific, short sequences and modifies the expression of genes by interacting with the polymerase machinery and/or affecting the accessibility of DNA. Based on these criteria, the existing collections of TFs cited above, and a recent study comparing eukaryotic transcriptional regulators [44], we classified TF families based on the presence of Pfam [72] protein domains with documented DNA binding

activity associated with transcriptional regulation. Note that this approach is taken to enable comparisons of TF family size across species and not to develop a new classification of TFs. For detailed classification information, we refer the readers to excellent plant TF databases (e.g. [46–48]).

Here we present a general overview of TF families that are: (1) common among multiple eukaryotic lineages, (2) shared between Viridiplantae species (green algae and/or land plants) and one or two major groups of eukaryotes, (3) specific to Viridiplantae, (4) specific to land plants, or (5) conspicuously absent from plants (Fig. 2). The number of proteins in each DBD family was obtained from pre-computed protein domain data in the Pfam database (v.29.0). There are two major challenges in comparing DBD content across species. The first is that genome assembly and annotation quality is expected to vary widely among species. The second challenge is that some DBDs are only present in one or very few species in a major eukaryotic group (e.g. Viridiplantae) and frequently are only represented by one protein. This pattern can be authentic and may have arisen by horizontal gene transfer or, much less likely, parallel losses in multiple lineages. But genomes are frequently contaminated unintentionally or due to the presence of endophytic species. Thus a domain family is only considered to be present in a major group if >20% of the species possess the domain in question. Although this scheme will lead to false negatives, we intend to err on the side of correctly calling the presence of a DBD family in a major lineage than the other way around.

2.2. TF families that are shared between major eukaryotic lineages

Out of 99 TF DBD families, 36 are present in all or most major eukaryotic lineages (referred to as common families, Fig. 2), indicating that these TF families were likely established prior to the divergence of these lineages and may have similar biochemical functions that have been conserved for over a billion years [73]. In eukaryotes, the presence of a nuclear membrane allows cytosolic functions to be separated from genomic activities, including transcriptional regulation. Coincidentally, these common TF families are absent in prokaryotes [44], suggesting that this partitioning may have created an environment that allowed subsequent radiation of TF families. Because early eukaryotes were unicellular organisms, many common TF family members are expected to play regulatory roles in cellular processes fundamental to unicellular life history. Consistent with this expectation, Myb [74], E2F [75] and basic Helix-Loop-Helix (bHLH, [76]) family members regulate cell proliferation, and AT-rich Interaction domain (ARID, [77]), Core-Binding Factor, Beta subunit (CBFB)/Nuclear transcription Factor-Y (NF-Y, [78,79]), and bHLH [80] members are involved in cell cycle control. RWP-RK members are found in the Viridiplantae, red algae, and a few other eukaryotic groups (Fig. 2). Members of the RWP-RK family have a deeply conserved role in regulating the gametophyte to sporophyte transition across land plants [81,82], and an RWP-RK TF is involved in gamete specification in *C. reinhardtii* [83,84].

Another unique feature of eukaryotes is the presence of chromatin and nucleosome structure [85,86]. As expected, multiple common TF families whose members are important for remodeling chromatin structures, including HMG-box [87,88], YL1 [89], ARID [77], CBFB/NF-Y [90,91], and GATA [92] are absent from prokaryotes. Although these common families are well characterized in other model systems, plant members of some common families have yet to be studied experimentally. One example is YL1, which has no published study regarding its function in plants.

Regardless of complexity, both unicellular and multicellular organisms must be able to perceive biotic and abiotic signals from the environment. The regulation of genes in response to communication between individual cells is enabled by common TFs such as Homeobox [93] and MADS-box [28]. In addition, Heat Shock Factor (HSF, [94]) and Zinc finger (Zf, [95]) TFs are important regulatory components of cellular stress response, and bZIP family members are important

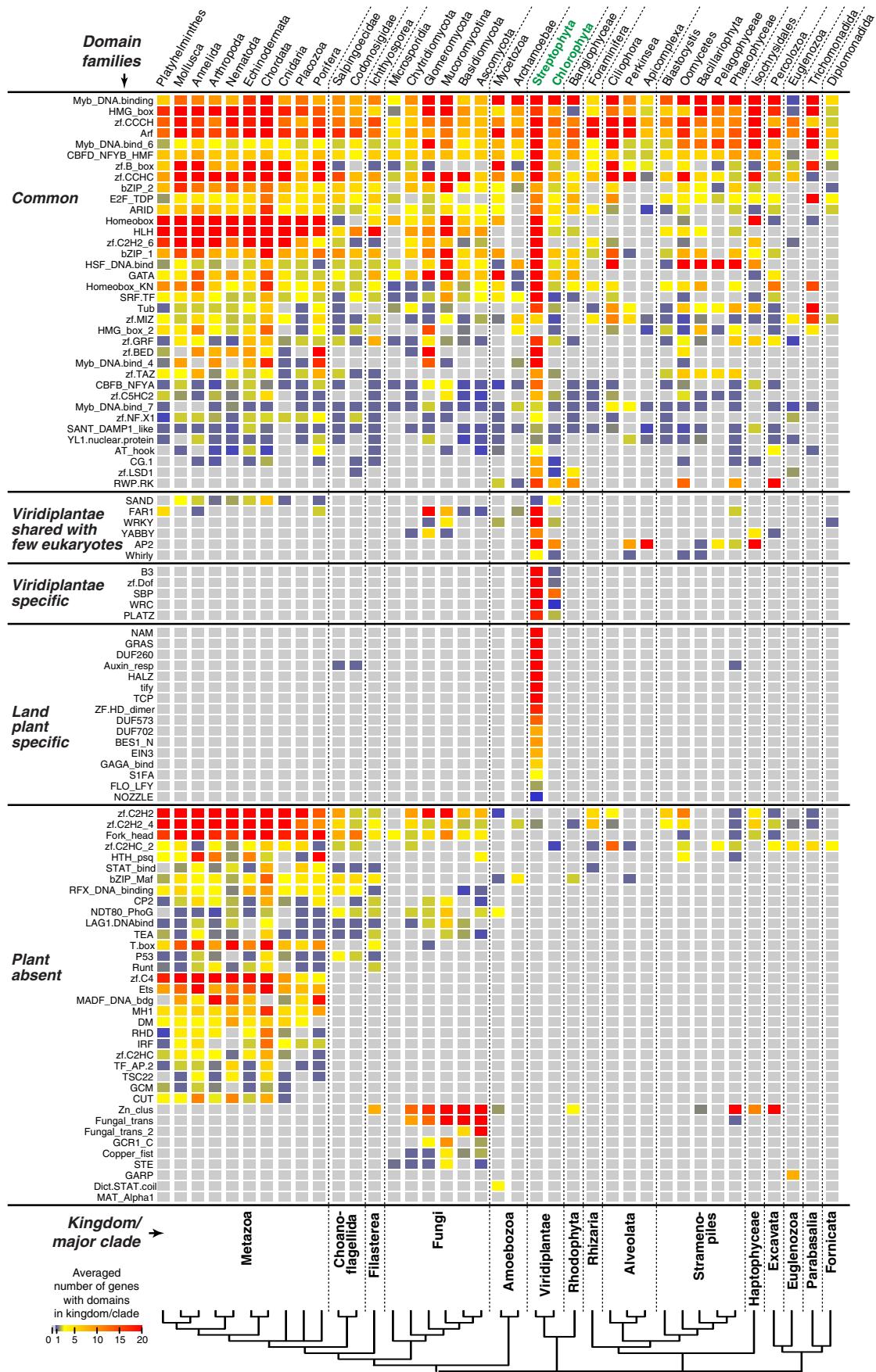
for the endoplasmic reticulum stress response [96]. Although TF families that are found in most eukaryotes are expected to be important for cellular functions, they have often gained novel regulatory roles. For example, cell fate determination, cellular differentiation, and/or multiple stages of multicellular eukaryotic development are regulated by several common TF families including Myb [97], bHLH [76], bZIP [98], ARID [77], and C2H2 Zf [95]. Myb [74] and CBFB/NF-Y [99] also regulate apoptosis and cell senescence. In some cases, common TF families are not involved in the same developmental processes across eukaryotes and instead participate in organism-specific developmental and physiological processes. For example, members of the CG-1 family are involved in sensory mechanisms, embryo development, and growth control in animals but stress tolerance and auxin-dependent development in plants [100]. These common TF families are also involved in many aspects of plant-specific functions that will be discussed in Section 3.

2.3. Viridiplantae TF families shared with a limited number of other eukaryotes

Six plant TF families are shared with only a few other eukaryotic lineages. One example is the SAND (Sp100, AIRE-1, NucP41/75, DEAF-1) domain family that is found in land plants and green algae but in no other eukaryotic lineage except for metazoans (Fig. 2). SAND domain proteins are involved in chromatin-dependent transcriptional regulation [101] and transcription elongation [102]. In plants, the SAND domain-containing protein ULTRAPETALA1 is part of the trithorax group protein complex that removes the repressive Polycomb complex [103]. Plant SAND proteins have diverse roles in regulating inflorescence and floral meristem size [104], floral meristem determinacy [105], gynoecium development [106], and apocarotenoid biosynthesis [107]. As discussed below, like most other lineage limited TF families, the origin of the SAND domain remains unclear.

Other examples of plant TF families shared with few other eukaryotes include YABBY, WRKY, Whirly and AP2/ERF (Fig. 2). Although their functions in plant development and environmental response are well established, with the notable exception of AP2/ERF genes in alveolates [108], it remains unclear what biological processes members of these gene families regulate in other eukaryotes. Plant YABBY family members regulate multiple aspects of plant development [109–112], seed shattering [113], and plant hormone gibberellin biosynthesis [114]. WRKY TFs regulate defense responses against pathogens, leaf senescence, trichome development, and biosynthesis of specialized metabolites [115]. Whirly family members are required for plastid genome stability [116,117], chloroplast biogenesis [118], telomere length homeostasis [119], disease resistance [120], and leaf senescence [121]. Plant AP2/ERF TFs are involved in a myriad of developmental processes, regulation of primary and specialized metabolism, and biotic and abiotic stress responses [122]. In alveolates, which include the malaria parasite *Plasmodium*, AP2/ERF members are involved in processes that are not apparently related to their plant homologs including stage-specific gene activation [123], heterochromatin formation, genome integrity [124] and parasite growth [125].

The fact that these TF families are shared among few non-plant eukaryotes raises interesting questions about their origin and the reason(s) underlying their limited distribution. In some cases, the origin of a TF domain can be hypothesized based on its structural similarity to a protein domain that has a broader taxa distribution. An example is the YABBY domain, which has a characteristic tandem helical structure with an intervening loop. This structure is similar to the bHLH domain [126], which is found in essentially all eukaryotes. Thus, the YABBY domain may be a divergent version of the bHLH domain, but the possibility of convergent evolution cannot be ruled out. In other cases, DBDs are similar to existing domains that do not appear to be involved in DNA binding. The SAND and the WRKY domains belong to the general alpha and beta protein and the small protein classes in the Superfamily



database, respectively. But in both cases it remains unclear which protein domain is ancestral. TF domains may also be completely novel. An example is the Whirly domain, whose quaternary structure has a whirligig appearance that is not found in any other known protein [127].

The reason for the limited distribution of some TF families is also not clear. However, there is speculation that horizontal gene transfer may explain this pattern for some families. The distribution of WRKY is highly restricted, with family members only found in plants, three fungal species, four slime mold species (Amoebozoa), and *Giardia* (Fornicata) (Fig. 2). It has been suggested that this restricted distribution may be due to horizontal gene transfer to non-plants [128]. Like WRKY, the distribution of AP2/ERF in non-plant species is restricted. In addition to a few other eukaryotes, AP2/ERF homologs are present in bacteria, although only in 44 of 1609 bacterial species with pre-computed protein domain data in Pfam. One trivial explanation is contamination of plant materials in the bacterial samples that were sequenced. But this is unlikely because the bacterial species with AP2/ERFs have no apparent affiliation with plants (except *Agrobacterium vitis*, which has two AP2/ERF genes), and no other plant TF is found in these bacteria. Alternatively, mobile DNA elements may have contributed to horizontal transfer of AP2 from bacteria or ciliates to plants [129].

2.4. Plant-specific TF families

There are 21 plant-specific TF families, and of these five families are present in both land plant and green algal lineages (Viridiplantae, Fig. 2) and the rest are present only in land plants (Land plant, Fig. 2). The B3 domain was originally identified in the maize transcriptional activator Viviparous 1 [130]. Several other members of this family have been characterized including, ABA INSENSITIVE 3 (ABI3), LEAFY COTYLEDON2 (LEC2), and FUSCA3, which like Vp1 are involved in seed development, auxin response factors (ARFs), which bind to a DNA element upstream of auxin-responsive genes, and RAV (Related to ABI3/VP1) [131]. RAV TFs are unique because they also contain an AP2 domain and bind to bipartite recognition sequences [132]. ARF TFs also contain a second land plant-specific Auxin_resp domain [133]. The Dof (DNA-binding with one finger) domain is a variant of the C2C2-type zinc finger and is capable of binding to DNA and interacting with other proteins [134–136]. PLATZ (PLAnt AT-rich sequence and Zinc-binding) proteins are a class of zinc-dependent DNA-binding proteins [137]. SBP (SQUAMOSA promoter Binding Protein) TFs were first isolated as regulators of the *SQUAMOSA* gene in *Antirrhinum majus* [138]. The WRC (Trp/Arg/Cys) domain contains a zinc-finger motif as well as a nuclear localization signal and was first studied in rice [139]. Regarding the 16 land-plant specific TF domain families, in some cases the DBD used to define TF families (Fig. 2) differs from the published TF family name. GeBP (GLABROUS1 enhancer-binding protein, [140]) TFs contain a DUF573 domain, HD-ZIP (Homeodomain–leucine zipper, [141]) TFs contain a HALZ (Homeobox associated leucine zipper) domain, LBD (LOB Domain, [142]) TFs contain DUF260 domain, and SRS (SHI RELATED SEQUENCE, [143]) TFs contain a DUF702 domain. The PlantTFDB has a detailed table listing the TF families and their corresponding DBDs [48].

Members of plant-specific TF families have been investigated in land plants via biochemical and genetic approaches. Although the five families shared between land plants and green algae likely regulate similar biological processes, unfortunately the functions of green algal members in four families (B3, PLATZ, WRC, and Dof) have not yet been elucidated. The only family with known functions in both land plants and green algae is SBP. In green algae, the SBP protein Copper Response Regulator 1 is required for regulating copper sensing pathway genes

[144,145]. Similarly, in *A. thaliana*, the SBP-like 7 gene is important for copper deficiency response [144]. Thus, the function of SBP in the regulation of metal ion homeostasis may be ancestral.

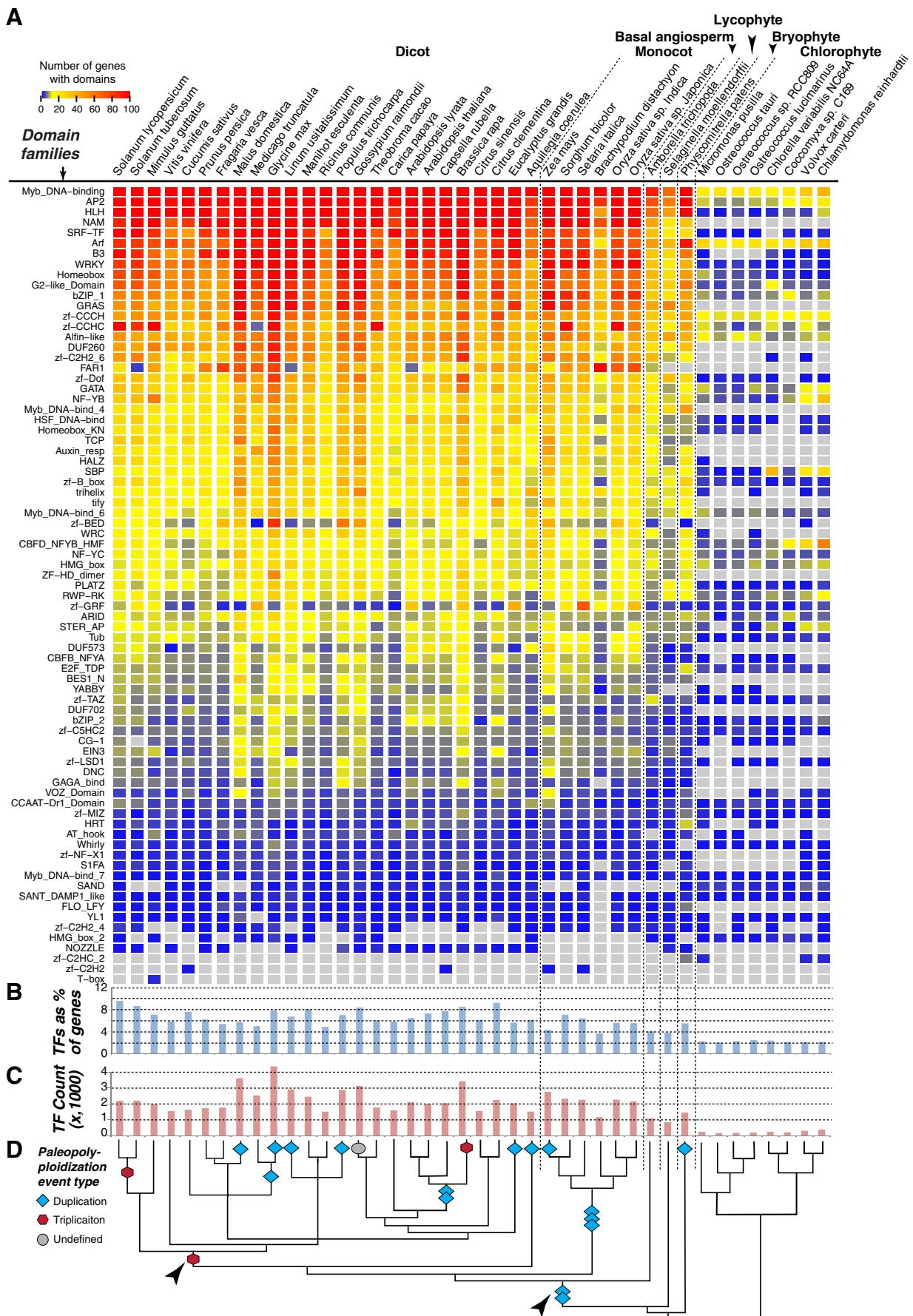
Plant-specific TF families play key roles in the regulation of all stages of plant developmental processes, signaling pathways, stress responses, and biosynthesis of bioactive metabolites that, in some cases, may be crucial for plant adaptation to terrestrial environments [43]. Early developmental processes including seed development, maturation, and germination are regulated by B3, Dof and TCP (TB1 CYC PCF) proteins [146–148]. Lateral organ growth, leaf development and senescence are also controlled by B3, Dof, and TCP as well as LBD, SBP and NAC (NAM, ATAF1,2 CUC2) proteins [146–151]. Maintenance of meristems and regulation of vegetative development are partly controlled by NAC, GRAS (GAI RGA SCR), GeBP, and HD-ZIP proteins [140,141,150,152]. At a later stage, floral development is regulated by members of the B3, SBP, TCP, NAC and LFY (LEAFY) families [146,147,150,153,154], and ovule development is regulated partly by the NZZ/SPL family [155]. In addition to development, B3, Dof, SBP, TCP, NAC, EIN3 (ETHYLENE INSENSITIVE3), GRAS, HD-ZIP, GeBP, and BES1 (BRI1-EMS-Suppressor 1) proteins are important regulators in multiple hormonal signaling pathways, including auxin, abscisic acid, cytokinin, brassinosteroid, phytochrome, gibberellin, salicylic acid, and ethylene [140,141,146–148,150,152,156–158] as well as nodulation [159]. Aside from internal signals such as hormones, SBP, TCP, NAC and GRAS proteins are important for regulating environmental responses to, for example, abiotic stress conditions, pathogens, copper and blue light [146,150,156,160]. Dof and TCP proteins are also involved in biosynthesis of bioactive metabolites, including glucosinolates, brassinosteroids, jasmonic acid and flavonoids [146,148].

How did plant-specific TFs originate? They could have arisen de novo, from existing protein domains without roles in DNA binding, or from other TF families. Although evidence for de novo formation of plant-specific TFs is lacking, there is ample evidence suggesting that some plant-specific TF DBDs originated from homing endonucleases and other TF families. For example, the B3 domain is thought to have originated from restriction endonucleases because the structure of the B3 domain is highly similar to those of the noncatalytic DBDs of EcoRII and Bfil [43]. The NAC domain was likely derived from the WRKY domain during the evolution of land plants [43]. In turn, the WRKY domain may be derived from the $\beta\beta\alpha$ -metal catalytic domain of homing endonucleases (this is also likely the case for C2H2 Zfs, [43]). The SBP domain may have originated from the fusion between a C2H2 Zf and a Zn ribbon, which is a domain found in I-Ppol endonucleases [43]. Plant-specific TFs may also be derived from other bacterial protein families. For example, the GRAS domain may have originally emerged in bacteria and is structurally similar to the Rossmann fold of the methyltransferase superfamily [161]. These examples show that some Viridiplantae and land plant-specific TFs have arisen from existing domains. There are two potential explanations for their absence in other eukaryotes. The first is that some of these plant-specific families may have been established prior to the divergence of major eukaryotic lineages and, over time, evidence for homology has been erased due to sequence divergence. The second explanation is that plants acquired some domains, particularly those related to endonucleases, from prokaryotes through horizontal gene transfer.

2.5. TF families that are absent from plants

Some TF families that are found in one or more eukaryotic lineages are absent from plants (plant absent, Fig. 2). This may be due to loss in the plant lineage, but in some cases the absence in plants is simply due to the fact that homologous DBDs are classified into related domain families that are taxa-specific. One example is the Maf TF family, which

Fig. 2. Average size of each TF domain family in major eukaryotic lineages. The dataset was downloaded from the Pfam database [72], and TF domain family designations are based on three plant TF databases [46–48] and a comparative study [44]. In each cell, the average number of genes containing each TF domain family is shown for each major lineage. Color scale: average number of genes. Numbers greater than 20 are treated as 20. Gray: no gene with the domain in question. The phylogenetic tree of organisms is summarized from multiple studies [128,342–345].



is found mainly in metazoans and Amoebozoa (bZIP_Maf, Fig. 2). Maf TFs contain a variant of the basic leucine ZIPper (bZIP) domain that allows binding to longer DNA sequences [162]. This domain is likely a modified version of an ancestral bZIP domain. Considering that bZIP is present in plants (Fig. 2), bZIP_Maf likely arose specifically in the lineage leading to amoebozoans and metazoans.

In other cases, the domain families that are absent in plants likely arose after the divergence of non-plant and plant lineages via generation of novel proteins or co-option of existing non-TF protein domains in non-plant lineages. For example, the Fork Head (also called Winged Helix) TFs are shared by animal, fungi and other non-plant organisms (Fork_head, Fig. 2). The Fork Head domain may be related to the more ancient catabolite gene activator protein family of helix-turn-helix proteins [163]. In animals, Fork Head TFs are involved in cell fate decisions during embryogenesis [164] as well as the regulation of metabolism, cellular proliferation, stress tolerance and potentially lifespan [165]. In fungi, Fork Head TFs regulate processes relevant to adaptation to the environment, biotic interactions, cell cycle and cellular events [166]. Another example of a TF family not found in plants is the metazoan-specific Ets DBD, which is structurally related to the fork head domain. Ets family members are involved in cell growth, apoptosis, cellular lineage development and differentiation, and tumor malignancy [167]. The Copper Fist family, which is specific to fungi, is involved in copper sensing and utilization [168]. The copper fist is structurally similar to metal binding proteins, metallothioneins [169]. Thus, the Copper Fist is likely derived from co-option of an ancestral metallothionein, which lead to novel functions in transcriptional regulation. There are many other examples of TF families that are absent from plants. The presence of TF families that are specific to certain eukaryotes illustrates that new, lineage-specific TFs have been independently generated, in most cases via unknown mechanisms, over the course of eukaryotic evolution.

3. Pattern of plant TF evolution

3.1. Variation in TF family size among plant species

The number of TFs varies greatly between plant species, from only 151 in *Ostreococcus tauri* to >4000 in *Glycine max* (Fig. 3A, B, C). This variation in TF number is highly correlated with the total number of genes in each species ($R^2 = 0.768$) as well as with the presence of recent paleopolyploidization event(s) in the lineage ($R^2 = 0.611$ when excluding the relatively ancient events indicated by arrowheads, Fig. 3D). Angiosperms, which as a whole have undergone at least two paleopolyploidization events [170], have ~2100 more TFs than non-angiosperm plant species (ANOVA, p -value = 3.30e – 11). Additionally, within angiosperms, there are significantly more TFs (1106 more on average, p -value = 3.52e – 05) in species with lineage-specific whole genome duplication (WGD) events (e.g. the duplication event in the lineage leading to maize). Angiosperms with more ancient WGD events (e.g. two rounds of duplication in the lineage leading to Brassicaceae species) have ~350 more TFs, but this value is not significant (Fig. 3B, C, D). Consistent with earlier findings that WGD events contributed to the expansion of TF families more significantly than to the expansion of most other families [51,52], angiosperm TFs account for significantly more gene content (~3.7%, p -value = 3.11e – 09) than non-angiosperm TFs.

Individual TF families show substantial variation in copy number across species. Some TF families, such as Myb, HLH, and AP2/ERF are found in all major land plant and chlorophyte lineages and have 213,

137, and 165 copies on average in angiosperm genomes, respectively, but only 24, 4, and 11 copies on average in chlorophytes (Fig. 3). Conversely, the E2FTDP and CG-1 families are present in low numbers in all plant species examined (Fig. 3). Considering the lower number of TFs in chlorophyte green algae, the large variation in TF family size across land plant families and species is evidence that differential expansion of TF families has occurred. For example, the MADS-box (SRF-TF domain) family has only one copy in nearly all chlorophyte species examined. Moderate expansion of the MADS-box family took place in early land plant lineages (18 and 24 copies in *Selaginella moellendorffii* and *Physcomitrella patens*, respectively). The average number of MADS-box genes is 98 in angiosperms, suggesting further expansion in vascular plants (Fig. 3). The WRKY family follows a similar pattern [171]. In contrast, the AP2/ERF family has highly expanded in both *A. thaliana* and *P. patens* [171] in comparison to green algae (Fig. 3), indicating that the major expansion of the AP2/ERF family had already taken place prior to the divergence of major land plant lineages ~400–500 million years ago (MYA) [172,173]. Alternatively, the expansion could have been lineage-specific. However this seems unlikely given the similar distribution of genes among AP2/ERF subfamilies in *A. thaliana*, rice, and cucumber [174]. Differential expansion is also evident when comparing more closely related lineages. For example, the NAC family has undergone greater expansion in dicots compared to monocots [175]. On the other hand, the WRKY family tends to be larger among monocots (98 on average) compared to dicots (84 on average). The variation in TF family size among angiosperms is highly correlated with their history of WGD (Fig. 3D). Nonetheless, there exists substantial variation in TF family size that cannot be fully explained by WGD alone. This variation is a consequence of differences in the rate of gene duplication and gene death. In addition, expansion of gene families can also be due to duplicate retention after acquisition of novel functions that confer an evolutionary advantage. In the following sections, the mechanisms that contribute to TF duplication and duplicate retention are discussed.

3.2. Mechanisms of TF duplication

Duplication of genes can occur through a variety of mechanisms, including WGD [176,177], tandem duplication [178], transposon mediated duplication [179], and retroduplication [180,181]. However, these mechanisms do not contribute equally to the duplication of genes as a whole [177,182], and it is expected that they contribute differently to TF duplication as well. WGD is prevalent in plants species [183], and WGD duplicates make up the majority of duplicated genes. As a consequence, in *A. thaliana* for example, there are more than twice as many WGD duplicate genes [177] than there are tandemly duplicated genes [182]. Given that WGD involves the entire genome, all genes have an equal chance of being duplicated. Nonetheless, in *A. thaliana*, TF WGD duplicates tend to be retained at a higher rate compared to other types of genes [51]. The enrichment of TF WGD duplicates has been observed in other species as well, including poplar, rice, soybean, maize, and sorghum [54], and WGD is the underlying cause of ~ 90% of the inferred expansion of TFs over the last 350 MY [184].

In contrast, TFs are not enriched among tandem duplicates in *A. thaliana* [185]. Although tandem duplication does not contribute as prominently to the duplicate gene repertoire in plants, a substantial number of TF duplicates are derived from tandem duplication [186]. For example, 35% of Myb TFs in poplar are tandemly duplicated [187]. Retroduplicates are less common than tandem duplicates [182]. For example, there are only 46 duplicates associated with Pack-MULE

Fig. 3. Distribution and sizes of TF domain families in plant species. (A) The number of genes containing a particular DBD in a plant species. All protein sequences from each species were obtained from Phytozome v.11 and searched against v.30 profile HMMs from Pfam [72], using HMMER v. 3.1 [346] with the “Trusted Cutoff” values as threshold. Additional HMMs for 10 documented plant TF families not available from Pfam were obtained from [70] with an E -value threshold of 1e – 4. Color scale: number of genes. Numbers greater than 100 are treated as 100. Gray: no gene with the domain in question. (B) TF content in a plant genome as the percentage of total genes. (C) TF content in a plant genome as the number of annotated TF genes. (D) Relationships between the plant species examined based on multiple studies [347–350] with paleopolyploidization events indicated. Arrowheads indicate the paleopolyploidization events excluded from the correlation analysis discussed in Section 3.1.

transposons in *A. thaliana* [188], although transposon-related duplicates are more common in other species [179,188,189]. Three partial TF sequences are also found in association with Helitrons in maize [189]. Duplication does not necessarily mean the duplicates will survive and contribute to expansion of families [184,185,190]. Thus, it is interesting that TFs in particular are highly enriched among retained WGD duplicates [51,52]. We next discuss the mechanisms underlying TF duplicate retention and loss.

3.3. TF duplicate retention and loss

Although there are abundant duplicate genes in plant genomes, not all duplicates will be retained. In fact, the predominant fate of duplicate genes is loss [184,185]. Loss of duplicates may include full deletion of the gene or pseudogenization through los-sof-function mutations. Thousands of pseudogenes have been identified in plant genomes [191,192]; however, TFs tend to be underrepresented among pseudogenes compared to other functional classes [191]. This is in spite of the fact that the size of some TF families is large, and gene family size is generally positively correlated with the number of pseudogenes [185]. The lack of TF pseudogenes, could be related to the fact that TFs are not enriched in tandem duplicates, and tandem genes tend to be enriched for pseudogenes [185]. In any case, these findings indicate that retained TF duplicates tend to remain intact, particularly if they were duplicated by WGD.

Because TF duplicates tend not to have pseudogenes, this implies that TFs tend to be either wholly lost or retained after duplication. The retention rate of TF duplicates from the most recent WGD in the *A. thaliana* lineage is 22–25%, which is higher than the genome wide average of 14–16% [51,52]. Additionally, a comparison of orthologous groups between *A. thaliana* and rice reveals that TF families have expanded more in a lineage-specific fashion relative to other functional groups [53] and that not all domains within the same TF family expand at the same rate [193]. What factors may contribute to this increased rate of retention? Efforts in modeling duplication retention have led to the findings that sequence conservation and protein domains/interactions are important in distinguishing retained duplicates from singletons [54,194]. Another important factor in duplicate retention is the function of the gene [51,52,194]. Thus, the high rate of retention for TFs may be largely attributed to their biochemical and biological functions.

TFs are important regulators of gene expression, and a large body of experimental studies has firmly established the relevance of TFs in all aspects of plant life. Given their importance, one may suggest that duplicate TFs are selected to be retained to buffer the effect of losing one copy, with duplicates functioning much like redundant parts in an engineered system [195,196]. In this context, loss of a TF duplicate will not have any fitness consequence. One type of evidence cited to support this notion is the overlapping expression/function of TF duplicates [197,198]. But this overlap could also be due to insufficient time for divergence. Another piece of evidence that duplicate genes buffer deleterious mutations is the finding in *A. thaliana* that phenotypic effects when one copy of a duplicated pair is disrupted are significantly smaller compared to those observed when a singleton gene is disrupted [199]. The issue is that the absence of a phenotype does not conclusively demonstrate the lack of a fitness effect. In addition, it is unclear how future loss of function could be selected for based on present redundancy. It is also unlikely that selection could preserve perfectly identical genes without some level of differentiation [200]. An alternative explanation for the retention of duplicates with seemingly identical functions is that there has been insufficient time for mutations to accumulate. Another possibility is that retention of TF duplicates is selected for.

TF duplicate retention via selection can result when one duplicate obtains a novel adaptive function post-duplication (neofunctionalization, Fig. 4) or because the increased dosage from two copies is itself adaptive [201]. Neofunctionalization of TF duplicates has contributed, for example, to the evolution of floral structures [202–205], vegetative organs [206],

and to the delay of flowering time in cereal crops [207]. Alternatively, duplication can be adaptive by allowing partitioning of conflicting ancestral sub-functions into different duplicates. This process, known as Escape from Adaptive Conflict, allows each sub-function to be independently improved [208,209]. However, the few confirmed examples in plants do not include TFs [209–211].

Duplicates can also be subjected to selection and retained for non-adaptive reasons. TF duplicates can be retained due to subfunctionalization where ancestral sub-functions are partitioned between the duplicates [212]. If all sub-functions are required for optimal fitness, then both copies will be kept. Subfunctionalization has likely occurred in TFs regulating floral development [198,202]. However, a global study of plant TF subfunctionalization is not available. Another possibility is gene (dosage) balance [213–216], which stipulates that simultaneous duplication of genes whose products are interacting are likely to be retained because of the need to maintain the stoichiometry, or balance, between interactors. The involvement of gene balance in TF retention is supported by analyses of protein-protein interaction data from non-plant species and theoretical models [217,218]. It should be noted that, if an ancestral TF has a large number of target genes, its duplicates are also predicted to be retained at a higher rate under the gene balance hypothesis. This notion is supported by a study of grass WGD duplicates where duplicate retention was shown to be influenced by dose-sensitive protein-DNA interactions [219]. One of the main challenges in our understanding of TF evolution is to determine the relative contributions of the different retention mechanisms. This is mainly due to the relative paucity of plant TF-TF and TF-target interaction data compared to other model systems.

3.4. Functional divergence between TF duplicates

Although duplicate TFs may have similar functions (especially immediately following a duplication event), the functions of duplicates are expected to diverge over time. Retained duplicate TFs may have divergent regulatory functions due to differences in expression profiles, binding sites, binding site affinities, and/or their interacting partners. After duplication, plant paralogous genes in general have been shown to diverge in their gene expression levels and patterns [220,221]. Significant divergence in TF duplicate expression has been noted in *A. thaliana* tomato, barley, orchid and poppy [198,202,222–224]. One apparent factor in determining expression divergence is the timing of duplication – more recent duplicates are expected to have more similar expression patterns. Although the correlation between a proxy for duplication time (synonymous substitution rate) and expression similarity is significant, the correlation is exceedingly weak [225], indicating that expression divergence is not simply due to the accumulation of neutral mutations. If expression divergence is random, each duplicate copy should be expressed in a similar number of tissues/conditions. In contrast, plant duplicates, including TFs, are characterized by divergent, asymmetric patterns of expression [220,226], where one duplicate tends to have a significantly higher level and broader range of expression than the other. This time dependent, asymmetric expression divergence can be due to retention of distinct ancestral expression states between duplicates (subfunctionalization) [212], and/or gain of novel functions with a net beneficial effect on fitness (neofunctionalization [201] and/or escape from adaptive conflict [209]).

Although TF duplicates can have diverged expression levels and patterns, the expression patterns of paralogous TFs are in fact more similar than other genes duplicated at a similar time. In *A. thaliana* TF duplicates have a higher degree of expression similarity during root development than other duplicates [227]. This higher than expected expression similarity is what would be expected if the retention of duplicate TFs is due to gene balance [213,215,216]. Gene balance likely explains the expression similarity and the long term retention of *DEFICIENS/APETALA3* (*DEF/AP3*) and *GLOBOSEA/PISTILLATA* (*GLO/PI*) duplicates, which are paralogous and function as floral organ identity B function

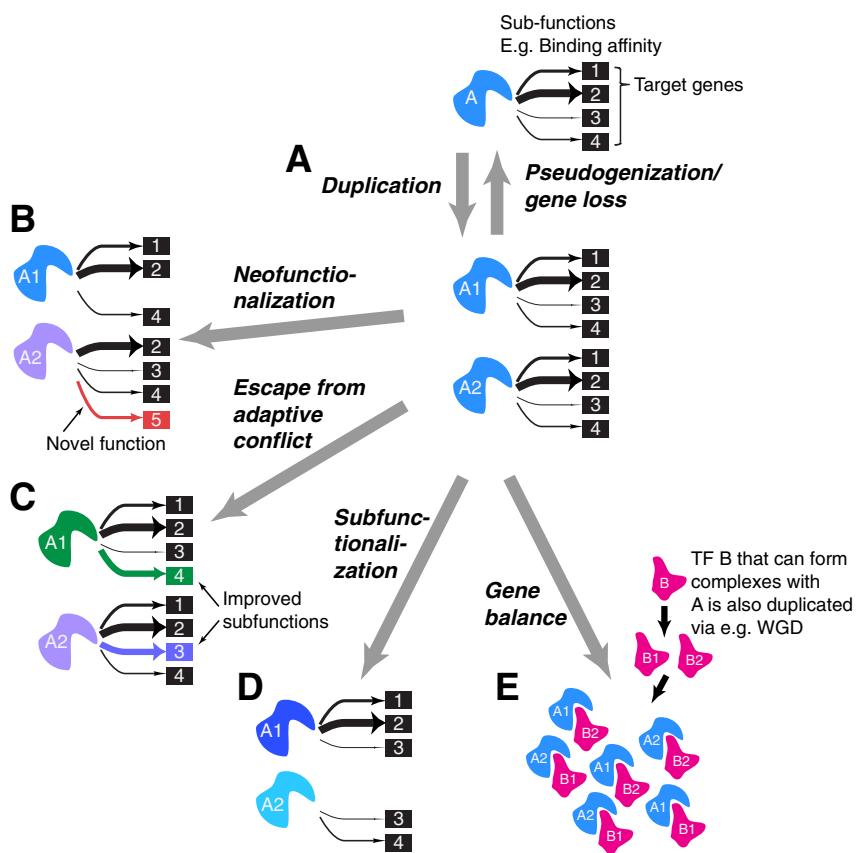


Fig. 4. The fate of duplicated TFs. (A) Following a duplication event, duplicate TFs can return to a single copy state through pseudogenization/gene loss. (B) The duplicates can also be retained by neofunctionalization, where one or both TFs obtain a novel function not present in the ancestor (e.g. a new binding target). (C) Retention due to escape from adaptive conflict. When a TF is not duplicated, improvement of one sub-function conflicts with another sub-function (e.g. increased binding affinity to one gene at the cost of reduced affinity to another). Once duplicated, improvement of either sub-function is feasible in different copies. (D) Subfunctionalization. The differential loss of subfunctions in each TF requires the retention of both copies to carry out the full set of ancestral subfunctions. (E) Gene balance. When TFs form heteromeric complexes, both TF duplicates are required to maintain the stoichiometry. If one copy is lost, the stoichiometric balance is disturbed and thus maladaptive.

genes [228]. Multimerization of AP3-PI is required for nuclear localization [230], and loss of one *GLO1* copy (a *PI* ortholog) would affect the stoichiometry with DEF/GLO complexes. Consistent with this, the severity of homeotic transformation in floral organs is correlated with the number of inactive *GLO1* and *GLO2* alleles in F2 hybrids of mutant petunia [229]. This example highlights the potential importance of gene balance to TF retention. Nonetheless, this example is anecdotal, and we currently do not have a good estimate of the quantitative contribution of gene balance genome wide. Other mechanisms in addition to gene balance likely contribute to TF duplicate retention. For example, although the AP3-PI duplicate pair serves as one of the best examples of gene balance, in species where multiple copies of AP3 and PI exist, preference for interactions between AP3 duplicates and PI duplicates has also been documented [231]. That is, an AP3 duplicate has preferential interaction with one of the PI duplicates. How can this diversification in interaction partners be reconciled with the importance of gene balance? It has been suggested that the apparent preferential interactions between duplicates represents specialization of duplicated complexes [228]. Because each complex consists of a specific AP3 and PI combination, the stoichiometry is maintained. In addition, this specialization can have adaptive consequences because, based on mathematical models, more specialized heterodimers are robust to perturbation [232].

Another mechanism for duplicate TF divergence is a changes in binding site preference that leads to the regulation of different target genes. For example, Pseudo-response Regulators (PPRs), which are important regulators of circadian rhythms, have overlapping but distinct binding sites in *A. thaliana* [233]. These PPRs share <50% of their target sites. TF duplicates can also differ significantly in their binding affinities. For

example, two AP2/ERF TF duplicates, DDF1 and DDF2, that arose from the most recent WGD in the *A. thaliana* lineage ~40–50 MYA, have nearly identical high affinity binding sequences, but one duplicate, DDF1, has much higher binding affinity than its paralog, DDF2 [224]. Although duplicate TFs tend to have different target genes (e.g. [233]), most have highly similar binding motifs [234]. Therefore, the large difference in their target genes can be attributed to affinity differences to the same motifs, divergent expression patterns, and/or other co-regulators. Another consideration is that duplicated genes tend differ greatly in the number and type of *cis*-regulatory elements [235]. This divergence in regulatory interactions, i.e. protein-protein and protein-DNA interaction partners between paralogous TFs, indicates that the selection to maintain duplicate TF expression patterns and regulatory targets to maintain gene balance is perhaps relatively transient. In the long run, other factors, such as neofunctionalization, escape from adaptive conflict, and/or subfunctionalization, subsequently drive divergence in regulatory interactions between duplicate TFs.

4. Roles of TFs in plant novelties and domestication

4.1. Contribution to plant diversity

The expansion of TF families in plants is correlated with increasing complexity and is suggested to be an important mechanism for generating diversity [70,236]. There is a large body of literature discussing the contribution of plant TFs to the evolution of novel traits. Here the goal is to provide a few of the most well-established and recent examples, and a number of excellent studies are unfortunately omitted due to

space limitations. One of the best known examples of the role of TF expansion in generating evolutionary novelty is the link between MADS-box TF expansion and the diversification of flowering plants [237,238]. The development and evolution of a wide range of floral structures can be explained by the “ABCE” model where floral organs are specified by specific combinations of genes: A + E function genes determine sepal identity; A + B + E function genes specify petals; B + C + E function genes specify stamens; C + E function genes determine carpel identity [29,239]. All of these floral organ identity genes except for the A function gene, *APETALA2*, are members of the MADS-box family [240,241]. MADS-box proteins form multimeric complexes composed of tetramers of homodimers and/or heterodimers that bind to regulatory regions of target genes [242]. The expansion of floral MADS-box genes by the WGD event predating the origin of angiosperms, and the subsequent diversification of both expression patterns and interaction partners, is thought to be critical to the origin and evolution of flowers [240,243,244].

4.2. Co-option of existing regulatory functions in the evolution of developmental novelties

One common theme in evolution is the recruitment of existing genes and pathways to generate novel morphologies [245,246]. For example, core networks of regulatory genes may be deeply conserved across land plants, but may be more elaborated in late diverging compared to early diverging lineages. The *ROOT HAIR DEFECTIVE SIX-LIKE* (RSL) genes encode bHLH TFs that regulate root hair development in *A. thaliana*. RSL TFs are also found in *P. patens* where they regulate the differentiation of chloronema cells to caulinema cells in a filamentous structure called the protonema. Both *A. thaliana* and *P. patens* RSL genes are expressed in response to auxin, but the *A. thaliana* RSL regulatory network is more hierarchical, suggesting that the RSL GRN has been remodeled [247]. In *A. thaliana*, RSL TFs positively regulate root hair development by up-regulating *LOTUS JAPONICUS ROOTHAIRLESS1-LIKE3* (LRL3) in response to auxin. In *P. patens* LRL TFs also regulate the development of a root-like structure, the rhizoid, in response to auxin, but independently of RSL TFs. The involvement of *P. patens* RSL and LRL in regulating the development of structures analogous to those in vascular plants via conserved but less complicated regulatory networks supports the idea that conserved core networks of TFs can be modified over time to increase both regulatory and organismal diversity [247,248].

Another example of a developmental regulator that has likely been co-opted to regulate novel but related pathways is *LEAFY COTYLEDON1* (LEC1), a CCAAT box binding factor HAP3 subunit, that specifies embryo identity [249]. The *LEC1* gene is also found in *Selaginella moellendorffii* where it is expressed in both sexual and asexual reproductive structures, but not in moss, suggesting that *LEC1* originated after the divergence of vascular and nonvascular plants but before the evolution of seed plants [250]. The expression pattern in *S. moellendorffii* suggests LEC1 has functions related to seed maturation [250]. In addition, the expression of LEC1-type HAP3 genes in non-seed plants is induced by drought stress [251]. Thus, it is likely that LEC1 was recruited from these ancestral developmental and stress response programs for a novel function in embryogenesis. Interestingly, loss of LEC1 function is associated with the evolution of constitutive asexual reproduction and loss of viable seed production in the *Kalanchoe* or “mother of thousands” genus [252]. Plantlets are produced in the leaf margins due to the activation of embryogenesis programs and organogenesis programs involving the TF *SHOOT MERISTEMLESS* (STM). Mutations that result in truncations of the LEC1 protein have occurred independently in several *Kalanchoe* species, and this likely provides a selective advantage by preventing the initiation of dormancy programs in these plantlets [253].

Recruitment of existing developmental programs can explain the independent origin of traits, such as leaves, over the course of plant evolution [254]. Early diverging vascular plants (lycophytes, such as clubmosses) have microphylls, which are small, simple leaves with a single vascular trace, whereas the later diverging vascular plants, ferns and

seed plants, have megaphylls, which are larger with a more complex vascular pattern. Evidence based on fossil records suggests that microphylls likely have a single origin, whereas megaphylls have several independent origins [255]. Despite having different structures and different origins, the formation of both microphylls and megaphylls involves a transition from an indeterminate apical growth program to a determinate lateral growth program. This transition is mediated by interactions between the Class I *KNOTTED1*-like homeobox (KNOX) TF, which maintains the undifferentiated state of the shoot apical meristem [256–258], and *AS1/ROUGH SHEATH2/PHANTASTICA* (ARP), a member of MYB TF family that negatively regulates the expression of the *KNOX* gene [259–264]. This KNOX-ARP interaction in leaf formation is conserved between eudicots and monocots [263,265], and gene expression and cross-species complementation experiments suggest that microphylls and megaphylls evolved via parallel recruitment of the KNOX-ARP pathway, the original function of which may have been leaf bifurcation [245,254].

Zygomorphy (bilaterally symmetric flowers), is another trait that arose independently several times during the evolution of angiosperms and is associated with pollination syndromes [266] and speciation [267]. The genetic basis of zygomorphy was first uncovered in *Antirrhinum majus* where four TFs control dorsoventral asymmetry. CYCLOIDEA (CYC) and DICHOTOMA (DICH) belong to the TCP family of TFs and control dorsal identity. RADIALIS (RAD) is a MYB TF and promotes ventral identity by restricting the expression of DIVARICATA (DIV), which also encodes MYB protein, to the ventral region [268]. The function of CYC-like genes in bilateral symmetry is conserved in other lineages with zygomorphic flowers, e.g., *Lotus japonicus* [269], pea [270], the legume *Cadia* [271], several species of Malpighiaceae [266] and *Primulina heterotricha* [272], suggesting that independent evolution of zygomorphic flowers was accomplished by parallel recruitment of a CYC-dependent developmental program [273,274].

4.3. Developmental novelties generated through diversification of TF regulatory networks

It is clear that diversification of transcriptional regulatory networks has led to the evolution of novel developmental pathways. TFs in these pathways can diverge via changes in expression pattern and/or protein function. Simple changes in expression patterns can have a dramatic impact on development. For example, variation in leaf structure between species results from differences in leaf margin outgrowth, which is driven by species-specific TF expression patterns [62]. A common mechanism for regulating outgrowth in compound leaves is reactivation of KNOX gene expression during leaf development. The temporal/spatial pattern of KNOX can give rise to many different leaf forms [245]. For example, the simple leaf of *A. thaliana* is due to exclusion of KNOX from the leaves, whereas in the closely related species *Cardamine hirsuta*, different KNOX regulation patterns result in the formation of a compound leaf [275]. KNOX function in compound leaf formation is conserved in some other species such as tomato [276], and *Medicago truncatula* [277]. KNOX function also appears to be conserved in moss and the fern *Ceratopteris richardii*. In these species KNOX genes are expressed in the sporophyte, and when over-expressed in *A. thaliana*, the resulting phenotypes resemble those of KNOX overexpression [278,279].

Differences in fruit development have arisen through variation in expression pattern and/or functions of AGAMOUS-like MADS-box TFs and their interactors [280]. For example, in *Medicago*, small changes in the coding sequence of the MADS-box TF, SHATTERPROOF, lead to a change in protein interactions. This change is correlated with a novel fruit morphology and seed dispersal strategy [281]. FRUITFULL (FUL) was identified in *A. thaliana* as a regulator of cell differentiation during fruit and leaf development [282]. *FUL* and its paralog *APETALA1* (AP1) are derived from a duplication event that is correlated with the origin of the core eudicots [283]. Whereas the pre-duplication *FUL* genes in *Aquilegia* are involved in leaf morphogenesis [284], the *AP1* and *FUL* duplicates have diverged in function with *AP1* specifying sepals and petals and *FUL*

regulating cauline leaf and fruit development [285]. In the case of AtAP1 and AtFUL, divergence is due to both changes in expression pattern and protein sequence [285]. In soybean, a recent change in the spatiotemporal expression of the *API/FUL* ortholog *Dt2* underlies a change in meristem determinacy. *Dt2* represses the expression of the ortholog of *TERMINAL FLOWER1*, *Dt1*, in the meristem. This disrupts the function of *Dt1* in specifying indeterminate growth in the meristem and leads to the semi-determinate growth pattern seen in soybean [286].

4.4. TFs and physiological innovations

The transition of plants from water to land required not only the evolution of developmental novelties, but also physiological innovations, such as the production of novel metabolites. The timing of the transition to land is correlated with the expansion of several TF families including Myb, bZIP, and bHLH as well as plant-specific TF families such as Dof, NAC, AP2/ERF, and WRKY [287], highlighting their potential importance in the adaptation of plants to non-aquatic environments. One physiological innovation considered crucial to the evolution of land plants is the biosynthesis of lignin, a polymer that provides stiffness, which allowed plants to stand upright on land [288]. The xylem cells in the vasculature have thick, rigid secondary cell walls, of which lignin is a major component. Members of a subfamily of NAC domain TFs, VND/NST/SND (VNS), are components of the regulatory network regulating secondary cell wall formation [289]. Consistent with the importance of this family in the evolution of secondary walls and plant adaptation to land, VNS genes are found in non-vascular plants where they are involved in water-conducting cell formation [290], and the VNS family has expanded and diversified in vascular plants, acquiring cell-type specific functions [289]. VNS genes act as master regulators of secondary cell wall formation. Targets of VNS TFs include the secondary regulators, MYB46 and MYB83, which in turn activate additional MYB TFs to more specifically regulate the synthesis of secondary cell wall components, such as lignin [289]. Proteins related to MYB46 have been identified in several vascular plants, suggesting that this regulatory network is conserved, and, given the large number of MYB genes in plants, has diversified [289].

Another primary metabolic pathway that has evolved independently in plants is fatty acid metabolism. Although both plants and animals produce lipids for energy storage, signaling, and membrane biogenesis, fatty acid metabolism in plants is distinct from animals and fungi. For example, fatty acid biosynthesis occurs in the chloroplasts instead of the cytosol, and the fatty acid synthase in plants is more similar to the *E. coli* enzyme [291]. Plant-specific TFs are important regulators of lipid biosynthesis. Overexpression of *Chlamydomonas reinhardtii* Dof leads to an increase in lipids [71]. Similarly, overexpression of soybean Dof4 and Dof11 in *A. thaliana* leads to increased fatty acids as well as total lipids in seeds [292]. The AP2/ERF TF, WRINKLED1 (WRI1), is a central regulator of fatty acid synthesis in *A. thaliana* [293], *Brassica napus* [294], maize [295,296], and oil palm [297]. WRI1 functions in the seed to activate fatty acid biosynthesis for triacylglycerol production, but along with the related TFs, WRI3 and WRI4, targets different enzymes in floral organs to regulate cutin biosynthesis [298]. Based on expression patterns, the *WRI1*, *WRI2* and *WRI3* orthologs in the basal angiosperm *Persea americana* (avocado) also appear to function in fatty acid biosynthesis, indicating a role for WRI genes in lipid metabolism prior to the divergence of monocots from dicots [299].

In addition to primary metabolites, plants synthesize a variety of compounds that are important for defense against herbivores and pathogens as well as for attracting symbionts [300]. There are estimated to be 200,000 specialized metabolites; however, this number may be an underestimate [300]. Glucosinolates, which are important in plant defense, are among the most extensively studied specialized metabolites [301]. Based on the presence/absence of biosynthetic enzymes, papaya only has the capability to produce aromatic glucosinolates [302], but a butterfly-plant arms race lead to the diversification of glucosinolates in the Brassicales [303]. Several studies have identified MYB TFs as

important regulators of glucosinolate biosynthesis [304–307]. In *A. thaliana*, the closely related MYB115 and MYB118 TFs, contributed to the evolution of a novel benzyloxy glucosinolate pathway [308]. The fact that benzyloxy glucosinolates are found in *A. thaliana* but not close relatives suggests that this pathway is newly arisen and that the TFs that regulate this pathway might also be specific to *A. thaliana*. However, MYB115 and MYB118 are conserved and were previously known to regulate embryogenesis, indicating that the regulation of novel genes by conserved TFs can lead to evolutionary novelty [308].

The regulation of the timing of life history traits such as germination and reproduction based on environmental cues such as temperature and day length is another physiological trait regulated by TFs that has been important in plant adaptation. For example, FLOWERING LOCUS C (FLC), a MADS-box TF, is a central regulator of flowering time and prevents early flowering by repressing floral identity genes. An extended cold treatment is required to reduce FLC levels, ensuring that plants overwinter before flowering [309]. FLC arose before the divergence of monocots and dicots [310] and also contributes to natural variation in flowering time [311,312] and seed germination [313] in *A. thaliana*. Another important regulator of flowering, CONSTANS (CO), triggers flowering in response to photoperiod [314]. Recent changes in the cis-regulatory sequence of CO also contribute to natural variation among *A. thaliana* accessions [315]. The role of CO as photoperiodic regulator of flowering evolved in the Brassicaceae after tandem duplication and likely contributed to range expansion [316]. CO-like genes with similar expression patterns and photoperiodic functions are found in potato, rice, and barley, but these functions likely evolved independently [316].

4.5. TFs and plant stress tolerance

TFs play a central role in the response to biotic stresses, such as insects and pathogens, and to abiotic stresses, such as drought, soil salinity, and extreme temperatures, that adversely affect plant growth and development [32]. TFs involved in stress response tend to belong to ancient families and also have lower binding specificity than TFs involved in development [133]. One of the main TF families involved in stress response is the AP2/ERF family [56], which has just 14 genes in *C. reinhardtii* but over 100 in dicots and monocots [317]. Of the two main groups within this family, ERF and AP2, the ERF group has undergone the most expansion in both monocots and dicots and is implicated in both biotic and abiotic stress responses [318,319]. The ERF group can be further divided into the ERF and C-repeat binding factor (CBF)/dehydration responsive element binding protein (DREB) subfamilies based on phylogenetic relationships [223]. In both monocots and dicots, ERF genes regulate responses to both abiotic and biotic stresses, whereas DREB genes primarily regulate responses to abiotic stress [318,319], although this distinction is not absolute [320]. Diversification of the ERF and DREB subfamilies occurred prior to the monocot-dicot split and may have contributed to the evolution of adaptive environmental responses [319,321].

Both ERF and DREB genes have contributed to the evolution of stress tolerance. The ERF gene *Sub1A* allows rice plants to survive hypoxia during flooding by inhibiting gibberellin-mediated elongation and carbon consumption [322,323]. The *Sub1A-1* allele is found only in submergence-tolerant rice cultivars and can confer submergence tolerance to intolerant cultivars indicating that it is a major determinant of submergence tolerance in rice [323]. Three DREB TFs, CBF1, CBF2, and CBF3, regulate cold responsive gene expression and have an important role in cold tolerance in *A. thaliana* [324]. Loss of CBF2 expression likely causes the low freezing tolerance of the *A. thaliana* Cvi accession [325]. CBF2 also maps to QTL for fitness [326] and freezing tolerance [327] in a recombinant inbred line population developed from locally adapted Swedish and Italian *A. thaliana* ecotypes. Less cold-responsive gene expression and lower cold tolerance in the Italian ecotype is due to a non-functional CBF2 protein, providing evidence that CBF2 has been important for local adaptation of *A. thaliana* [328]. Cold-responsive CBF genes

have been identified in several monocot and dicot species and are candidate genes for cold tolerance QTL in wheat [329] and barley [330], suggesting that CBF function is conserved across flowering plants [56].

4.6. Contribution of TFs to crop domestication

Similar to what has been observed for morphological evolution, targets of artificial selection during domestication tend to be transcriptional regulators [331,332]. In addition, TFs that occupy nodal positions in regulatory pathways (i.e. control downstream gene regulation in response to upstream signals) may be more likely to be targets of convergent evolution [333]. The *teosinte branched1* (*tb1*) locus in maize is a striking example of how a change in a single TF can have a large effect on a domestication trait. Compared to teosinte, maize has several improved morphological traits contributing to higher yield and suitability for human diet, such as an increase in apical dominance and naked grains. *Tb1*, a member of TCP transcription factor family with homology to the CYC gene of *A. majus*, represses the growth of axillary organs and promotes the formation of female inflorescences in maize [334]. The higher expression level of *Tb1* in maize is due to the upstream insertion of a *Hopscotch* transposable element, which serves as an enhancer [335]. There is evidence that *tb1* may also have been targeted during artificial selection of pearl millet [336]. Another TF that was targeted during the selection for more accessible grain during maize domestication was the *teosinte glume architecture1* (*tga1*) locus. *Tga1* encodes a SBP-box transcriptional regulator and is largely responsible for the exposed grain phenotype in maize [337,338]. In teosinte, *tga1* promotes longer inflorescence internodes, shorter rachillae, and longer, thicker and harder glumes [339]. A single fixed nucleotide difference that results in a Lys-to-Asn substitution between teosinte and maize, transforms maize *tga1* into a transcriptional repressor and likely underlies the origin of naked kernels [337].

The transcriptional regulation of grain shattering is another common target of artificial selection. In rice, a single amino acid substitution in the MYB3 DBD of Shattering4 dramatically reduces shattering, but does not eliminate it [340]. The *qSH1* gene, which encodes a BEL1-type homeobox TF, also regulates seed shattering in rice. A single nucleotide polymorphism in the promoter of *qSH1* eliminates its expression in the abscission zone, leading to the loss of an abscission zone and therefore loss of shattering [341]. In sorghum, seed shattering is controlled by a single gene, *Shattering1* (*Sh1*), which encodes a YABBY TF. Loss of shattering was selected independently three different times during sorghum domestication through mutations in the regulatory and coding regions of *Sh1*. Interestingly, the *Sh1* orthologs in maize and rice were also under parallel selection during domestication [113]. These examples illustrate the important role of TFs in crop domestication and show that simple changes in the coding sequence and regulatory regions of TFs can have a dramatic impact on plant morphology.

5. Future perspectives

TFs are among the best studied groups of genes in plants, and there is extensive literature detailing their loss- and gain-of-function phenotypes, biochemical interactions with proteins and DNA, patterns of expression, phylogenetic relationships, and evolutionary properties. The accumulation of plant genome sequences, experimental data for individual TFs, and functional genomic data provides a tremendous opportunity to integrate comparative and functional information to better our understanding of the evolutionary processes underlying the diversity of plant TFs. There are a number of outstanding questions. When were extant TF duplicates established, and how long have they persisted? Although the expansion history of individual TF families over the course of land plant evolution has been examined, there has yet to be a global study of all TF families to summarize the timing and the degree of TF family expansion as a whole. In addition, we need a better understanding of the roles of TFs in multiple species so that functional information can be superimposed with phylogenetic relationships to have a better idea

of when plant innovations were established and what the ancestral functions of ancient TFs were. This will not only provide key information on the origin of novel traits, but will also be important for using evolutionary relationships across species to infer TF functions.

One of the most intriguing features of plant TFs is their elevated retention rate post WGD, and several mechanisms (e.g. neofunctionalization, subfunctionalization, escape from adaptive conflict, gene balance) have been proposed to explain retention. Although multiple lines of evidence have implicated the involvement of one or more of these mechanisms, we do not have a good quantitative understanding of the contribution of each mechanism to TF retention. Because of the lack of mutant phenotypes in several apparent loss-of-function TF mutants, duplicate TFs, like many other plant duplicates, are frequently regarded as redundant. However, duplicates with purely redundant functions are not expected to survive in the long run. Thus, a major question is whether the surviving duplicate TFs are still present because there has not been enough time for the accumulation of disabling mutations or because they are under selection. Related to this, it remains unresolved in most cases whether the selective force driving TF duplicate retention is adaptive (e.g. neofunctionalization) or not (e.g. subfunctionalization).

Based on our current knowledge, duplicate TFs have fundamental roles in plant evolution. Multiple lines of evidence linking novel plant traits to TF duplicates suggest that these TFs must have neofunctionalized repeatedly over the course of land plant evolution. However, given that many plant innovations took place tens to hundreds of million years ago, the exact evolutionary paths leading to these novelties are difficult to trace. Thus, to pin down the selective forces driving TF retention and to determine how retained TF duplicates contribute to plant novelties, it is of interest to study relatively recently evolved traits that are polymorphic among plants in both the laboratory and the field. Integrated research programs that combine molecular genetics and biochemistry with a strong ecological and evolutionary perspective will be crucial.

Transparency document

The Transparency document associated with this article can be found, in the online version.

Acknowledgement

This work was supported by the National Science Foundation grants MCB-1119778 and IOS-1126998 and a Michigan State University Discretionary Funding Initiative grant to S.-H.S.

References

- [1] M. Van Lijsebettens, K.D. Grasser, Transcript elongation factors: shaping transcriptomes after transcript initiation, *Trends Plant Sci.* 19 (2014) 717–726.
- [2] A.C.M. Cheung, P. Cramer, A movie of RNA polymerase II transcription, *Cell* 149 (2012) 1431–1437.
- [3] J.C. Hong, General Aspects of Plant Transcription Factor Families, *Plant Transcription Factors*, Elsevier 2016, pp. 35–56.
- [4] J.R. Haag, C.S. Pikaard, Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing, *Nat. Rev. Mol. Cell Biol.* 12 (2011) 483–492.
- [5] A.T. Wierzbicki, J.R. Haag, C.S. Pikaard, Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes, *Cell* 135 (2008) 635–648.
- [6] F. Ariel, N. Romero-Barrios, T. Jégu, M. Benhamed, M. Crespi, Battles and hijacks: noncoding transcription in plants, *Trends Plant Sci.* 20 (2015) 362–371.
- [7] K.B. Singh, Transcriptional regulation in plants: the importance of combinatorial control, *Plant Physiol.* 118 (1998) 1111–1120.
- [8] K. Kaufmann, A. Pajoro, G.C. Angenent, Regulation of transcription in plants: mechanisms controlling developmental switches, *Nat. Rev. Genet.* 11 (2010) 830–842.
- [9] M. Ikeuchi, A. Iwase, K. Sugimoto, Control of plant cell differentiation by histone modification and DNA methylation, *Curr. Opin. Plant Biol.* 28 (2015) 60–67.
- [10] J. Du, L.M. Johnson, S.E. Jacobsen, D.J. Patel, DNA methylation pathways and their crosstalk with histone methylation, *Nat. Rev. Mol. Cell Biol.* 16 (2015) 519–532.
- [11] W. Zhou, Y. Zhu, A. Dong, W.-H. Shen, Histone H2A/H2B chaperones: from molecules to chromatin-based functions in plant growth and development, *Plant J.* 83 (2015) 78–95.

[12] M.-F. Wu, N. Yamaguchi, J. Xiao, B. Bargmann, M. Estelle, Y. Sang, D. Wagner, Auxin-regulated chromatin switch directs acquisition of flower primordium founder fate, *Elife* 4 (2015), e09269.

[13] S. Kagle, K. Rozwadowski, EAR motif-mediated transcriptional repression in plants: an underlying mechanism for epigenetic regulation of gene expression, *Epigenetics* 6 (2011) 141–146.

[14] Y. Jing, H. Sun, W. Yuan, Y. Wang, Q. Li, Y. Liu, Y. Li, W. Qian, SUVH2 and SUVH9 couple two essential steps for transcriptional gene silencing in *Arabidopsis*, *Mol. Plant* (2016), <http://dx.doi.org/10.1016/j.molp.2016.05.006>.

[15] J.B. Heo, Y.-S. Lee, S. Sung, Epigenetic regulation by long noncoding RNAs in plants, *Chromosom. Res.* 21 (2013) 685–693.

[16] D. Holoch, D. Moazed, RNA-mediated epigenetic regulation of gene expression, *Nat. Rev. Genet.* 16 (2015) 71–84.

[17] N.Y. Rodriguez-Granados, J.S. Ramirez-Prado, A. Veluchamy, D. Latrasse, C. Raynaud, M. Crespi, F. Ariel, M. Benhamed, Put your 3D glasses on: plant chromatin is on show, *J. Exp. Bot.* 67 (2016) 3205–3221.

[18] Z. Wu, R. Ietswaart, F. Liu, H. Yang, M. Howard, C. Dean, Quantitative regulation of FLC via coordinated transcriptional initiation and elongation, *Proc. Natl. Acad. Sci. U. S. A.* 113 (2016) 218–223.

[19] N.J. Proudfoot, A. Furger, M.J. Dye, Integrating mRNA processing with transcription, *Cell* 108 (2002) 501–512.

[20] S. Nagaya, K. Kawamura, A. Shinmyo, K. Kato, The HSP terminator of *Arabidopsis thaliana* increases gene expression in plant cells, *Plant Cell Physiol.* 51 (2010) 328–332.

[21] C. Guo, M. Spinelli, M. Liu, Q.Q. Li, C. Liang, A genome-wide study of “Non-3UTR” polyadenylation sites in *Arabidopsis thaliana*, *Sci. Rep.* 6 (2016) 28060.

[22] N.J. Proudfoot, Transcriptional termination in mammals: stopping the RNA polymerase II juggernaut, *Science* 352 (2016) (aad9926).

[23] D.S. Latchman, Eukaryotic transcription factors, *Biochem. J.* 270 (1990) 281–289.

[24] D.S. Latchman, Transcription factors: an overview, *Int. J. Biochem. Cell Biol.* 29 (1997) 1305–1312.

[25] G. Gill, Regulation of the initiation of eukaryotic transcription, *Essays Biochem.* 37 (2001) 33–43.

[26] G.J. Narlikar, H.-Y. Fan, R.E. Kingston, Cooperation between complexes that regulate chromatin structure and transcription, *Cell* 108 (2002) 475–487.

[27] P.N. Benfey, Toward a systems analysis of the root, *Cold Spring Harb. Symp. Quant. Biol.* 77 (2012) 91–96.

[28] D.E. Soltis, P.S. Soltis, V.A. Albert, D.G. Oppenheimer, C.W. dePamphilis, H. Ma, M.W. Frohlich, G. Theissen, Floral genome project research group, missing links: the genetic architecture of flowers [correction of flower] and floral diversification, *Trends Plant Sci.* 7 (2002) 22–31 (discussion 31–4).

[29] E.S. Coen, E.M. Meyerowitz, The war of the whorls: genetic interactions controlling flower development, *Nature* 353 (1991) 31–37.

[30] Y. Nakai, Y. Nakahira, H. Sumida, K. Takebayashi, Y. Nagasawa, K. Yamasaki, M. Akiyama, M. Ohme-Takagi, S. Fujiwara, T. Shiina, N. Mitsuda, E. Fukusaki, Y. Kubo, M.H. Sato, Vascular plant one-zinc-finger protein 1/2 transcription factors regulate abiotic and biotic stress responses in *Arabidopsis*, *Plant J.* 73 (2013) 761–775.

[31] K. Nakashima, H. Takasaki, J. Mizoi, K. Shinozaki, K. Yamaguchi-Shinozaki, NAC transcription factors in plant abiotic stress responses, *Biochim. Biophys. Acta* 1819 (2012) 97–103.

[32] K. Singh, R.C. Foley, L. Oñate-Sánchez, Transcription factors in plant defense and stress responses, *Curr. Opin. Plant Biol.* 5 (2002) 430–436.

[33] H. Jin, C. Martin, Multifunctionality and diversity within the plant MYB-gene family, *Plant Mol. Biol.* 41 (1999) 577–585.

[34] J.M. Gendron, J.L. Pruneda-Paz, C.J. Doherty, A.M. Gross, S.E. Kang, S.A. Kay, *Arabidopsis* circadian clock protein, TOC1, is a DNA-binding transcription factor, *Proc. Natl. Acad. Sci. U. S. A.* 109 (2012) 3167–3172.

[35] N. Nader, G.P. Chrousos, T. Kino, Circadian rhythm transcription factor CLOCK regulates the transcriptional activity of the glucocorticoid receptor by acetylating its hinge region lysine cluster: potential physiological implications, *FASEB J.* 23 (2009) 1572–1583.

[36] L. Corbesier, C. Vincent, S. Jang, F. Fornara, Q. Fan, I. Searle, A. Giakountis, S. Farrona, L. Gissot, C. Turnbull, G. Coupland, FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*, *Science* 316 (2007) 1030–1033.

[37] A. Bastidas, M. Yañez, S. Osorio, V. Arbona, A. Gómez-Cadenas, A.R. Fernie, J.A. Casaretto, The transcription factor AREB1 regulates primary metabolic pathways in tomato fruits, *J. Exp. Bot.* 65 (2014) 2351–2363.

[38] W.-C. Kim, J.-H. Ko, K.-H. Han, Identification of a cis-acting regulatory motif recognized by MYB46, a master transcriptional regulator of secondary wall biosynthesis, *Plant Mol. Biol.* 78 (2012) 489–501.

[39] M. Taylor-Teeple, L. Lin, M. de Lucas, G. Turco, T.W. Toal, A. Gaudinier, N.F. Young, G.M. Trabucco, M.T. Veling, R. Lamothe, P.P. Handakumbura, G. Xiong, C. Wang, J. Corwin, A. Tsoukalas, L. Zhang, D. Ware, M. Pauly, D.J. Kliebenstein, K. Dehesh, I. Tagkopoulos, G. Breton, J.L. Pruneda-Paz, S.E. Ahnert, S.A. Kay, S.P. Hazen, S.M. Brady, An *Arabidopsis* gene regulatory network for secondary cell wall synthesis, *Nature* 517 (2015) 571–575.

[40] S. Della Pina, E. Souer, R. Koes, Arguments in the evo-devo debate: say it with flowers! *J. Exp. Bot.* 65 (2014) 2231–2242.

[41] J.L. Riechmann, J. Heard, G. Martin, L. Reuber, C. Jiang, J. Keddie, L. Adam, O. Pineda, O.J. Ratcliffe, R.R. Samaha, R. Creelman, M. Pilgrim, P. Broun, J.Z. Zhang, D. Ghandehari, B.K. Sherman, G. Yu, *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes, *Science* 290 (2000) 2105–2110.

[42] N. Mitsuda, M. Ohme-Takagi, Functional analysis of transcription factors in *Arabidopsis*, *Plant Cell Physiol.* 50 (2009) 1232–1248.

[43] K. Yamasaki, T. Kigawa, M. Seki, K. Shinozaki, S. Yokoyama, DNA-binding domains of plant-specific transcription factors: structure, function, and evolution, *Trends Plant Sci.* 18 (2013) 267–276.

[44] A. de Mendoza, A. Sebé-Pedrós, M.S. Šesták, M. Matejcic, G. Torruella, T. Domazet-Loso, I. Ruiz-Trillo, Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages, *Proc. Natl. Acad. Sci. U. S. A.* 110 (2013) E4858–E4866.

[45] J.M. Franco-Zorrilla, I. López-Vidriero, J.L. Carrasco, M. Godoy, P. Vera, R. Solano, DNA-binding specificities of plant transcription factors and their potential to define target genes, *Proc. Natl. Acad. Sci. U. S. A.* 111 (2014) 2367–2372.

[46] D.M. Riaño-Pachón, S. Ruzicic, I. Dreyer, B. Mueller-Roeber, PlnTFDB: an integrative plant transcription factor database, *BMC Bioinform.* 8 (2007) 42.

[47] S. Richardt, D. Lang, R. Reski, W. Frank, S.A. Rensing, PlanTAPDB, a phylogeny-based resource of plant transcription-associated proteins, *Plant Physiol.* 143 (2007) 1452–1466.

[48] J. Jin, H. Zhang, L. Kong, G. Gao, J. Luo, PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors, *Nucleic Acids Res.* 42 (2014) D1182–D1187.

[49] E. Kejnovsky, I.J. Leitch, A.R. Leitch, Contrasting evolutionary dynamics between angiosperm and mammalian genomes, *Trends Ecol. Evol.* 24 (2009) 572–582.

[50] F. Murat, Y. Van de Peer, J. Salse, Decoding plant and animal genome plasticity from differential paleo-evolutionary patterns and processes, *Genome Biol. Evol.* 4 (2012) 917–928.

[51] G. Blanc, K.H. Wolfe, Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution, *Plant Cell* 16 (2004) 1679–1691.

[52] C. Seoighe, C. Gehring, Genome duplication led to highly selective expansion of the *Arabidopsis thaliana* proteome, *Trends Genet.* 20 (2004) 461–464.

[53] S.-H. Shiu, M.-C. Shih, W.-H. Li, Transcription factor families have much higher expansion rates in plants than in animals, *Plant Physiol.* 139 (2005) 18–26.

[54] W.-K. Jiang, Y.-L. Liu, E.-H. Xia, L.-Z. Gao, Prevalent role of gene features in determining evolutionary fates of whole-genome duplication duplicated genes in flowering plants, *Plant Physiol.* 161 (2013) 1844–1861.

[55] S. Kim, P.S. Soltis, K. Wall, D.E. Soltis, Phylogeny and domain evolution in the APETALA2-like gene family, *Mol. Biol. Evol.* 23 (2006) 107–120.

[56] J. Mizoi, K. Shinozaki, K. Yamaguchi-Shinozaki, AP2/ERF family transcription factors in plant abiotic stress responses, *Biochim. Biophys. Acta* 1819 (2012) 86–96.

[57] H. Shan, L. Zahn, S. Guindon, P.K. Wall, H. Kong, H. Ma, C.W. dePamphilis, J. Leebens-Mack, Evolution of plant MADS box transcription factors: evidence for shifts in selection associated with early angiosperm diversification and concerted gene duplications, *Mol. Biol. Evol.* 26 (2009) 2229–2244.

[58] L. Gramzow, M.S. Ritz, G. Theissen, On the origin of MADS-domain transcription factors, *Trends Genet.* 26 (2010) 149–153.

[59] A. Feller, K. Machemer, E.L. Braun, E. Grotewold, Evolutionary and comparative analysis of MYB and bHLH plant transcription factors, *Plant J.* 66 (2011) 94–116.

[60] H. Du, Z. Liang, S. Zhao, M.-G. Nan, L.-S.P. Tran, K. Lu, Y.-B. Huang, J.-N. Li, The evolutionary history of R2R3-MYB proteins across 50 eukaryotes: new insights into subfamily classification and expansion, *Sci. Rep.* 5 (2015) 11037.

[61] Y. Qin, X. Ma, G. Yu, Q. Wang, L. Wang, L. Kong, W. Kim, H.W. Wang, Evolutionary history of trihelix family and their functional diversification, *DNA Res.* 21 (2014) 499–510.

[62] R.A. Mentink, M. Tsiantis, From limbs to leaves: common themes in evolutionary diversification of organ form, *Front. Genet.* 6 (2015) 284.

[63] N.W. Albert, K.M. Davies, K.E. Schwinn, Gene regulation networks generate diverse pigmentation patterns in plants, *Plant Signal. Behav.* 9 (2014) e29526.

[64] J. Nardmann, W. Werr, The evolution of plant regulatory networks: what *Arabidopsis* cannot say for itself, *Curr. Opin. Plant Biol.* 10 (2007) 653–659.

[65] L. Liu, M.J. White, T.H. MacRae, Transcription factors and their genes in higher plants functional domains, evolution and regulation, *Eur. J. Biochem.* 262 (1999) 247–257.

[66] X. Han, D. Kumar, H. Chen, S. Wu, J.-Y. Kim, Transcription factor-mediated cell-to-cell signalling in plants, *J. Exp. Bot.* 65 (2014) 1737–1749.

[67] T. Laloum, S. De Mita, P. Gamas, M. Baudin, A. Niebel, CCAAT-box binding transcription factors in plants: Y so many? *Trends Plant Sci.* 18 (2013) 157–166.

[68] D.H. Gonzalez, *Plant Transcription Factors: Evolutionary, Structural and Functional Aspects*, Academic Press, 2015.

[69] A.-Y. Guo, X. Chen, G. Gao, H. Zhang, Q.-H. Zhu, X.-C. Liu, Y.-F. Zhong, X. Gu, K. He, J. Luo, PlantTFDB: a comprehensive plant transcription factor database, *Nucleic Acids Res.* 36 (2008) D966–D969.

[70] D. Lang, B. Weiche, G. Timmerhaus, S. Richardt, D.M. Riaño-Pachón, L.G.G. Corrêa, R. Reski, B. Mueller-Roeber, S.A. Rensing, Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: a timeline of loss, gain, expansion, and correlation with complexity, *Genome Biol. Evol.* 2 (2010) 488–503.

[71] T. Nagata, N. Toshifumi, H.-S. Aeni, K. Shoshi, The evolutionary diversification of genes that encode transcription factor proteins in plants, *Plant Transcription Factors* 2016, pp. 73–97.

[72] R.D. Finn, P. Coggill, R.Y. Eberhardt, S.R. Eddy, J. Mistry, A.L. Mitchell, S.C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G.A. Salazar, J. Tate, A. Bateman, The Pfam protein families database: towards a more sustainable future, *Nucleic Acids Res.* 44 (2016) D279–D285.

[73] F. Burki, The eukaryotic tree of life from a global phylogenomic perspective, *Cold Spring Harb. Perspect. Biol.* 6 (2014) a016147.

[74] K. Weston, Myb proteins in life, death and differentiation, *Curr. Opin. Genet. Dev.* 8 (1998) 76–81.

[75] Z.H. Fang, Z.C. Han, The transcription factor E2F: a crucial switch in the control of homeostasis and tumorigenesis, *Histol. Histopathol.* 21 (2006) 403–413.

[76] V. Ledent, M. Vervoort, The basic helix-loop-helix protein family: comparative genomics and phylogenetic analysis, *Genome Res.* 11 (2001) 754–770.

[77] R.D. Kortschak, P.W. Tucker, R. Saint, ARID proteins come in from the desert, *Trends Biochem. Sci.* 25 (2000) 294–299.

[78] P. Benatti, D. Dolfini, A. Viganò, M. Ravo, A. Weisz, C. Imbriano, Specific inhibition of NF-Y subunits triggers different cell proliferation defects, *Nucleic Acids Res.* 39 (2011) 5356–5368.

[79] G. Caretti, V. Salsi, C. Vecchi, C. Imbriano, R. Mantovani, Dynamic recruitment of NF-Y and histone acetyltransferases on cell-cycle promoters, *J. Biol. Chem.* 278 (2003) 30435–30440.

[80] R.W. Deed, E. Hara, G.T. Atherton, G. Peters, J.D. Norton, Regulation of Id3 cell cycle function by Cdk-2-dependent phosphorylation, *Mol. Cell. Biol.* 17 (1997) 6815–6821.

[81] M. Rövekamp, J.L. Bowman, U. Grossniklaus, *Marchantia* MpRKD regulates the gametophyte-sporophyte transition by keeping egg cells quiescent in the absence of fertilization, *Curr. Biol.* 26 (2016) 1782–1789.

[82] S. Koi, T. Hisanaga, K. Sato, M. Shimamura, K.T. Yamato, K. Ishizaki, T. Kohchi, K. Nakajima, An evolutionarily conserved plant RKD factor controls germ cell differentiation, *Curr. Biol.* 26 (2016) 1775–1781.

[83] H. Lin, U.W. Goodenough, Gametogenesis in the *Chlamydomonas reinhardtii* minus mating type is controlled by two genes, MID and MTD1, *Genetics* 176 (2007) 913–925.

[84] P.J. Ferris, U.W. Goodenough, Mating type in *Chlamydomonas* is specified by mid, the minus-dominance gene, *Genetics* 146 (1997) 859–869.

[85] A. Kanhere, M. Bansal, Structural properties of promoters: similarities and differences between prokaryotes and eukaryotes, *Nucleic Acids Res.* 33 (2005) 3165–3175.

[86] K. Struhl, Fundamentally different logic of gene regulation in eukaryotes and prokaryotes, *Cell* 98 (1999) 1–4.

[87] P.P. Schrumpfová, M. Fojtová, P. Mokroš, K.D. Grasser, J. Fajkus, Role of HMGB proteins in chromatin dynamics and telomere maintenance in *Arabidopsis thaliana*, *Curr. Protein Pept. Sci.* 12 (2011) 105–111.

[88] M. Stros, D. Launholt, K.D. Grasser, The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins, *Cell. Mol. Life Sci.* 64 (2007) 2590–2606.

[89] Y. Cai, J. Jin, L. Florens, S.K. Swanson, T. Kusch, B. Li, J.L. Workman, M.P. Washburn, R.C. Conaway, J.W. Conaway, The mammalian Y1 protein is a shared subunit of the TRRAP/TIP60 histone acetyltransferase and SRCAP complexes, *J. Biol. Chem.* 280 (2005) 13665–13670.

[90] S. Jin, K.W. Scotto, Transcriptional regulation of the MDR1 gene by histone acetyltransferase and deacetylase is mediated by NF-Y, *Mol. Cell. Biol.* 18 (1998) 4377–4384.

[91] M.W. Linhoff, K.L. Wright, J.P. Ting, CCAAT-binding factor NF-Y and RFX are required for in vivo assembly of a nucleoprotein complex that spans 250 base pairs: the invariant chain promoter as a model, *Mol. Cell. Biol.* 17 (1997) 4589–4596.

[92] C. Scazzocchio, The fungal GATA factors, *Curr. Opin. Microbiol.* 3 (2000) 126–131.

[93] J.C. Pearson, D. Lemons, W. McGinnis, Modulating Hox gene functions during animal body patterning, *Nat. Rev. Genet.* 6 (2005) 893–904.

[94] S.G. Ahn, P.C. Liu, K. Klyachko, R.I. Morimoto, D.J. Thiele, The loop domain of heat shock transcription factor 1 dictates DNA-binding specificity and responses to heat stress, *Genes Dev.* 15 (2001) 2134–2145.

[95] X. Deng, J. Yang, X. Wu, Y. Li, X. Fei, A C2H2 zinc finger protein FEMU2 is required for fox1 expression in *Chlamydomonas reinhardtii*, *PLoS One* 9 (2014), e112977.

[96] T. Hai, C.D. Wolfgang, D.K. Marsee, A.E. Allen, U. Sivaprasad, ATF3 and stress responses, *Gene Expr.* 7 (1999) 321–335.

[97] H. Du, L. Zhang, L. Liu, X.-F. Tang, W.-J. Yang, Y.-M. Wu, Y.-B. Huang, Y.-X. Tang, Biochemical and molecular characterization of plant MYB transcription factor family, *Biochemistry* 74 (2009) 1–11.

[98] C.D. Deppmann, R.S. Alvarez, E.J. Taparowsky, Cross-species annotation of basic leucine zipper factor interactions: insight into the evolution of closed interaction networks, *Mol. Biol. Evol.* 23 (2006) 1480–1492.

[99] K. Matuoka, K.Y. Chen, Nuclear factor Y (NF-Y) and cellular senescence, *Exp. Cell Res.* 253 (1999) 365–371.

[100] A. Finkler, R. Ashery-Padan, H. Fromm, CAMTAs: calmodulin-binding transcription activators from plants to human, *FEBS Lett.* 581 (2007) 3893–3898.

[101] M.J. Bottomley, M.W. Collard, J.I. Huggenvik, Z. Liu, T.J. Gibson, M. Sattler, The SAND domain structure defines a novel DNA-binding fold in transcriptional regulation, *Nat. Struct. Biol.* 8 (2001) 626–633.

[102] P. Peterson, T. Org, A. Rebane, Transcriptional regulation by AIRE: molecular mechanisms of central tolerance, *Nat. Rev. Immunol.* 8 (2008) 948–957.

[103] C.C. Carles, J.C. Fletcher, The SAND domain protein ULTRAPETALA1 acts as a trithorax group factor to regulate cell fate in plants, *Genes Dev.* 23 (2009) 2723–2728.

[104] C.C. Carles, D. Choffnes-Inada, K. Reville, K. Lertpiriyapong, J.C. Fletcher, ULTRAPETALA1 encodes a SAND domain putative transcriptional regulator that controls shoot and floral meristem activity in *Arabidopsis*, *Development* 132 (2005) 897–911.

[105] C.C. Carles, K. Lertpiriyapong, K. Reville, J.C. Fletcher, The ULTRAPETALA1 gene functions early in *Arabidopsis* development to restrict shoot apical meristem activity and acts through WUSCHEL to regulate floral meristem determinacy, *Genetics* 167 (2004) 1893–1903.

[106] M.M. Monfared, C.C. Carles, P. Rossignol, H.R. Pires, J.C. Fletcher, The ULT1 and ULT2 trxG genes play overlapping roles in *Arabidopsis* development and gene regulation, *Mol. Plant* 6 (2013) 1564–1579.

[107] N. Ashraf, D. Jain, R.A. Vishwakarma, Identification, cloning and characterization of an uppetala transcription factor CsULT1 from *Crocus*: a novel regulator of apocarotenoid biosynthesis, *BMC Plant Biol.* 15 (2015) 25.

[108] H.J. Painter, T.L. Campbell, M. Llinás, The Apicomplexan AP2 family: integral factors regulating *Plasmodium* development, *Mol. Biochem. Parasitol.* 176 (2011) 1–7.

[109] C.M. Ha, J.H. Jun, J.C. Fletcher, Control of *Arabidopsis* leaf morphogenesis through regulation of the YABBY and KNOX families of transcription factors, *Genetics* 186 (2010) 197–206.

[110] B. Cong, L.S. Barrero, S.D. Tanksley, Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication, *Nat. Genet.* 40 (2008) 800–804.

[111] M. Dai, Y. Hu, Y. Zhao, D.-X. Zhou, Regulatory networks involving YABBY genes in rice shoot development, *Plant Signal. Behav.* 2 (2007) 399–400.

[112] J.-Y. Lee, S.F. Baum, J. Alvarez, A. Patel, D.H. Chitwood, J.L. Bowman, Activation of CRABS CLAW in the nectaries and carpels of *Arabidopsis*, *Plant Cell* 17 (2005) 25–36.

[113] Z. Lin, X. Li, L.M. Shannon, C.-T. Yeh, M.L. Wang, G. Bai, Z. Peng, J. Li, H.N. Trick, T.E. Clemente, J. Doebley, P.S. Schnable, M.R. Tuinstra, T.T. Tesso, F. White, J. Yu, Parallel domestication of the Shattering1 genes in cereals, *Nat. Genet.* 44 (2012) 720–724.

[114] M. Dai, Y. Zhao, Q. Ma, Y. Hu, P. Hedden, Q. Zhang, D.-X. Zhou, The rice YABBY1 gene is involved in the feedback regulation of gibberellin metabolism, *Plant Physiol.* 144 (2007) 121–133.

[115] T. Eulgem, P.J. Rushton, S. Robatzek, I.E. Somssich, The WRKY superfamily of plant transcription factors, *Trends Plant Sci.* 5 (2000) 199–206.

[116] L. Cappadocia, A. Maréchal, J.-S. Parent, E. Lepage, J. Sygusch, N. Brisson, Crystal structures of DNA-Whirly complexes and their role in *Arabidopsis* organelle genome repair, *Plant Cell* 22 (2010) 1849–1867.

[117] A. Maréchal, J.-S. Parent, F. Véronneau-Lafortune, A. Joyeux, B.F. Lang, N. Brisson, Whirly proteins maintain plastid genome stability in *Arabidopsis*, *Proc. Natl. Acad. Sci. U. S. A.* 106 (2009) 14693–14698.

[118] J. Prikryl, K.P. Watkins, G. Friso, K.J. van Wijk, A. Barkan, A member of the Whirly family is a multifunctional RNA- and DNA-binding protein that is essential for chloroplast biogenesis, *Nucleic Acids Res.* 36 (2008) 5152–5165.

[119] H.H. Yoo, C. Kwon, M.M. Lee, I.K. Chung, Single-stranded DNA binding factor AtWHY1 modulates telomere length homeostasis in *Arabidopsis*, *Plant J.* 49 (2007) 442–451.

[120] D. Desveaux, R. Subramaniam, C. Després, J.-N. Mess, C. Lévesque, P.R. Fobert, J.L. Dangl, N. Brisson, A “Whirly” transcription factor is required for salicylic acid-dependent disease resistance in *Arabidopsis*, *Dev. Cell* 6 (2004) 229–240.

[121] Y. Miao, J. Jiang, Y. Ren, Z. Zhao, The single-stranded DNA-binding protein WHIRLY1 represses WRKY53 expression and delays leaf senescence in a developmental stage-dependent manner in *Arabidopsis*, *Plant Physiol.* 163 (2013) 746–756.

[122] F. Licausi, M. Ohme-Takagi, P. Perata, APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs, *New Phytol.* 199 (2013) 639–649.

[123] M. Yuda, S. Iwanaga, S. Shigenobu, G.R. Mair, C.J. Janse, A.P. Waters, T. Kato, I. Kaneko, Identification of a transcription factor in the mosquito-invasive stage of malaria parasites, *Mol. Microbiol.* 71 (2009) 1402–1414.

[124] C. Flueck, R. Bartfai, I. Niederwieser, K. Witmer, B.T.F. Alako, S. Moes, Z. Bozdech, P. Jenoe, H.G. Stunnenberg, T.S. Voss, A major role for the *Plasmodium falciparum* ApiAP2 protein PfSIP2 in chromosome end biology, *PLoS Pathog.* 6 (2010), e1000784.

[125] J.B. Radke, O. Lucas, E.K. De Silva, Y. Ma, W.J. Sullivan Jr., L.M. Weiss, M. Llinás, M.W. White, ApiAP2 transcription factor restricts development of the *Toxoplasma* tissue cyst, *Proc. Natl. Acad. Sci. U. S. A.* 110 (2013) 6871–6876.

[126] J.F. Golz, A. Hudson, Plant development: YABBYs claw to the fore, *Curr. Biol.* 9 (1999) R861–R863.

[127] D. Desveaux, J. Allard, N. Brisson, J. Sygusch, A new family of plant transcription factors displays a novel ssDNA-binding surface, *Nat. Struct. Biol.* 9 (2002) 512–517.

[128] C.I. Rinerson, R.C. Rabara, P. Tripathi, Q.J. Shen, P.J. Rushton, The evolution of WRKY transcription factors, *BMC Plant Biol.* 15 (2015) 66.

[129] S.R. Wessler, Homing into the origin of the AP2 DNA binding domain, *Trends Plant Sci.* 10 (2005) 54–56.

[130] D.R. McCarty, T. Hattori, C.B. Carson, V. Vasil, M. Lazar, I.K. Vasil, The Viviparous-1 developmental gene of maize encodes a novel transcriptional activator, *Cell* 66 (1991) 895–905.

[131] K. Swaminathan, K. Peterson, T. Jack, The plant B3 superfamily, *Trends Plant Sci.* 13 (2008) 647–655.

[132] Y. Kagaya, K. Ohmiya, T. Hattori, RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants, *Nucleic Acids Res.* 27 (1999) 470–478.

[133] J. Jin, K. He, X. Tang, Z. Li, L. Lv, Y. Zhao, J. Luo, G. Gao, An *Arabidopsis* transcriptional regulatory map reveals distinct functional and evolutionary features of novel transcription factors, *Mol. Biol. Evol.* 32 (2015) 1767–1773.

[134] S. Yanagisawa, Dof domain proteins: plant-specific transcription factors associated with diverse phenomena unique to plants, *Plant Cell Physiol.* 45 (2004) 386–391.

[135] S. Yanagisawa, Dof DNA-binding domains of plant transcription factors contribute to multiple protein–protein interactions, *Eur. J. Biochem.* 250 (1997) 403–410.

[136] S. Yanagisawa, A novel DNA-binding domain that may form a single zinc finger motif, *Nucleic Acids Res.* 23 (1995) 3403–3410.

[137] Y. Nagano, H. Furuhashi, T. Inaba, Y. Sasaki, A novel class of plant-specific zinc-dependent DNA-binding protein that binds to A/T-rich DNA sequences, *Nucleic Acids Res.* 29 (2001) 4097–4105.

[138] J. Klein, H. Saedler, P. Huijser, A new family of DNA binding proteins includes putative transcriptional regulators of the *Antirrhinum majus* floral meristem identity gene SQUAMOSA, *Mol. Gen. Genet.* 250 (1996) 7–16.

[139] E. van der Knaap, J.H. Kim, H. Kende, A novel gibberellin-induced gene from rice and its potential regulatory role in stem growth, *Plant Physiol.* 122 (2000) 695–704.

[140] F. Chevalier, D. Perazza, F. Laporte, G. Le Hénaff, P. Hornitschek, J.-M. Bonneville, M. Herzog, G. Vachon, GeBP and GeBP-like proteins are noncanonical leucine-zipper transcription factors that regulate cytokinin response in *Arabidopsis*, *Plant Physiol.* 146 (2008) 1142–1154.

[141] M. Capella, C. Matias, P.A. Ribone, A.L. Arce, R.L. Chan, Homeodomain-leucine zipper transcription factors: structural features of these proteins, unique to plants, *Plant Transcription Factors* 2016, pp. 113–126.

[142] G. Rubin, T. Tohge, F. Matsuda, K. Saito, W.-R. Scheible, Members of the LBD family of transcription factors repress anthocyanin synthesis and affect additional nitrogen responses in *Arabidopsis*, *Plant Cell* 21 (2009) 3567–3584.

[143] I. Fridborg, S. Kuusk, M. Robertson, E. Sundberg, The *Arabidopsis* protein SHI represses gibberellin responses in *Arabidopsis* and barley, *Plant Physiol.* 127 (2001) 937–948.

[144] J. Kropat, S. Tottey, R.P. Birkenbihl, N. Depèze, P. Huijser, S. Merchant, A regulator of nutritional copper signaling in *Chlamydomonas* is an SBP domain protein that recognizes the GTAC core of copper response element, *Proc. Natl. Acad. Sci. U. S. A.* 102 (2005) 18730–18735.

[145] H. Yamasaki, M. Hayashi, M. Fukazawa, Y. Kobayashi, T. Shikanai, SQUAMOSA promoter binding protein-like7 is a central regulator for copper homeostasis in *Arabidopsis*, *Plant Cell* 21 (2009) 347–361.

[146] S. Li, The *Arabidopsis thaliana* TCP transcription factors: a broadening horizon beyond development, *Plant Signal. Behav.* 10 (2015), e1044192.

[147] E.A.C. Romanel, C.G. Schrago, R.M. Couñago, C.A.M. Russo, M. Alves-Ferreira, Evolution of the B3 DNA binding superfamily: new insights into REM family gene diversification, *PLoS One* 4 (2009), e5791.

[148] M. Shigyo, N. Tabei, T. Yoneyama, S. Yanagisawa, Evolutionary processes during the formation of the plant-specific Dof transcription factor family, *Plant Cell Physiol.* 48 (2007) 179–185.

[149] M. Riese, O. Zobell, H. Saedler, P. Huijser, SBP-domain transcription factors as possible effectors of cryptochrome-mediated blue light signalling in the moss *Physcomitrella patens*, *Planta* 227 (2008) 505–515.

[150] A.N. Olsen, H.A. Ernst, L.L. Leggio, K. Skriver, NAC transcription factors: structurally distinct, functionally diverse, *Trends Plant Sci.* 10 (2005) 79–87.

[151] Y. Matsumura, H. Iwakawa, Y. Machida, C. Machida, Characterization of genes in the ASYMMETRIC LEAVES2/LATERAL ORGAN BOUNDARIES (AS2/LOB) family in *Arabidopsis thaliana*, and functional and molecular comparisons between AS2 and other family members, *Plant J.* 58 (2009) 525–537.

[152] C. Bolle, The role of GRAS proteins in plant signal transduction and development, *Planta* 218 (2004) 683–692.

[153] N.S. Siriwardana, R.S. Lamb, The poetry of reproduction: the role of LEAFY in *Arabidopsis thaliana* flower formation, *Int. J. Dev. Biol.* 56 (2012) 207–221.

[154] A. Yamaguchi, M.-F. Wu, L. Yang, G. Wu, R.S. Poethig, D. Wagner, The microRNA-regulated SBP-Box transcription factor SPL3 is a direct upstream activator of LEAFY, FRUITFULL, and APETALA1, *Dev. Cell* 17 (2009) 268–278.

[155] B. Wei, J. Zhang, C. Pang, H. Yu, D. Guo, H. Jiang, M. Ding, Z. Chen, Q. Tao, H. Gu, L.-J. Qu, G. Qin, The molecular mechanism of sporocysteless/nozzle in controlling *Arabidopsis* ovule development, *Cell Res.* 25 (2015) 121–134.

[156] A.-Y. Guo, Q.-H. Zhu, X. Gu, S. Ge, J. Yang, J. Luo, Genome-wide identification and evolutionary analysis of the plant specific SBP-box transcription factor family, *Gene* 418 (2008) 1–8.

[157] B.M. Binder, J.M. Walker, J.M. Gagne, T.J. Emborg, G. Hemmann, A.B. Bleeker, R.D. Vierstra, The *Arabidopsis* EIN3 binding F-Box proteins EBF1 and EBF2 have distinct but overlapping roles in ethylene signaling, *Plant Cell* 19 (2007) 509–523.

[158] L. Li, X.W. Deng, It runs in the family: regulation of brassinosteroid signaling by the BZR1-BES1 class of transcription factors, *Trends Plant Sci.* 10 (2005) 266–268.

[159] P. Kaló, C. Gleason, A. Edwards, J. Marsh, R.M. Mitra, S. Hirsch, J. Jakab, S. Sims, S.R. Long, J. Rogers, G.B. Kiss, J.A. Downie, G.E.D. Oldroyd, Nodulation signaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulators, *Science* 308 (2005) 1786–1789.

[160] B. Fode, T. Siemsen, C. Thurow, R. Weigel, C. Gatz, The *Arabidopsis* GRAS protein SCL14 interacts with class II TGA transcription factors and is essential for the activation of stress-inducible promoters, *Plant Cell* 20 (2008) 3122–3135.

[161] D. Zhang, L.M. Iyer, L. Aravind, Bacterial GRAS domain proteins throw new light on gibberellin acid response mechanisms, *Bioinformatics* 28 (2012) 2407–2411.

[162] H. Kurokawa, H. Motohashi, S. Sueno, M. Kimura, H. Takagawa, Y. Kanno, M. Yamamoto, T. Tanaka, Structural basis of alternative DNA recognition by Maf transcription factors, *Mol. Cell. Biol.* 29 (2009) 6232–6244.

[163] H. Liang, E.T. Olejniczak, X. Mao, D.G. Nettesheim, L. Yu, C.B. Thompson, S.W. Fesik, The secondary structure of the ets domain of human Fli-1 resembles that of the helix-turn-helix DNA-binding motif of the *Escherichia coli* catabolite gene activator protein, *Proc. Natl. Acad. Sci. U. S. A.* 91 (1994) 11655–11659.

[164] U. Häcker, U. Grossniklaus, W.J. Gehring, H. Jäckle, Developmentally regulated *Drosophila* gene family encoding the fork head domain, *Proc. Natl. Acad. Sci. U. S. A.* 89 (1992) 8754–8758.

[165] A. van der Horst, B.M.T. Burgering, Stressing the role of FoxO proteins in lifespan and disease, *Nat. Rev. Mol. Cell Biol.* 8 (2007) 440–450.

[166] J.-J. Wang, L. Qiu, Q. Cai, S.-H. Ying, M.-G. Feng, Transcriptional control of fungal cell cycle and cellular events by Fkh2, a forkhead transcription factor in an insect pathogen, *Sci. Rep.* 5 (2015) 10108.

[167] T. Oikawa, T. Yamada, Molecular biology of the Ets family of transcription factors, *Gene* 303 (2003) 11–34.

[168] E. Shelest, Transcription factors in fungi, *FEMS Microbiol. Lett.* 286 (2008) 145–151.

[169] R.B. Turner, D.L. Smith, M.E. Zawrotny, M.F. Summers, M.C. Posewitz, D.R. Winge, Solution structure of a zinc domain conserved in yeast copper-regulated transcription factors, *Nat. Struct. Biol.* 5 (1998) 551–555.

[170] Y. Jiao, N.J. Wickett, S. Ayyampalayam, A.S. Chanderbali, L. Landherr, P.E. Ralph, L.P. Tomsho, Y. Hu, H. Liang, P.S. Soltis, D.E. Soltis, S.W. Clifton, S.E. Schlarbaum, S.C. Schuster, H. Ma, J. Leebens-Mack, C.W. dePamphilis, Ancestral polyploidy in seed plants and angiosperms, *Nature* 473 (2011) 97–100.

[171] S.A. Rensing, D. Lang, A.D. Zimmer, A. Terry, A. Salamov, H. Shapiro, T. Nishizawa, P.-F. Perroud, E.A. Lindquist, Y. Kamiyama, T. Tanahashi, K. Sakakibara, T. Fujita, K. Oishi, T. Shin-I, Y. Kuroki, A. Toyoda, Y. Suzuki, S.-I. Hashimoto, K. Yamaguchi, S. Sugano, Y. Kohara, A. Fujiyama, A. Anterola, S. Aoki, N. Ashton, W.B. Barbazuk, E. Barker, J.L. Bennetzen, R. Blankenship, S.H. Cho, S.K. Dutcher, M. Estelle, J.A. Fawcett, H. Gundlach, K. Hanada, A. Heyl, K.A. Hicks, J. Hughes, M. Lohr, K. Mayer, A. Melkozernov, T. Murata, D.R. Nelson, B. Pils, M. Prigge, B. Reiss, T. Renner, S. Rombauds, P.J. Rushton, A. Sanderfoot, G. Schween, S.-H. Shiu, K. Stueber, F.L. Theodoulou, H. Tu, Y. Van de Peer, P.J. Verrier, E. Waters, A. Wood, L. Yang, D. Cove, A.C. Cuming, M. Hasebe, S. Lucas, B.D. Mishler, R. Reski, I.V. Grigoriev, R.S. Quatrano, J.L. Boore, The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants, *Science* 319 (2008) 64–69.

[172] M.J. Sanderson, J.L. Thorne, N. Wikström, K. Bremer, Molecular evidence on plant divergence times, *Am. J. Bot.* 91 (2004) 1656–1665.

[173] P.G. Gensel, The earliest land plants, *Annu. Rev. Ecol. Evol. Syst.* 39 (2008) 459–477.

[174] L. Hu, S. Liu, Genome-wide identification and phylogenetic analysis of the ERF gene family in cucumbers, *Genet. Mol. Biol.* 34 (2011) 624–633.

[175] A. Pereira-Santana, L.D. Alcaraz, E. Castaño, L. Sanchez-Calderon, F. Sanchez-Teyer, L. Rodriguez-Zapata, Comparative genomics of NAC transcriptional factors in angiosperms: implications for the adaptation and diversification of flowering plants, *PLoS One* 10 (2015), e0141866.

[176] G. Blanc, K. Hökamp, K.H. Wolfe, A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome, *Genome Res.* 13 (2003) 137–144.

[177] J.E. Bowers, B.A. Chapman, J. Rong, A.H. Paterson, Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events, *Nature* 422 (2003) 433–438.

[178] J. Zhang, Z. Jianzhi, Evolution by gene duplication: an update, *Trends Ecol. Evol.* 18 (2003) 292–298.

[179] N. Jiang, Z. Bao, X. Zhang, S.R. Eddy, S.R. Wessler, Pack-MULE transposable elements mediate gene evolution in plants, *Nature* 431 (2004) 569–573.

[180] G. Drouin, G.A. Dover, Independent gene evolution in the potato actin gene family demonstrated by phylogenetic procedures for resolving gene conversions and the phylogeny of angiosperm actin genes, *J. Mol. Evol.* 31 (1990) 132–150.

[181] J. Brosius, Retroposons—seeds of evolution, *Science* 251 (1991) 753.

[182] J. Yu, T. Ke, S. Tehrim, F. Sun, B. Liao, W. Hua, PTGBase: an integrated database to study tandem duplicated genes in plants, *Database* 2015 (2015), <http://dx.doi.org/10.1093/database/bav017>.

[183] D.E. Soltis, V.A. Albert, J. Leebens-Mack, C.D. Bell, A.H. Paterson, C. Zheng, D. Sankoff, C.W. Depamphilis, P.K. Wall, P.S. Soltis, Polyploidy and angiosperm diversification, *Am. J. Bot.* 96 (2009) 336–348.

[184] S. Maere, S. De Bodt, J. Raes, T. Casneuf, M. Van Montagu, M. Kuiper, Y. Van de Peer, Modeling gene and genome duplications in eukaryotes, *Proc. Natl. Acad. Sci. U. S. A.* 102 (2005) 5454–5459.

[185] K. Hanada, C. Zou, M.D. Lehti-Shiu, K. Shinozaki, S.-H. Shiu, Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli, *Plant Physiol.* 148 (2008) 993–1003.

[186] X. Li, X. Duan, H. Jiang, Y. Sun, Y. Tang, Z. Yuan, J. Guo, W. Liang, L. Chen, J. Yin, H. Ma, J. Wang, D. Zhang, Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and *Arabidopsis*, *Plant Physiol.* 141 (2006) 1167–1184.

[187] O. Wilkins, H. Nahal, J. Foong, N.J. Provart, M.M. Campbell, Expansion and diversification of the *Populus R2R3-MYB* family of transcription factors, *Plant Physiol.* 149 (2009) 981–993.

[188] N. Jiang, A.A. Ferguson, R.K. Slotkin, D. Lisch, Pack-Mutator-like transposable elements (Pack-MULEs) induce directional modification of genes through biased insertion and DNA acquisition, *Proc. Natl. Acad. Sci. U. S. A.* 108 (2011) 1537–1542.

[189] C. Du, N. Fefelova, J. Caronna, L. He, H.K. Dooner, The polychromatic Helitron landscape of the maize genome, *Proc. Natl. Acad. Sci. U. S. A.* 106 (2009) 19916–19921.

[190] Z. Li, J. Defoort, S. Tasdighian, S. Maere, Y. Van de Peer, R. De Smet, Gene duplicability of core genes is highly consistent across all angiosperms, *Plant Cell* 28 (2016) 326–344.

[191] X. Guo, Z. Zhang, M.B. Gerstein, D. Zheng, Small RNAs originated from pseudogenes: cis- or trans-acting? *PLoS Comput. Biol.* 5 (2009), e1000449.

[192] C. Zou, M.D. Lehti-Shiu, F. Thibaud-Nissen, T. Prakash, C.R. Buell, S.-H. Shiu, Evolutionary and expression signatures of pseudogenes in *Arabidopsis* and rice, *Plant Physiol.* 151 (2009) 3–15.

[193] J. Nam, J. Kim, S. Lee, G. An, H. Ma, M. Nei, Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms, *Proc. Natl. Acad. Sci. U. S. A.* 101 (2004) 1910–1915.

[194] G.D. Moghe, D.E. Hufnagel, H. Tang, Y. Xiao, I. Dworkin, C.D. Town, J.K. Conner, S.-H. Shiu, Consequences of whole-genome triplication as revealed by comparative genomic analyses of the wild radish *Raphanus raphanistrum* and three other Brassicaceae species, *Plant Cell* 26 (2014) 1925–1937.

[195] H.H. McAdams, A. Arkin, It's a noisy business! Genetic regulation at the nanomolar scale, *Trends Genet.* 15 (1999) 65–69.

[196] H. Kitano, Biological robustness, *Nat. Rev. Genet.* 5 (2004) 826–837.

[197] Z.F. Zhou, B.L. Xiao, G.Y. Zhang, L.Z. Zhuang, A study of the effect of B-EP and naloxone on the function of the hypothalamo-pituitary-testicular axis of the rat, *J. Androl.* 11 (1990) 233–239.

[198] G. de Martino, I. Pan, E. Emmanuel, A. Levy, V.F. Irish, Functional analyses of two tomato APETALA3 genes demonstrate diversification in their roles in regulating floral development, *Plant Cell* 18 (2006) 1833–1845.

[199] K. Hanada, T. Kuromori, F. Myouga, T. Toyoda, W.-H. Li, K. Shinozaki, Evolutionary persistence of functional compensation by duplicate genes in *Arabidopsis*, *Genome Biol. Evol.* 1 (2009) 409–414.

[200] M.A. Nowak, M.C. Boerlijst, J. Cooke, J.M. Smith, Evolution of genetic redundancy, *Nature* 388 (1997) 167–171.

[201] S. Ohno, *Evolution by Gene Duplication*, Springer, 1970.

[202] A.S. Rijkpema, S. Royaert, J. Zethof, G. van der Weerden, T. Gerats, M. Vandebussche, Analysis of the *Petunia* TM6 MADS box gene reveals functional divergence within the DEF/AP3 lineage, *Plant Cell* 18 (2006) 1819–1832.

[203] E.M. Kramer, L. Holappa, B. Gould, M.A. Jaramillo, D. Setnikov, P.M. Santiago, Elaboration of B gene function to include the identity of novel floral organs in the lower eudicot *Aquilegia*, *Plant Cell* 19 (2007) 750–766.

[204] D.A. Rasmussen, E.M. Kramer, E.A. Zimmer, One size fits all? Molecular evidence for a commonly inherited petal identity program in Ranunculales, *Am. J. Bot.* 96 (2009) 96–109.

[205] M. Mondragón-Palomino, G. Theissen, Conserved differential expression of paralogous DEFICIENS- and GLOBOSA-like MADS-box genes in the flowers of Orchidaceae: refining the “orchid code”, *Plant J.* 66 (2011) 1008–1019.

[206] C. Furumizu, J.P. Alvarez, K. Sakakibara, J.L. Bowman, Antagonistic roles for KNOX1 and KNOX2 genes in patterning the land plant body plan following an ancient gene duplication, *PLoS Genet.* 11 (2015), e1004980.

[207] H. Liu, H. Liu, L. Zhou, Z. Zhang, X. Zhang, M. Wang, H. Li, Z. Lin, Parallel domestication of the heading date 1 gene in cereals, *Mol. Biol. Evol.* 32 (2015) 2726–2737.

[208] C.T. Hittinger, S.B. Carroll, Gene duplication and the adaptive evolution of a classic genetic switch, *Nature* 449 (2007) 677–681.

[209] D.L. Des Marais, M.D. Rausher, Escape from adaptive conflict after duplication in an anthocyanin pathway gene, *Nature* 454 (2008) 762–765.

[210] R. Huang, F. Hippauf, D. Rohrbeck, M. Haustein, K. Wenke, J. Feike, N. Sorrelle, B. Piechulla, T.J. Barkman, Enzyme functional evolution through improved catalysis of ancestrally nonpreferred substrates, *Proc. Natl. Acad. Sci. U. S. A.* 109 (2012) 2966–2971.

[211] T. Sikosek, H.S. Chan, E. Bornberg-Bauer, Escape from Adaptive Conflict follows from weak functional trade-offs and mutational robustness, *Proc. Natl. Acad. Sci. U. S. A.* 109 (2012) 14889–14893.

[212] A. Force, M. Lynch, F.B. Pickett, A. Amores, Y.L. Yan, J. Postlethwait, Preservation of duplicate genes by complementary, degenerative mutations, *Genetics* 151 (1999) 1531–1545.

[213] J.A. Birchler, N.C. Riddle, D.L. Auger, R.A. Veitia, Dosage balance in gene regulation: biological implications, *Trends Genet.* 21 (2005) 219–226.

[214] M. Freeling, B.C. Thomas, Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity, *Genome Res.* 16 (2006) 805–814.

[215] J.A. Birchler, R.A. Veitia, The gene balance hypothesis: from classical genetics to modern genomics, *Plant Cell* 19 (2007) 395–402.

[216] J.A. Birchler, R.A. Veitia, The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution, *New Phytol.* 186 (2010) 54–62.

[217] R.A. Veitia, Exploring the etiology of haploinsufficiency, *BioEssays* 24 (2002) 175–184.

[218] B. Papp, C. Pál, L.D. Hurst, Dosage sensitivity and the evolution of gene families in yeast, *Nature* 424 (2003) 194–197.

[219] J.C. Schnable, B.S. Pedersen, S. Subramaniam, M. Freeling, Dose-sensitivity, conserved non-coding sequences and duplicate gene retention through multiple tetraploidies in the grasses, *Front. Plant Sci.* 2 (2011) 2.

[220] E.W. Ganko, B.C. Meyers, T.J. Vision, Divergence in expression between duplicated genes in *Arabidopsis*, *Mol. Biol. Evol.* 24 (2007) 2298–2309.

[221] S. Renny-Byfield, J.P. Gallagher, C.E. Grover, E. Szadkowski, J.T. Page, J.A. Udall, X. Wang, A.H. Paterson, J.F. Wendel, Ancient gene duplicates in *Gossypium* (cotton) exhibit near-complete expression divergence, *Genome Biol. Evol.* 6 (2014) 559–571.

[222] S. Drea, L.C. Hileman, G. de Martino, V.F. Irish, Functional analyses of genetic pathways controlling petal specification in poppy, *Development* 134 (2007) 4157–4166.

[223] S. Sakuma, M. Pourkheirandish, G. Hensel, J. Kumlehn, N. Stein, A. Tagiri, N. Yamaji, J.F. Ma, H. Sassa, T. Koba, T. Komatsuda, Divergence of expression pattern contributed to neofunctionalization of duplicated HD-Zip I transcription factor in barley, *New Phytol.* 197 (2013) 939–948.

[224] M.D. Lehti-Shiu, S. Uygun, G.D. Moghe, N. Panchy, L. Fang, D.E. Hufnagel, H.L. Jasicki, M. Feig, S.-H. Shiu, Molecular evidence for functional divergence and decay of a transcription factor derived from whole-genome duplication in *Arabidopsis thaliana*, *Plant Physiol.* 168 (2015) 1717–1734.

[225] G. Haberer, T. Hindemitt, B.C. Meyers, K.F.X. Mayer, Transcriptional similarities, dissimilarities, and conservation of cis-elements in duplicated genes of *Arabidopsis*, *Plant Physiol.* 136 (2004) 3009–3022.

[226] M. Ha, E.-D. Kim, Z.J. Chen, Duplicate genes increase expression diversity in closely related species and allopolyploids, *Proc. Natl. Acad. Sci. U. S. A.* 106 (2009) 2295–2300.

[227] A.L. Hughes, R. Friedman, Expression patterns of duplicate genes in the developing root in *Arabidopsis thaliana*, *J. Mol. Evol.* 60 (2005) 247–256.

[228] K. Geuten, T. Vlaeene, V.F. Irish, Robustness and evolvability in the B-system of flower development, *Ann. Bot.* 107 (2011) 1545–1556.

[229] M. Vandebussche, J. Zethof, S. Royaert, K. Weterings, T. Gerats, The duplicated B-class heterodimer model: whorl-specific effects and complex genetic interactions in *Petunia hybrida* flower development, *Plant Cell* 16 (2004) 741–754.

[230] B. McGonigle, K. Bouhidel, V.F. Irish, Nuclear localization of the *Arabidopsis* APETALA3 and PISTILLATA homeotic gene products depends on their simultaneous expression, *Genes Dev.* 10 (1996) 1812–1821.

[231] V.F. Irish, Evolution of petal identity, *J. Exp. Bot.* 60 (2009) 2517–2527.

[232] T. Lenser, G. Theissen, P. Dittrich, Developmental robustness by obligate interaction of class B floral homeotic genes and proteins, *PLoS Comput. Biol.* 5 (2009), e1000264.

[233] T.L. Liu, L. Newton, M.-J. Liu, S.-H. Shiu, E.M. Farré, A G-box-like motif is necessary for transcriptional regulation by circadian pseudo-response regulators in *Arabidopsis*, *Plant Physiol.* 170 (2016) 528–539.

[234] M.T. Weirauch, A. Yang, M. Albu, A.G. Cote, A. Montenegro-Montero, P. Drewe, H.S. Nafafabadi, S.A. Lambert, I. Mann, K. Cook, H. Zheng, A. Goity, H. van Bakel, J.-C. Lozano, M. Galli, M.G. Lewsey, E. Huang, T. Mukherjee, X. Chen, J.S. Reece-Hoyes, S. Govindarajan, G. Shaulsky, A.J.M. Walhout, F.-Y. Bouget, G. Ratsch, L.F. Larondo, J.R. Ecker, T.R. Hughes, Determination and inference of eukaryotic transcription factor sequence specificity, *Cell* 158 (2014) 1431–1443.

[235] A.A. Arsovski, J. Pradnik, X.Q. Guo, S. Wang, K.L. Adams, Evolution of Cis-regulatory elements and regulatory networks in duplicated genes of *Arabidopsis*, *Plant Physiol.* 169 (2015) 2982–2991.

[236] S.A. Rensing, Gene duplication as a driver of plant morphogenetic evolution, *Curr. Opin. Plant Biol.* 17 (2014) 43–48.

[237] C. Smaczniak, R.G.H. Immink, G.C. Angenent, K. Kaufmann, Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies, *Development* 139 (2012) 3081–3098.

[238] A. Becker, G. Theissen, The major clades of MADS-box genes and their role in the development and evolution of flowering plants, *Mol. Phylogenet. Evol.* 29 (2003) 464–489.

[239] B.A. Krizek, J.C. Fletcher, Molecular mechanisms of flower development: an armchair guide, *Nat. Rev. Genet.* 6 (2005) 688–698.

[240] G. Theissen, R. Melzer, Molecular mechanisms underlying origin and diversification of the angiosperm flower, *Ann. Bot.* 100 (2007) 603–619.

[241] G. Theissen, Development of floral organ identity: stories from the MADS house, *Curr. Opin. Plant Biol.* 4 (2001) 75–85.

[242] G. Theissen, H. Saedler, Plant biology: floral quartets, *Nature* 409 (2001) 469–471.

[243] R. Melzer, A. Härtel, F. Rümpler, S. Kim, P.S. Soltis, D.E. Soltis, G. Theißen, DEF- and GLO-like proteins may have lost most of their interaction partners during angiosperm evolution, *Ann. Bot.* 114 (2014) 1431–1443.

[244] L. Li, X.-X. Yu, C.-C. Guo, X.-S. Duan, H.-Y. Shan, R. Zhang, G.-X. Xu, H.-Z. Kong, Interactions among proteins of floral MADS-box genes in *Nuphar pumila* (Nymphaeaceae) and the most recent common ancestor of extant angiosperms help understand the underlying mechanisms of the origin of the flower: evolution of PPLs among floral MADS-box genes, *J. Syst. Evol.* 53 (2015) 285–296.

[245] N.D. Pires, L. Dolan, Morphological evolution in land plants: new designs with old genes, *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 367 (2012) 508–518.

[246] J.R. True, S.B. Carroll, Gene co-option in physiological and morphological evolution, *Annu. Rev. Cell Dev. Biol.* 18 (2002) 53–80.

[247] N.D. Pires, K. Yi, H. Breuninger, B. Catarino, B. Menand, L. Dolan, Recruitment and remodeling of an ancient gene regulatory network during land plant evolution, *Proc. Natl. Acad. Sci. U. S. A.* 110 (2013) 9571–9576.

[248] T.H.Y. Tam, B. Catarino, L. Dolan, Conserved regulatory mechanism controls the development of cells with rooting functions in land plants, *Proc. Natl. Acad. Sci. U. S. A.* 112 (2015) E3959–E3968.

[249] T. Lotan, M. Ohto, K.M. Yee, M.A. West, R. Lo, R.W. Kwong, K. Yamagishi, R.L. Fischer, R.B. Goldberg, J.J. Harada, *Arabidopsis* LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells, *Cell* 93 (1998) 1195–1205.

[250] R.C. Kirkbride, R.L. Fischer, J.J. Harada, LEAFY COTYLEDON1, a key regulator of seed development, is expressed in vegetative and sexual propagules of *Selaginella moellendorffii*, *PLoS One* 8 (2013), e67971.

[251] Z. Xie, X. Li, B.J. Glover, S. Bai, G.-Y. Rao, J. Luo, J. Yang, Duplication and functional diversification of HAP3 genes leading to the origin of the seed-developmental regulatory gene, LEAFY COTYLEDON1 (LEC1), in nonseed plant genomes, *Mol. Biol. Evol.* 25 (2008) 1581–1592.

[252] H.M.P. Garcés, C.E.M. Champagne, B.T. Townsley, S. Park, R. Malhó, M.C. Pedroso, J.J. Harada, N.R. Sinha, Evolution of asexual reproduction in leaves of the genus *Kalanchoe*, *Proc. Natl. Acad. Sci. U. S. A.* 104 (2007) 15578–15583.

[253] H.M.P. Garcés, D. Koenig, B.T. Townsley, M. Kim, N.R. Sinha, Truncation of LEAFY COTYLEDON1 protein is required for asexual reproduction in *Kalanchoe daigremontiana*, *Plant Physiol.* 165 (2014) 196–206.

[254] C.J. Harrison, S.B. Corley, E.C. Moylan, D.L. Alexander, R.W. Scotland, J.A. Langdale, Independent recruitment of a conserved developmental mechanism during leaf evolution, *Nature* 434 (2005) 509–514.

[255] A.M.F. Tomescu, Megaphylls, microphylls and the evolution of leaf development, *Trends Plant Sci.* 14 (2009) 5–12.

[256] C. Lincoln, J. Long, J. Yamaguchi, K. Serikawa, S. Hake, A knotted1-like homeobox gene in *Arabidopsis* is expressed in the vegetative meristem and dramatically alters leaf morphology when overexpressed in transgenic plants, *Plant Cell* 6 (1994) 1859–1876.

[257] N.R. Sinha, R.E. Williams, S. Hake, Overexpression of the maize homeo box gene, KNOTTED-1, causes a switch from determinate to indeterminate cell fates, *Genes Dev.* 7 (1993) 787–795.

[258] E. Vollbrecht, B. Veit, N. Sinha, S. Hake, The developmental gene Knotted-1 is a member of a maize homeobox gene family, *Nature* 350 (1991) 241–243.

[259] M.E. Byrne, R. Barley, M. Curtis, J.M. Arroyo, M. Dunham, A. Hudson, R.A. Martienssen, Asymmetric leaves1 mediates leaf patterning and stem cell function in *Arabidopsis*, *Nature* 408 (2000) 967–971.

[260] N. Ori, Y. Eshed, G. Chuck, J.L. Bowman, S. Hake, Mechanisms that control knox gene expression in the *Arabidopsis* shoot, *Development* 127 (2000) 5523–5532.

[261] M. Tsiantis, R. Schneeberger, J.F. Golz, M. Freeling, J.A. Langdale, The maize rough sheath2 gene and leaf development programs in monocot and dicot plants, *Science* 284 (1999) 154–156.

[262] M.C. Timmermans, A. Hudson, P.W. Becroft, T. Nelson, ROUGH SHEATH2: a Myb protein that represses knox homeobox genes in maize lateral organ primordia, *Science* 284 (1999) 151–153.

[263] R. Waites, H.R. Selvadurai, I.R. Oliver, A. Hudson, The PHANTASTICA gene encodes a MYB transcription factor involved in growth and dorsoventrality of lateral organs in *Antirrhinum*, *Cell* 93 (1998) 779–789.

[264] C. Pozzi, L. Rossini, F. Agosti, Patterns and symmetries in leaf development, *Semin. Cell Dev. Biol.* 12 (2001) 363–372.

[265] T.L. Phelps-Durr, J. Thomas, P. Vahab, M.C.P. Timmermans, Maize rough sheath2 and its *Arabidopsis* orthologue ASYMMETRIC LEAVES1 interact with HIRA, a predicted histone chaperone, to maintain knox gene silencing and determinacy during organogenesis, *Plant Cell* 17 (2005) 2886–2898.

[266] W. Zhang, E.M. Kramer, C.C. Davis, Floral symmetry genes and the origin and maintenance of zygomorphy in a plant-pollinator mutualism, *Proc. Natl. Acad. Sci. U. S. A.* 107 (2010) 6388–6393.

[267] R.D. Sargent, Floral symmetry affects speciation rates in angiosperms, *Proc. Biol. Sci.* 271 (2004) 603–608.

[268] S.B. Corley, R. Carpenter, L. Copsey, E. Coen, Floral asymmetry involves an interplay between TCP and MYB transcription factors in *Antirrhinum*, *Proc. Natl. Acad. Sci. U. S. A.* 102 (2005) 5068–5073.

[269] X. Feng, Z. Zhao, Z. Tian, S. Xu, Y. Luo, Z. Cai, Y. Wang, J. Yang, Z. Wang, L. Weng, J. Chen, L. Zheng, X. Guo, J. Luo, S. Sato, S. Tabata, W. Ma, X. Cao, X. Hu, C. Sun, D. Luo, Control of petal shape and floral zygomorphy in *Lotus japonicus*, *Proc. Natl. Acad. Sci. U. S. A.* 103 (2006) 4970–4975.

[270] Z. Wang, Y. Luo, X. Li, L. Wang, S. Xu, J. Yang, L. Weng, S. Sato, S. Tabata, M. Ambrose, C. Ramau, X. Feng, X. Hu, D. Luo, Genetic control of floral zygomorphy in pea (*Pisum sativum* L.), *Proc. Natl. Acad. Sci. U. S. A.* 105 (2008) 10414–10419.

[271] H.L. Citerne, R.T. Pennington, Q.C.B. Cronk, An apparent reversal in floral symmetry in the legume *Cadia* is a homeotic transformation, *Proc. Natl. Acad. Sci. U. S. A.* 103 (2006) 12017–12020.

[272] X. Yang, H.-B. Pang, B.-L. Liu, Z.-J. Qiu, Q. Gao, L. Wei, Y. Dong, Y.-Z. Wang, Evolution of double positive autoregulatory feedback loops in CYCLOIDEA2 clade genes is associated with the origin of floral zygomorphy, *Plant Cell* 24 (2012) 1834–1847.

[273] L.C. Hileman, Trends in flower symmetry evolution revealed through phylogenetic and developmental genetic advances, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369 (2014), <http://dx.doi.org/10.1098/rstb.2013.0348>.

[274] P. Cubas, Floral zygomorphy, the recurring evolution of a successful trait, *BioEssays* 26 (2004) 1175–1184.

[275] A. Hay, M. Tsiantis, The genetic basis for differences in leaf form between *Arabidopsis thaliana* and its wild relative *Cardamine hirsuta*, *Nat. Genet.* 38 (2006) 942–947.

[276] E. Shani, Y. Burko, L. Ben-Yaakov, Y. Berger, Z. Amsellelem, A. Goldshmidt, E. Sharon, N. Ori, Stage-specific regulation of *Solanum lycopersicum* leaf maturation by class 1 KNOTTED1-LIKE HOMEOBOX proteins, *Plant Cell* 21 (2009) 3078–3092.

[277] J. Peng, J. Yu, H. Wang, Y. Guo, G. Li, G. Bai, R. Chen, Regulation of compound leaf development in *Medicago truncatula* by fused compound leaf1, a class M KNOX gene, *Plant Cell* 23 (2011) 3929–3943.

[278] K. Sakakibara, T. Nishiyama, H. Deguchi, M. Hasebe, Class 1 KNOX genes are not involved in shoot development in the moss *Physcomitrella patens* but do function in sporophyte development, *Evol. Dev.* 10 (2008) 555–566.

[279] R. Sano, C.M. Juárez, B. Hass, K. Sakakibara, M. Ito, J.A. Banks, M. Hasebe, KNOX homeobox genes potentially have similar function in both diploid unicellular and multicellular meristems, but not in haploid meristems, *Evol. Dev.* 7 (2005) 69–78.

[280] C. Dardick, A.M. Callahan, Evolution of the fruit endocarp: molecular mechanisms underlying adaptations in seed protection and dispersal strategies, *Front. Plant Sci.* 5 (2014) 284.

[281] C. Fourquin, C. del Cerro, F.C. Victoria, A. Viallette-Guiraud, A.C. de Oliveira, C. Ferrández, A change in SHATTERPROOF protein lies at the origin of a fruit morphological novelty and a new strategy for seed dispersal in *medicago* genus, *Plant Physiol.* 162 (2013) 907–917.

[282] Q. Gu, C. Ferrández, M.F. Yanofsky, R. Martienssen, The FRUITFULL MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development, *Development* 125 (1998) 1509–1517.

[283] A. Litt, V.F. Irish, Duplication and diversification in the APETALA1/FRUITFULL floral homeotic gene lineage: implications for the evolution of floral development, *Genetics* 165 (2003) 821–833.

[284] N. Pabón-Mora, B. Sharma, L.D. Holappa, E.M. Kramer, A. Litt, The Aquilegia FRUITFULL-like genes play key roles in leaf morphogenesis and inflorescence development, *Plant J.* 74 (2013) 197–212.

[285] E.W. McCarthy, A. Mohamed, A. Litt, Functional divergence of APETALA1 and FRUITFULL is due to changes in both regulation and coding sequence, *Front. Plant Sci.* 6 (2015) 1076.

[286] Y. Liu, D. Zhang, J. Ping, S. Li, Z. Chen, J. Ma, Innovation of a regulatory mechanism modulating semi-determinate stem growth through artificial selection in soybean, *PLoS Genet.* 12 (2016), e1005818.

[287] M.T. Weirauch, T.R. Hughes, A catalogue of eukaryotic transcription factor types, their evolutionary origin, and species distribution, *Subcell. Biochem.* 52 (2011) 25–73.

[288] J.-K. Weng, C. Chapple, The origin and evolution of lignin biosynthesis, *New Phytol.* 187 (2010) 273–285.

[289] Y. Nakano, M. Yamaguchi, H. Endo, N.A. Rejab, M. Ohtani, NAC-MYB-based transcriptional regulation of secondary cell wall biosynthesis in land plants, *Front. Plant Sci.* 6 (2015) 288.

[290] B. Xu, M. Ohtani, M. Yamaguchi, K. Toyooka, M. Wakazaki, M. Sato, M. Kubo, Y. Nakano, R. Sano, Y. Hiwatashi, T. Murata, T. Kurata, A. Yoneda, K. Kato, M. Hasebe, T. Demura, Contribution of NAC transcription factors to plant adaptation to land, *Science* 343 (2014) 1505–1508.

[291] K.M. Schmid, J.B. Ohlrogge, Chapter 4 Lipid metabolism in plants, *Biochemistry of Lipids, Lipoproteins and Membranes*, fourth ed. Elsevier 2002, pp. 93–126.

[292] H.-W. Wang, B. Zhang, Y.-J. Hao, J. Huang, A.-G. Tian, Y. Liao, J.-S. Zhang, S.-Y. Chen, The soybean Dof-type transcription factor genes, GmDof4 and GmDof11, enhance lipid content in the seeds of transgenic *Arabidopsis* plants, *Plant J.* 52 (2007) 716–729.

[293] A. Cernac, C. Benning, WRINKLED1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in *Arabidopsis*, *Plant J.* 40 (2004) 575–585.

[294] J. Liu, W. Hua, G. Zhan, F. Wei, X. Wang, G. Liu, H. Wang, Increasing seed mass and oil content in transgenic *Arabidopsis* by the overexpression of *wri1*-like gene from *Brassica napus*, *Plant Physiol. Biochem.* 48 (2010) 9–15.

[295] B. Pouvreau, S. Baud, V. Vernoud, V. Morin, C. Py, G. Gendrot, J.-P. Pichon, J. Rouster, W. Paul, P.M. Rogowsky, Duplicate maize Wrinkled1 transcription factors activate target genes involved in seed oil biosynthesis, *Plant Physiol.* 156 (2011) 674–686.

[296] B. Shen, W.B. Allen, P. Zheng, C. Li, K. Glassman, J. Ranch, D. Nubel, M.C. Tarczynski, Expression of ZmLEC1 and ZmWRI1 increases seed oil production in maize, *Plant Physiol.* 153 (2010) 980–987.

[297] W. Ma, Q. Kong, V. Arondel, A. Kilaru, P.D. Bates, N.A. Thrower, C. Benning, J.B. Ohlrogge, Wrinkled1, a ubiquitous regulator in oil accumulating tissues from *Arabidopsis* embryos to oil palm mesocarp, *PLoS One* 8 (2013), e68887.

[298] A. To, J. Joubès, G. Barthole, A. Lécureuil, A. Scagnelli, S. Jasinski, L. Lepiniec, S. Baud, WRINKLED transcription factors orchestrate tissue-specific regulation of fatty acid biosynthesis in *Arabidopsis*, *Plant Cell* 24 (2012) 5007–5023.

[299] A. Kilaru, X. Cao, P.B. Dabbs, H.-J. Sung, M.M. Rahman, N. Thrower, G. Zynda, R. Podicheti, E. Ibarra-Laclette, L. Herrera-Estrella, K. Mockaitis, J.B. Ohlrogge, Oil biosynthesis in a basal angiosperm: transcriptome analysis of *Persea Americana* mesocarp, *BMC Plant Biol.* 15 (2015) 203.

[300] E. Pichersky, E. Lewinsohn, Convergent evolution in plant specialized metabolism, *Annu. Rev. Plant Biol.* 62 (2011) 549–566.

[301] C.D. Grubb, S. Abel, Glucosinolate metabolism and its control, *Trends Plant Sci.* 11 (2006) 89–100.

[302] M. Bekaert, P.P. Edger, C.M. Hudson, J.C. Pires, G.C. Conant, Metabolic and evolutionary costs of herbivory defense: systems biology of glucosinolate synthesis, *New Phytol.* 196 (2012) 596–605.

[303] P.P. Edger, H.M. Heidel-Fischer, M. Bekaert, J. Rota, G. Glöckner, A.E. Platts, D.G. Heckel, J.P. Der, E.K. Wafula, M. Tang, J.A. Hofberger, A. Smithson, J.C. Hall, M. Blanchette, T.E. Bureau, S.I. Wright, C.W. dePamphilis, M.E. Schranz, M.S. Barker, G.C. Conant, N. Wahlberg, H. Vogel, J.C. Pires, C.W. Wheat, The butterfly plant arms-race escalated by gene and genome duplications, *Proc. Natl. Acad. Sci. U. S. A.* 112 (2015) 8362–8366.

[304] H. Frerigmann, T. Gigolashvili, MYB34, MYB51, and MYB122 distinctly regulate indolic glucosinolate biosynthesis in *Arabidopsis thaliana*, *Mol. Plant* 7 (2014) 814–828.

[305] I.E. Sonderby, M. Burow, H.C. Rowe, D.J. Kliebenstein, B.A. Halkier, A complex interplay of three R2R3 MYB transcription factors determines the profile of aliphatic glucosinolates in *Arabidopsis*, *Plant Physiol.* 153 (2010) 348–363.

[306] M.Y. Hirai, K. Sugiyama, Y. Sawada, T. Tohge, T. Obayashi, A. Suzuki, R. Araki, N. Sakurai, H. Suzuki, K. Aoki, H. Goda, O.I. Nishizawa, D. Shibata, K. Saito, Omics-based identification of *Arabidopsis* Myb transcription factors regulating aliphatic glucosinolate biosynthesis, *Proc. Natl. Acad. Sci. U. S. A.* 104 (2007) 6478–6483.

[307] J.L. Celenza, J.A. Quiel, G.A. Smolen, H. Merrith, A.R. Silvestro, J. Normanly, J. Bender, The *Arabidopsis* ATR1 Myb transcription factor controls indolic glucosinolate homeostasis, *Plant Physiol.* 137 (2005) 253–262.

[308] Y. Zhang, B. Li, D. Huai, Y. Zhou, D.J. Kliebenstein, The conserved transcription factors, MYB115 and MYB118, control expression of the newly evolved benzoyloxy glucosinolate pathway in *Arabidopsis thaliana*, *Front. Plant Sci.* 6 (2015) 343.

[309] S.D. Michaels, R.M. Amasino, FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering, *Plant Cell* 11 (1999) 949–956.

[310] P. Ruelens, R.A. de Maagd, S. Proost, G. Theißen, K. Geuten, K. Kaufmann, FLOWERING LOCUS C in monocots and the tandem origin of angiosperm-specific MADS-box genes, *Nat. Commun.* 4 (2013) 2280.

[311] C. Shindo, M.J. Aranzana, C. Lister, C. Baxter, C. Nicholls, M. Nordborg, C. Dean, Role of FRIGIDA and FLOWERING LOCUS C in determining variation in flowering time of *Arabidopsis*, *Plant Physiol.* 138 (2005) 1163–1173.

[312] A.L. Caicedo, J.R. Stinchcombe, K.M. Olsen, J. Schmitt, M.D. Purugganan, Epistatic interaction between *Arabidopsis* FRI and FLC flowering time genes generates a latitudinal cline in a life history trait, *Proc. Natl. Acad. Sci. U. S. A.* 101 (2004) 15670–15675.

[313] G.C.K. Chiang, D. Barua, E.M. Kramer, R.M. Amasino, K. Donohue, Major flowering time gene, flowering locus C, regulates seed germination in *Arabidopsis thaliana*, *Proc. Natl. Acad. Sci. U. S. A.* 106 (2009) 11661–11666.

[314] J. Putterill, F. Robson, K. Lee, R. Simon, G. Coupland, The CONSTANS gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors, *Cell* 80 (1995) 847–857.

[315] U. Rosas, Y. Mei, Q. Xie, J.A. Banta, R.W. Zhou, G. Seufferheld, S. Gerard, L. Chou, N. Bhambhra, J.D. Parks, J.M. Flowers, C.R. McClung, Y. Hanzawa, M.D. Purugganan, Variation in *Arabidopsis* flowering time associated with cis-regulatory variation in CONSTANS, *Nat. Commun.* 5 (2014) 3651.

[316] S. Simon, M. Rühl, A. de Montaigu, S. Wöltzel, G. Coupland, Evolution of CONSTANS regulation and function after gene duplication produced a photoperiodic flowering switch in the Brassicaceae, *Mol. Biol. Evol.* 32 (2015) 2284–2301.

[317] K.-J. Dietz, M.O. Vogel, A. Viehhauser, AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling, *Protoplasma* 245 (2010) 3–14.

[318] S. Dey, A. Corina Vlot, Ethylene responsive factors in the orchestration of stress responses in monocotyledonous plants, *Front. Plant Sci.* 6 (2015) 640.

[319] N. Guttersen, T.L. Reuber, Regulation of disease resistance pathways by AP2/ERF transcription factors, *Curr. Opin. Plant Biol.* 7 (2004) 465–471.

[320] S. Sun, J.-P. Yu, F. Chen, T.-J. Zhao, X.-H. Fang, Y.-Q. Li, S.-F. Sui, TINY, a dehydration-responsive element (DRE)-binding protein-like transcription factor connecting the DRE- and ethylene-responsive element-mediated signaling pathways in *Arabidopsis*, *J. Biol. Chem.* 283 (2008) 6261–6271.

[321] T. Nakano, K. Suzuki, T. Fujimura, H. Shinshi, Genome-wide analysis of the ERF gene family in *Arabidopsis* and rice, *Plant Physiol.* 140 (2006) 411–432.

[322] T. Fukao, J. Bailey-Serres, Submergence tolerance conferred by *Sub1A* is mediated by *SLR1* and *SLRL1* restriction of gibberellin responses in rice, *Proc. Natl. Acad. Sci. U. S. A.* 105 (2008) 16814–16819.

[323] K. Xu, X. Xu, T. Fukao, P. Canlas, R. Maghirang-Rodriguez, S. Heuer, A.M. Ismail, J. Bailey-Serres, P.C. Ronald, D.J. Mackill, *Sub1A* is an ethylene-response-factor-like gene that confers submergence tolerance to rice, *Nature* 442 (2006) 705–708.

[324] M.F. Thomashow, Molecular basis of plant cold acclimation: insights gained from studying the CBF cold response pathway, *Plant Physiol.* 154 (2010) 571–577.

[325] C. Alonso-Blanco, C. Gomez-Mena, F. Llorente, M. Koornneef, J. Salinas, J.M. Martínez-Zapater, Genetic and molecular analyses of natural variation indicate CBF2 as a candidate gene for underlying a freezing tolerance quantitative trait locus in *Arabidopsis*, *Plant Physiol.* 139 (2005) 1304–1312.

[326] J. Ågren, C.G. Oakley, J.K. McKay, J.T. Lovell, D.W. Schemske, Genetic mapping of adaptation reveals fitness tradeoffs in *Arabidopsis thaliana*, *Proc. Natl. Acad. Sci. U. S. A.* 110 (2013) 21077–21082.

[327] C.G. Oakley, J. Ågren, R.A. Atchison, D.W. Schemske, QTL mapping of freezing tolerance: links to fitness and adaptive trade-offs, *Mol. Ecol.* 23 (2014) 4304–4315.

[328] M.A. Gehan, S. Park, S.J. Gilmour, C. An, C.-M. Lee, M.F. Thomashow, Natural variation in the C-repeat binding factor cold response pathway correlates with local adaptation of *Arabidopsis* ecotypes, *Plant J.* 84 (2015) 682–693.

[329] A.-N. Sieber, C.F.H. Longin, W.L. Leiser, T. Würschum, Copy Number Variation of CBF-A14 at the Fr-A2 Locus Determines Frost Tolerance in Winter Durum Wheat, *Theor. Appl. Genet.* 2016, <http://dx.doi.org/10.1007/s00122-016-2685-3>.

[330] E. Francia, D. Barabaschi, A. Tondelli, G. Laidò, F. Rizza, A.M. Stanca, M. Busconi, C. Fogher, E.J. Stockinger, N. Pecchioni, Fine mapping of a HvCBF gene cluster at the frost resistance locus Fr-H2 in barley, *Theor. Appl. Genet.* 115 (2007) 1083–1091.

[331] K.M. Olsen, J.F. Wendel, Crop plants as models for understanding plant adaptation and diversification, *Front. Plant Sci.* 4 (2013) 290.

[332] J. Doebley, Plant science. Unfallen grains: how ancient farmers turned weeds into crops, *Science* 312 (2006) 1318–1319.

[333] T. Lenser, G. Theißen, Molecular mechanisms involved in convergent crop domestication, *Trends Plant Sci.* 18 (2013) 704–714.

[334] J. Doebley, A. Stec, L. Hubbard, The evolution of apical dominance in maize, *Nature* 386 (1997) 485–488.

[335] A. Studer, Q. Zhao, J. Ross-Ibarra, J. Doebley, Identification of a functional transposon insertion in the maize domestication gene *tb1*, *Nat. Genet.* 43 (2011) 1160–1163.

[336] M.-S. Remigereau, G. Lakis, S. Rekima, M. Leveugle, M.C. Fontaine, T. Langin, A. Sarr, T. Robert, Cereal domestication and evolution of branching: evidence for soft selection in the *Tb1* orthologue of pearl millet (*Pennisetum glaucum* [L.] R. Br.), *PLoS One* 6 (2011), e22404.

[337] H. Wang, A.J. Studer, Q. Zhao, R. Meeley, J.F. Doebley, Evidence that the origin of naked kernels during maize domestication was caused by a single amino acid substitution in *tga1*, *Genetics* 200 (2015) 965–974.

[338] J. Dorweiler, A. Stec, J. Kermicle, J. Doebley, Teosinte glume architecture 1: a genetic locus controlling a key step in maize evolution, *Science* 262 (1993) 233–235.

[339] J. Dorweiler, J. Doebley, Developmental analysis of teosinte glume architecture 1: a key locus in the evolution of maize (Poaceae), *Am. J. Bot.* 84 (1997) 1313.

[340] C. Li, A. Zhou, T. Sang, Rice domestication by reducing shattering, *Science* 311 (2006) 1936–1939.

[341] S. Konishi, T. Izawa, S.Y. Lin, K. Ebana, Y. Fukuta, T. Sasaki, M. Yano, An SNP caused loss of seed shattering during rice domestication, *Science* 312 (2006) 1392–1396.

[342] J.T. Cannon, B.C. Vellutini, J. Smith 3rd, F. Ronquist, U. Jondelius, A. Hejnol, *Xenacoelomorpha* is the sister group to *Nephrozoa*, *Nature* 530 (2016) 89–93.

[343] I. Ebersberger, R. de Matos Simoes, A. Kupczok, M. Guba, E. Kothe, K. Voigt, A. von Haeseler, A consistent phylogenetic backbone for the fungi, *Mol. Biol. Evol.* 29 (2012) 1319–1334.

[344] J.W. Brown, U. Sorhannus, A molecular genetic timescale for the diversification of autotrophic stramenopiles (Ochrophyta): substantive underestimation of putative fossil ages, *PLoS One* 5 (2010), <http://dx.doi.org/10.1371/journal.pone.0012759>.

[345] O.N. Kuvardina, B.S. Leander, V.V. Aleshin, A.P. Myl'nikov, P.J. Keeling, T.G. Simidyanov, The phylogeny of colpodellids (Alveolata) using small subunit rRNA gene sequences suggests they are the free-living sister group to apicomplexans, *J. Eukaryot. Microbiol.* 49 (2002) 498–504.

[346] R.D. Finn, J. Clements, S.R. Eddy, HMMER web server: interactive sequence similarity searching, *Nucleic Acids Res.* 39 (2011) W29–W37.

[347] G. Wu, D.E. Hufnagel, A.K. Denton, S.-H. Shiu, Retained duplicate genes in green alga *Chlamydomonas reinhardtii* tend to be stress responsive and experience frequent response gains, *BMC Genomics* 16 (2015) 149.

[348] D.M. Goodstein, S. Shu, R. Howson, R. Neupane, R.D. Hayes, J. Fazo, T. Mitros, W. Dirks, U. Hellsten, N. Putnam, D.S. Rokhsar, Phytozome: a comparative platform for green plant genomics, *Nucleic Acids Res.* 40 (2012) D1178–D1186.

[349] E. Lyons, B. Pedersen, J. Kane, M. Alam, R. Ming, H. Tang, X. Wang, J. Bowers, A. Paterson, D. Lisch, M. Freeling, Finding and comparing syntenic regions among *Arabidopsis* and the outgroups papaya, poplar, and grape: CoGe with rosids, *Plant Physiol.* 148 (2008) 1772–1781.

[350] E. Lyons, M. Freeling, How to usefully compare homologous plant genes and chromosomes as DNA sequences, *Plant J.* 53 (2008) 661–673.