Contrasted evolutionary trajectories of plant transcription factors

Xuelei Lai, Hicham Chahtane, Raquel Martin-Arevalillo, Chloe Zubieta, François Parcy

Current Opinion in Plant Biology

Contrasted evolutionary trajectories of plant transcription factors

--Manuscript Draft--

Manuscript Number:	COPLBI-D-19-00035R1
Full Title:	Contrasted evolutionary trajectories of plant transcription factors
Article Type:	54: Genome studies and molecular genetics 2020
Short Title:	Plant transcription factor evolution
Keywords:	transcription factors; Evolution; green lineage
Corresponding Author:	Francois Parcy Centre National de la Recherche Scientifique FRANCE
Corresponding Author's Institution:	Centre National de la Recherche Scientifique
Corresponding Author E-Mail:	francois.parcy@cea.fr
First Author:	Francois Parcy
Order of Authors:	Francois Parcy
	Xuelei Lai
	Hicham Chahtane
	Raquel Martin-Arevalillo
	Chloe Zubieta
Abstract:	Because of their prominent roles in plant development, transcription factors (TF) play central roles as drivers of innovation in the evolution of the green lineage (viridiplantae). The advent of massive sequencing combined with comparative genetics/genomics allows a rigorous investigation of how TF families have contributed to plant diversification from charophyte algae to bryophytes to angiosperms. Here, we review recent progress on TF family reconstruction and the identification of distantly related TFs present throughout the evolutionary timeline from algae to angiosperms. These data provide examples of contrasting evolutionary trajectories of TF families and illustrate how conserved TFs adopt diverse roles over the course of evolution.
Author Comments:	

Contrasted evolutionary trajectories of plant transcription factors

Xuelei Lai[1#], Hicham Chahtane[1#], Raquel Martin-Arevalillo[2], Chloe Zubieta[1] and François Parcy[1]

[1] Laboratoire Physiologie Cellulaire et Végétale, Université Grenoble Alpes, CNRS, CEA, INRAE, IRIG-LPCV, 38000 Grenoble, France

[2] Laboratoire de Reproduction et de Développement des Plantes, INRAE, CNRS, ENS de Lyon, UCB Lyon 1, Université de Lyon, France

Corresponding author: Parcy, François (francois.parcy@cea.fr)

[#] These authors contributed equally to this work.

Abstract

Because of their prominent roles in plant development, transcription factors (TF) play central roles as drivers of innovation in the evolution of the green lineage (viridiplantae). The advent of massive sequencing combined with comparative genetics/genomics allows a rigorous investigation of how TF families have contributed to plant diversification from charophyte algae to bryophytes to angiosperms. Here, we review recent progress on TF family reconstruction and the identification of distantly related TFs present throughout the evolutionary timeline from algae to angiosperms. These data provide examples of contrasting evolutionary trajectories of TF families and illustrate how conserved TFs adopt diverse roles over the course of evolution.

Introduction

In plants, as in all other organisms, transcriptional regulation is crucial for most biological processes, from basic metabolism to complex organ development. *Trans*- and *cis*-regulatory factors, known as transcription factors (TFs) and *cis*-regulatory elements (*cis*-elements), respectively, play pivotal roles in this transcriptional regulation. TFs are DNA-binding proteins that bind to specific *cis*-elements and directly regulate the transcription of DNA to mRNA. Because of their importance, the evolution of TFs and their cognate *cis*-elements is tightly linked

to the increase of organismal complexity and morphological innovations that occurred during evolution [1–3] .

With the unprecedented pace of sequencing of genomes and transcriptomes, the repertoire of TFs from a wide variety of plant species (ranging from green algae or bryophytes to angiosperms) is now better characterized [4**, 5**, 6**, 7, 8]. Overall, plant genomes contain a higher number of TFs, more divergent TF families and more unique DNA binding domains (DBD) as compared with genomes from other eukaryotic organisms [9]. Newly obtained sequence information is also useful in constructing more reliable phylogenetic trees, providing solid ground to understand plant evolution [6**]. Today, only the phylogeny of bryophytes (moss + liverworts + hornworts) remains uncertain: bryophytes could be monophyletic or alternatively, they could be paraphyletic with the liverworts + moss clade sister to land plants or to tracheophytes [10, 11, 12*].

Based on more robust phylogenies, how plant evolution was driven by TF family expansion, alteration of binding affinity, novel *cis*-element recognition and diversification of protein partners can be better understood. For example, cross-species comparison of TFs gives clues as to which genetic events (*i.e*. duplications, gene loss) gave rise to alterations in TF families that could be at the source of developmental novelty during evolution.

Evolution of plant TF families

The wealth of recent sequence data has confirmed previous finding as to the origin of some TF, such as the birth of the TCP family in the charophyte lineage [13]. But more importantly, it also sheds new light on the evolution of plant TF genes [6**]. Originally, mainly due to lack of sequence information in charophyte algae, several plant specific TF families (such as LEAFY, NAC, GRAS, MIKC-type MADS and ARF) were thought to have arisen after the water-to-land transition. However, newly available sequences revealed that many of these families were already present in the most recent common ancestor (MRCA) of Streptophyta (Figure 1) [14]. It is likely that very few TF families originated in the MRCA of land plants (such as GeBP) or later after the water-to-land transition (such as VOZ) (Figure 1). Even YABBY TFs, sometimes proposed as land plant specific, are found in most branches of eukaryotic life [7,8,15]. Thus, the origin of a new type of TF family *per se* did not likely play a major role in terrestrialization. Moreover, once plants were able to grow on land, they have largely innovated by building on the

TF families that were already present in the MRCA of land plants. Innovations that occurred all along the streptophyte lineage (including for example stomata, vascular tissues, roots, reproductive cones or flowers) used repeated TF family expansions (through local or genome-wide duplications) and diversification [16] (Figure 1). TF duplications offer the opportunity for changes affecting the biochemical properties of the TF protein itself (such as changes in DNA binding specificity and protein interaction partners) or its expression pattern (allowing the interaction with a novel set of accessible *cis*-elements or available protein partners). Some families experienced strong expansion at different stages of plant evolution. The basic leucine zipper domain (bZIP) expanded in charophytes, while basic helix-loop-helix (bHLH) family members increased in land plants. The MADS TF expanded in spermatophyte and grew from a few members to dozens and sometimes over a hundred members in angiosperms (Figure 1), where they participate in a wide variety of processes including development, immune response, stress response and light and hormonal signaling [17–19].

The new sequence information not only helps to identify the TF family members but also retraces how TFs have diversified their function by the modular assembly of novel protein domains. This is the case for the Auxin Response Factor (ARF) family implicated in transcriptional responses to the auxin phytohormone. In addition to their B3 DBD, ARFs possess a Phox and Bem1 (PB1) domain that, in land plants, mediates ARF interaction with a similar domain present in the auxin/indole acetic acid (Aux/IAA) repressors [20]. In the absence of auxin, ARFs form complexes with Aux/IAA proteins and are transcriptionally inactive. Auxin triggers Aux/IAA degradation, releasing ARFs for transcriptional activation. The auxin pathway originated only in land plants but was assembled using several components present before: i) ARF proteins (possessing both the B3 DBD and PB1 domain) already present in charophytes but with no described role nor link to auxin, and ii) Aux/IAA and auxin receptor both generated through modifications of proteins present in charophytes [5**,21–23]. Similar to several other hormonal pathways at work in land plants, the ARF TF family later expanded to diversify their roles in numerous developmental or signaling cascades. [5**,24–26].

Conservation of TF function over large evolutionary scales

With the advent of genetic analyses in the moss *Physcomitrella patens* (*P. patens*) and the liverwort *Marchantia polymorpha* (*M. polymorpha*), it became possible to compare TF function over the large evolutionary distance that separates early diverging land plants from angiosperms, and to infer TF roles in the MRCA of land plant or understand how their function evolved [27]. For TFs involved in vital processes present in all plants, it can be anticipated that both their biological roles and molecular function (e.g. *cis*-element recognition) could be conserved. It is indeed the case, for example, for PHYTOCHROME INTERACTING FACTOR (PIFs) that are bHLH TFs involved in light signaling. They are found in most plants from charophytes to angiosperms [28*]. Several PIF target genes and even the modulation of PIF protein stability by light via protein-protein interactions with phytochromes are properties conserved between *P. patens* and the flowering plant Arabidopsis [29,30]. This conservation is highlighted by experiments in which the Arabidopsis *pif* quadruple mutant was rescued by the expression of either one of the four PIF orthologs of *P. patens* [31]. In *M. polymorpha*, the unique PIF protein is involved in gemma germination, which corresponds to an asexual mode of propagation, and the light-dependent gametangiophore formation (containing the sexual organs). These developmental responses to light involve a similar PIF-phytochrome module as in Arabidopsis. Thus, this module is conserved at least since plants conquered the land [32,33]. It is even present in streptophytes, however its functionality remains unclear [28*].

The DUO POLLEN 1 (DUO1) MYB TF provides another example where a protein and its biological role were conserved between bryophytes and angiosperms. DUO1 is involved in male gamete development in both *M. polymorpha* and Arabidopsis [34*]. Reciprocal complementation reveals a conserved function of DUO1 between bryophyte and angiosperms in sperm differentiation. Interestingly, whereas the DUO1 from the charophyte algae *Chara braunii* is also functional in *M. polymorpha*, the protein from a zygnematale species is not functional. This is probably due to several substitutions in its DBD, an event that coincides with the loss of sperm mobility in this group [34*].

The functional conservation of TFs in seemingly divergent developmental processes has also been demonstrated in several cases, where the parallel between bryophytes and angiosperms is less obvious. The TCP family of TF, for instance, controls sporophyte branching in angiosperms and in moss *P. patens* [35] although their branching structures (axillary meristems or branch

initials) are very different. Conservation can also occur between the moss gametophyte (haploid tissue) and the angiosperm sporophyte (diploid). For example, the bHLH TFs ROOT HAIR DEFECTIVE 6-LIKE and LOTUS JAPONICUS ROOTHAIRLESS1-LIKE control root hair development in Arabidopsis and an analogous structure (rhizoids) involved in water and nutrient uptake in the gametophyte of *P. patens* [36] or *M. polymorpha* [37]. The molecular mechanisms of reproductive development, involving the bHLH TF BONOBO during the germ cell differentiation process, also appear to be partly conserved among land plants [38]. Finally, the B3 TF ABSCISIC ACID INSENSITIVE 3 (ABI3) involved in abscisic acid (ABA) dependent acquisition of desiccation tolerance and dormancy in seeds of angiosperms such as Arabidopsis and maize is also necessary for ABA induced desiccation tolerance in the gametophyte of *P. patens* [39] and in gemmae dormancy in Marchantia [40]. This provides another example of a presumptive gene network conserved over large evolutionary distance but at work in a different phase of the life cycle.

Long range evolution of floral TFs

TFs important for angiosperm flower development provide interesting examples of how members of an ancestral TF family were able to fulfill new developmental roles over the course of evolution. We describe here two contrasting examples, the MIKC-type MADS TF subfamily and LEAFY (LFY).

With multiple roles linked to angiosperm reproduction (e.g. control of flowering time, flower meristem identity, floral organ identity, ovule and fruit development), the MIKC-type MADS, or type II MADS, illustrates well how the expansion of a TF family has provided members involved in multiple angiosperm specific innovations [19,41]. While an ancient MADS is present in the MRCA of streptophytes and chlorophytes, this TF possesses only the MADS DBD (M domain) and C-terminal domain and is designated a type I MADS TF. The type II or MIKC-type MADS emerged prior to the split between charophytes and land plants and is characterized by the addition of an Intervening (I) and Keratin-like domain (K) (involved in dimer and tetramer formation) to a MADS type I TF present in chlorophytes [19,42]. The acquisition of novel functions by MIKC-type MADS TFs in land plants, in angiosperms in particular, involved several amplifications of the MIKC-type MADS subfamily followed by changes of their biochemical properties- namely changes in their oligomerization state and protein-protein

interaction specificity. These changes progressively allowed the formation of the heterotetramers made of four different proteins that are required, for example, in petal and stamen development in angiosperms [41]. Early on, MADS TFs from chlorophytes (devoid of the K domain necessary for tetramer formation) (Figure 2) or the unique MIKC copy from *M. polymorpha* could only form homodimers [43]. MIKC-type MADS from the gymnosperm *Gnetum gnemon* were reported to form heterotetramers [44], indicating that MADS tetramerization capacity was born sometime between chlorophytes and gymnosperms (Figure 2). This tetramerization capacity became highly promiscuous in the angiosperm-specific SEPALATTA (SEP) MADS clade proteins (also referred to as 'glue' in multiple hetero-tetrameric complexes formation [45]) which are able to recruit up to three different protein partners as opposed to the one or two protein types recruited by type II MADS in gymnosperms (Figure 2). Even later, the gamma genome triplication event that occurred in angiosperms increased the number of so-called "hubs" (such as SEPs), members that are highly connected in TF protein-protein interaction networks [46]. MIKC-type MADS TF also likely diversified their DNA binding properties. For years, it was thought all MADS TFs have the same DNA binding specificity, recognizing a *cis*-element called the CArG box. However, ChIP-seq analyses in Arabidopsis revealed that MADS TFs do not bind the same regulatory regions even if large overlaps sometimes exist. This specificity might be explained by slight differences between the CArG box sequences recognized by each tetramer [47] resulting in the initiation of different developmental programs for each floral organ type. Taken together, the MADS TFs fully employed classical patterns of family expansion and neofunctionalization via gene duplication events. This mechanism is in direct contrast to LFY, an ancient TF that is encoded by a single copy gene in most streptophytes.

LFY plays a key role during flower meristem emergence and fate in flowering plants and is also present in the MRCA of Streptophyta. LFY is a unique case where genetic studies have been performed in a moss, a fern and numerous angiosperms. Studies in *P. patens* and *Ceratopteris richardii* revealed LFY's role in cell division of the moss sporophyte and in the apical growth of gametophyte and sporophyte axes in fern [48,49**]. Studies are lacking to elucidate to what extend LFY control cell division/apical growth in early land plants or even in charophytes. However, it clearly seems that the floral function of LFY in angiosperm was co-opted from an ancestral vegetative role. LFY likely started acquiring a function in reproductive structures of gymnosperms where there is evidence that it controls expression of some MIKC-MADS TF

[50,51] and this role further expanded in angiosperms. The ancestral meristematic function of LFY became less essential in angiosperms but is still obvious in rice (for tiller growth), legumes (in compound leaves) and even Arabidopsis early flower development [50,52,53]. LFY is an outlier to the general theme of TF families diversifying via the addition of protein-protein interaction domains to a core DBD. Even in algae, mosses and ferns, LFY possesses a Sterile Alpha Motif domain, a highly conserved eukaryotic oligomerization domain [54]. The LFY family did not follow a classical expansion scheme and duplications were rarely retained except at the base of liverworts/moss and in gymnosperms, but the reasons for this remain elusive [55]. Interestingly, despite a highly conserved DBD sequence, LFY DNA binding specificity changed several times in evolution [56]. High-throughput methods allowing the determination of TF DNA binding specificity (such as DNA affinity purification sequencing (DAP-seq) [57,58] or systematic evolution of ligands by exponential enrichment followed by sequencing (SELEX-seq) [59]) should thus be useful to thoroughly characterize the DNA binding properties of TF families even in cases where sequence conservation suggest they recognize the same *cis*-elements.

Conclusion

Analyzing how TF evolution has contributed to plant innovation is entering a golden age. The advent of massive genomic data from a wide range of species (1kp/10kp projects [60,61]) allows the resolution of both phylogenies and genome content. The new tools for gene inactivation or gene modification are hitting more and more species although the capacity to easily transform charophytes algae, hornworts, ferns and gymnosperms remains a limitation. Finally, the combination of new sequence datasets, high-throughput *in vitro* genomics such as SELEX-seq or DAP-seq, and computational methods will help speed up the process of determining TF binding sites and potentially target genes [62]. These data will decipher how TF properties (e.g. DNA binding) have evolved and how this evolution has contributed to innovations all along the green lineage.

Acknowledgments

Our work is supported by the French National Agency for Research (FloPiNet ANR-16-CE92-0023-01; UbiFlor ANR-17-CE20-0014; ChromAuxi ANR-18-CE12-0014; GRAL, ANR-10-LABX-49-01 and ANR-17-EURE-0003). We thank R. Dumas for proofreading, L. Verhage for

her input and we apologize to our colleagues whose work could not be cited due to space limitations.

References

1. Das Gupta M, Tsiantis M: **Gene networks and the evolution of plant morphology**. *Current Opinion in Plant Biology* 2018, **45**:82–87.

2. Rebeiz M, Tsiantis M: **Enhancer evolution and the origins of morphological novelty**. *Current Opinion in Genetics & Development* 2017, **45**:115–123.

3. Schmitz JF, Zimmer F, Bornberg-Bauer E: **Mechanisms of transcription factor evolution in Metazoa**. *Nucleic Acids Res* 2016, **44**:6287–6297.

4**. Nishiyama T, Sakayama H, de Vries J, Buschmann H, Saint-Marcoux D, Ullrich KK, Haas FB, Vanderstraeten L, Becker D, Lang D, et al.: **The Chara Genome: Secondary Complexity and Implications for Plant Terrestrialization**. *Cell* 2018, **174**:448-464.e24.

This genomic paper provides an overview of regulators present in *Chara braunii* genome and summarizes the evolutionary novelty enabling terrestrialization.

5**. Bowman JL, Kohchi T, Yamato KT, Jenkins J, Shu S, Ishizaki K, Yamaoka S, Nishihama R, Nakamura Y, Berger F, et al.: **Insights into Land Plant Evolution Garnered from the Marchantia polymorpha Genome**. *Cell* 2017, **171**:287-304.e15.

This paper analyzed the genome of the liverwort *M. polymorpha*, a member of a basal land plant lineage. The data sheds light on the evolution of land plants and presents an overview of the putative hormonal signaling toolkit evolution.

6**. One Thousand Plant Transcriptomes Initiative: **One thousand plant transcriptomes and the phylogenomics of green plants**. *Nature* 2019, **574**:679–685.

A massive sequence data of streptophytic plants from different lineage refine the green lineage phylogeny, refresh the origins of some gene families and discuss the occurrence of whole-genome duplications in plants.

7. Liang Z, Geng Y, Ji C, Du H, Wong CE, Zhang Q, Zhang Y, Zhang P, Riaz A, Chachar S, et al.: ***Mesostigma viride* Genome and Transcriptome Provide Insights into the Origin and Evolution of Streptophyta**. *Adv Sci* 2020, **7**:1901850.

8. Cheng S, Xian W, Fu Y, Marin B, Keller J, Wu T, Sun W, Li X, Xu Y, Zhang Y, et al.: **Genomes of Subaerial Zygnematophyceae Provide Insights into Land Plant Evolution**. *Cell* 2019, **179**:1057-1067.e14.

9. de Mendoza A, Sebe-Pedros A, Sestak MS, Matejcic M, Torruella G, Domazet-Loso T, Ruiz-Trillo I: **Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages**. *Proceedings of the National Academy of Sciences* 2013, **110**:E4858–E4866.

10. Puttick MN, Morris JL, Williams TA, Cox CJ, Edwards D, Kenrick P, Pressel S, Wellman CH, Schneider H, Pisani D, et al.: **The Interrelationships of Land Plants and the Nature of the Ancestral Embryophyte**. *Current Biology* 2018, **28**:733-745.e2.

11. Wickett NJ, Mirarab S, Nguyen N, Warnow T, Carpenter E, Matasci N, Ayyampalayam S, Barker MS, Burleigh JG, Gitzendanner MA, et al.: **Phylotranscriptomic analysis of the origin and early diversification of land plants**. *Proc Natl Acad Sci USA* 2014, **111**:E4859–E4868.

12*.Delaux P-M, Hetherington AJ, Coudert Y, Delwiche C, Dunand C, Gould S, Kenrick P, Li F-W, Philippe H, Rensing SA, et al.: **Reconstructing trait evolution in plant evo–devo studies**. *Current Biology* 2019, **29**:R1110–R1118.

This review paper provides a nice workflow and guidelines to study comparative genomic approaches.

13. Navaud O, Dabos P, Carnus E, Tremousaygue D, Hervé C: **TCP Transcription Factors Predate the Emergence of Land Plants**. *J Mol Evol* 2007, **65**:23–33.

14. Wilhelmsson PKI, Mühlich C, Ullrich KK, Rensing SA: **Comprehensive Genome-Wide Classification Reveals That Many Plant-Specific Transcription Factors Evolved in Streptophyte Algae**. *Genome Biology and Evolution* 2017, **9**:3384–3397.

15. Carmell MA, Dokshin GA, Skaletsky H, Hu Y-C, van Wolfswinkel JC, Igarashi KJ, Bellott DW, Nefedov M, Reddien PW, Enders GC, et al.: **A widely employed germ cell marker is an ancient disordered protein with reproductive functions in diverse eukaryotes**. *eLife* 2016, **5**:e19993.

16. Rosin FM, Kramer EM: **Old dogs, new tricks: Regulatory evolution in conserved genetic modules leads to novel morphologies in plants**. *Developmental Biology* 2009, **332**:25–35.

17. Dröge-Laser W, Snoek BL, Snel B, Weiste C: **The Arabidopsis bZIP transcription factor family — an update**. *Current Opinion in Plant Biology* 2018, **45**:36–49.

18. Feller A, Machemer K, Braun EL, Grotewold E: **Evolutionary and comparative analysis of MYB and bHLH plant transcription factors: Plant MYB and bHLH factors**. *The Plant Journal* 2011, **66**:94–116.

19. Kaufmann K, Melzer R, Theißen G: **MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants**. *Gene* 2005, **347**:183–198.

20. Guilfoyle TJ: **The PB1 Domain in Auxin Response Factor and Aux/IAA Proteins: A Versatile Protein Interaction Module in the Auxin Response**. *Plant Cell* 2015, **27**:33–43.

21. Mutte SK, Kato H, Rothfels C, Melkonian M, Wong GK-S, Weijers D: **Origin and evolution of the nuclear auxin response system**. *eLife* 2018, **7**:e33399.

22. Flores-Sandoval E, Eklund DM, Hong S-F, Alvarez JP, Fisher TJ, Lampugnani ER, Golz JF, Vázquez-Lobo A, Dierschke T, Lin S-S, et al.: **Class C ARFs evolved before the origin of land plants and antagonize differentiation and developmental transitions in *Marchantia polymorpha***. *New Phytol* 2018, **218**:1612–1630.

23. Martin-Arevalillo R, Thévenon E, Jégu F, Vinos-Poyo T, Vernoux T, Parcy F, Dumas R: **Evolution of the Auxin Response Factors from charophyte ancestors**. *PLoS Genet* 2019, **15**:e1008400.

24. Piya S, Shrestha SK, Binder B, Stewart CN, Hewezi T: **Protein-protein interaction and gene co-expression maps of ARFs and Aux/IAAs in Arabidopsis**. *Front Plant Sci* 2014, **5**.

25. Trigg SA, Garza RM, MacWilliams A, Nery JR, Bartlett A, Castanon R, Goubil A, Feeney J, O'Malley R, Huang SSC, et al.: **CrY2H-seq: A massively multiplexed assay for deep-coverage interactome mapping**. *Nature Methods* 2017, **14**:819–825.

26. Vernoux T, Brunoud G, Farcot E, Morin V, Van den Daele H, Legrand J, Oliva M, Das P, Larrieu A, Wells D, et al.: **The auxin signalling network translates dynamic input into robust patterning at the shoot apex**. *Mol Syst Biol* 2011, **7**:508.

27. Bowman JL, Briginshaw LN, Florent SN: **Evolution and co-option of developmental regulatory networks in early land plants**. In *Current Topics in Developmental Biology*. . Elsevier; 2019:35–53.

28*. Han X, Chang X, Zhang Z, Chen H, He H, Zhong B, Deng XW: **Origin and Evolution of Core Components Responsible for Monitoring Light Environment Changes during Plant Terrestrialization**. *Molecular Plant* 2019, **12**:847–862.

This paper summarizes the genes orthologs involved in light signaling and shed light on the origin and evolution of several key players such as UVR8 or the module SPA-COP1.

29. Possart A, Xu T, Paik I, Hanke S, Keim S, Hermann H-M, Wolf L, Hiß M, Becker C, Huq E, et al.: **Characterization of Phytochrome Interacting Factors from the Moss *Physcomitrella patens* Illustrates Conservation of Phytochrome Signaling Modules in Land Plants**. *Plant Cell* 2017, **29**:310–330.

30. Pham VN, Kathare PK, Huq E: **Phytochromes and Phytochrome Interacting Factors**. *Plant Physiol* 2018, **176**:1025–1038.

31. Xu T, Hiltbrunner A: **PHYTOCHROME INTERACTING FACTORs from *Physcomitrella patens* are active in *Arabidopsis* and complement the *pif* quadruple mutant**. *Plant Signaling & Behavior* 2017, **12**:e1388975.

32. Inoue K, Nishihama R, Kataoka H, Hosaka M, Manabe R, Nomoto M, Tada Y, Ishizaki K, Kohchi T: **Phytochrome Signaling Is Mediated by PHYTOCHROME INTERACTING FACTOR in the Liverwort *Marchantia polymorpha***. *Plant Cell* 2016, **28**:1406–1421.

33. Inoue K, Nishihama R, Araki T, Kohchi T: **Reproductive Induction is a Far-Red High Irradiance Response that is Mediated by Phytochrome and PHYTOCHROME INTERACTING FACTOR in Marchantia polymorpha**. *Plant and Cell Physiology* 2019, **60**:1136–1145.

34*. Higo A, Kawashima T, Borg M, Zhao M, López-Vidriero I, Sakayama H, Montgomery SA, Sekimoto H, Hackenberg D, Shimamura M, et al.: **Transcription factor DUO1 generated by neo-functionalization is associated with evolution of sperm differentiation in plants**. *Nat Commun* 2018, **9**:5283.

This paper reconstructed the evolutionary scenario of DUO1 based on biochemistry and multidirectional complementation experiments. This allow the authors to understand some gain and loss of functions related to gametogenesis and shed light on sperm cell mobility trait.

35. Ortiz-Ramírez C, Hernandez-Coronado M, Thamm A, Catarino B, Wang M, Dolan L, Feijó JA, Becker JD: **A Transcriptome Atlas of Physcomitrella patens Provides Insights into the Evolution and Development of Land Plants**. *Molecular Plant* 2016, **9**:205–220.

36. Tam THY, Catarino B, Dolan L: **Conserved regulatory mechanism controls the development of cells with rooting functions in land plants**. *Proc Natl Acad Sci USA* 2015, **112**:E3959–E3968.

37. Proust H, Honkanen S, Jones VAS, Morieri G, Prescott H, Kelly S, Ishizaki K, Kohchi T, Dolan L: **RSL Class I Genes Controlled the Development of Epidermal Structures in the Common Ancestor of Land Plants**. *Current Biology* 2016, **26**:93–99.

38. Yamaoka S, Nishihama R, Yoshitake Y, Ishida S, Inoue K, Saito M, Okahashi K, Bao H, Nishida H, Yamaguchi K, et al.: **Generative Cell Specification Requires Transcription Factors Evolutionarily Conserved in Land Plants**. *Current Biology* 2018, **28**:479-486.e5.

39. Khandelwal A, Cho SH, Marella H, Sakata Y, Perroud P-F, Pan A, Quatrano RS: **Role of ABA and ABI3 in Desiccation Tolerance**. *Science* 2010, **327**:546–546.

40. Eklund DM, Kanei M, Flores-Sandoval E, Ishizaki K, Nishihama R, Kohchi T, Lagercrantz U, Bhalerao RP, Sakata Y, Bowman JL: **An Evolutionarily Conserved Abscisic Acid Signaling Pathway Regulates Dormancy in the Liverwort Marchantia polymorpha**. *Current Biology* 2018, **28**:3691-3699.e3.

41. Theißen G, Melzer R, Rümpler F: **MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution**. *Development* 2016, **143**:3259–3271.

42. Tanabe Y, Hasebe M, Sekimoto H, Nishiyama T, Kitani M, Henschel K, Munster T, Theissen G, Nozaki H, Ito M: **Characterization of MADS-box genes in charophycean green algae and its implication for the evolution of MADS-box genes**. *Proceedings of the National Academy of Sciences* 2005, **102**:2436–2441.

43. Zobell O, Faigl W, Saedler H, Munster T: **MIKC* MADS-Box Proteins: Conserved Regulators of the Gametophytic Generation of Land Plants**. *Molecular Biology and Evolution* 2010, **27**:1201–1211.

44. Wang Y-Q, Melzer R, Theißen G: **Molecular interactions of orthologues of floral homeotic proteins from the gymnosperm Gnetum gnemon provide a clue to the evolutionary origin of 'floral quartets': Quartet formation of gymnosperm MADS proteins**. *The Plant Journal* 2010, **64**:177–190.

45. Immink RG, Tonaco IA, de Folter S, Shchennikova A, van Dijk AD, Busscher-Lange J, Borst JW, Angenent GC: **SEPALLATA3: the "glue" for MADS box transcription factor complex formation**. *Genome Biology* 2009, **10**:R24.

46. Zhang Z, Coenen H, Ruelens P, Hazarika RR, Al Hindi T, Oguis GK, Vandeperre A, van Noort V, Geuten K: **Resurrected Protein Interaction Networks Reveal the Innovation Potential of Ancient Whole Genome Duplication**. *Plant Cell* 2018, doi:10.1105/tpc.18.00409.

47. Smaczniak C, Muiño JM, Chen D, Angenent GC, Kaufmann K: **Differences in DNA-binding specificity of floral homeotic protein complexes predict organ-specific target genes**. *The Plant Cell* 2017, doi:10.1105/tpc.17.00145.

48. Tanahashi T, Sumikawa N, Kato M, Hasebe M: **Diversification of gene function: homologs of the floral regulator FLO/LFY control the first zygotic cell division in the moss Physcomitrella patens**. *Development* 2005, **132**:1727–1736.

49**. Plackett AR, Conway SJ, Hazelton KDH, Rabbinowitsch EH, Langdale JA: **LEAFY maintains apical stem cell activity during shoot development in the fern**. *eLife* 2018, **7**:e39625.

This paper revealed a role for LEAFY in maintaining cell division in the apical stem cells of both haploid and diploid phases of the lifecycle in the fern *Ceratopteris richardii*. The study supports an evolutionary trajectory in which an ancestral LEAFY module that promotes cell proliferation was progressively co-opted, adapted and specialized as novel shoot developmental contexts emerged.

50. Moyroud E, Kusters E, Monniaux M, Koes R, Parcy F: **LEAFY blossoms**. *Trends in Plant Science* 2010, **15**:346–352.

51. Moyroud E, Monniaux M, Thévenon E, Dumas R, Scutt CP, Frohlich MW, Parcy F: **A link between LEAFY and B-gene homologues in Welwitschia mirabilis sheds light on ancestral mechanisms prefiguring floral development**. *New Phytologist* 2017, doi:10.1111/nph.14483.

52. Chahtane H, Vachon G, Le Masson M, Thévenon E, Périgon S, Mihajlovic N, Kalinina A, Michard R, Moyroud E, Monniaux M, et al.: **A variant of LEAFY reveals its capacity to stimulate meristem development by inducing RAX1**. *Plant Journal* 2013, **74**:678–689.

53. Yamaguchi N, Wu MF, Winter CM, Berns MC, Nole-Wilson S, Yamaguchi A, Coupland G, Krizek BA, Wagner D: **A Molecular Framework for Auxin-Mediated Initiation of Flower Primordia**. *Developmental Cell* 2013, **24**:271–282.

54. Sayou C: **A SAM oligomerization domain shapes the genomic binding landscape of the LEAFY transcription factor**. *Nature Communications* 2016,

55. Gao B, Chen M, Li X, Zhang J: **Ancient duplications and grass-specific transposition influenced the evolution of LEAFY transcription factor genes**. *Commun Biol* 2019, **2**:237.

56. Sayou C, Monniaux M, Nanao MH, Moyroud E, Brockington SF, Thévenon E, Chahtane H, Warthmann N, Melkonian M, Zhang Y, et al.: **A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity**. *Science* 2014, **343**:645–648.

57. Bartlett A, O'Malley RC, Huang SC, Galli M, Nery JR, Gallavotti A, Ecker JR: **Mapping genome-wide transcription-factor binding sites using DAP-seq**. *Nature Protocols* 2017, **12**:1659–1672.

58. O'Malley RC, Huang S shan C, Song L, Lewsey MG, Bartlett A, Nery JR, Galli M, Gallavotti A, Ecker JR: **Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape**. *Cell* 2016, **166**:1598.

59. Jolma A, Kivioja T, Toivonen J, Cheng L, Wei G, Enge M, Taipale M, Vaquerizas JM, Yan J, Sillanpää MJ, et al.: **Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities**. *Genome Research* 2010, **20**:861–873.

60. Cheng S, Melkonian M, Smith SA, Brockington S, Archibald JM, Delaux P-M, Li F-W, Melkonian B, Mavrodiev EV, Sun W, et al.: **10KP: A phylodiverse genome sequencing plan**. *GigaScience* 2018, **7**.

61. Matasci N, Hung L-H, Yan Z, Carpenter EJ, Wickett NJ, Mirarab S, Nguyen N, Warnow T, Ayyampalayam S, Barker M, et al.: **Data access for the 1,000 Plants (1KP) project**. *GigaSci* 2014, **3**:17.

62. Lai X, Stigliani A, Vachon G, Carles C, Smaczniak C, Zubieta C, Kaufmann K, Parcy F: **Building Transcription Factor Binding Site Models to Understand Gene Regulation in Plants**. *Molecular Plant* 2019, **12**:743–763.

Figures and legends

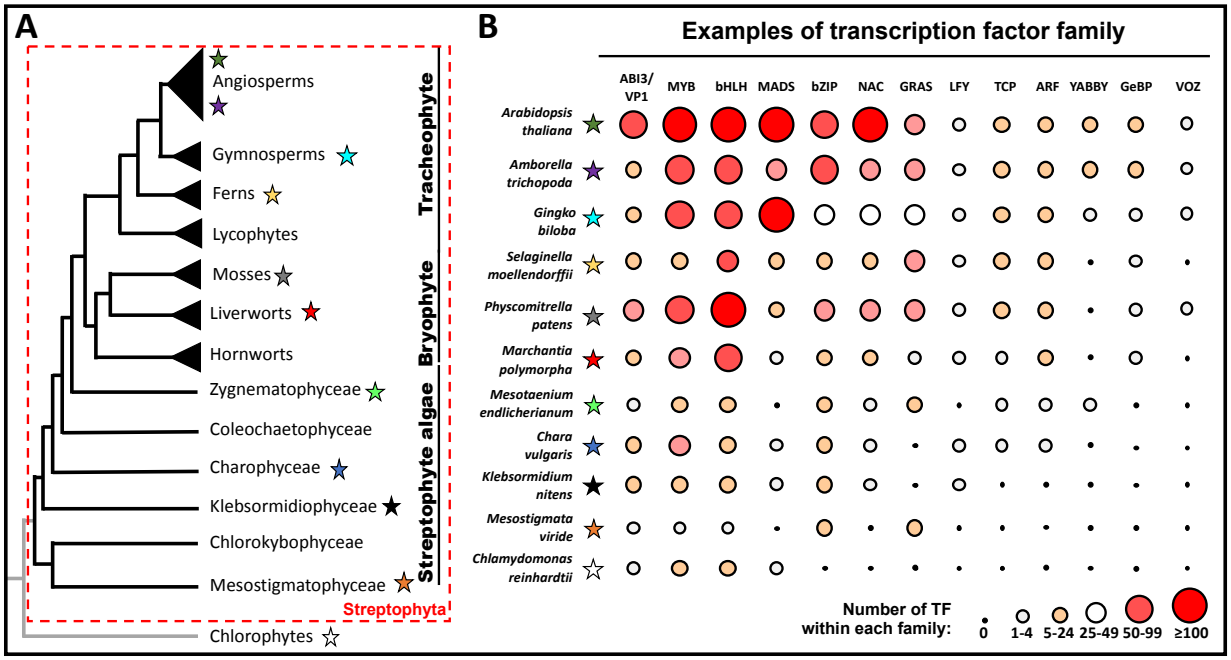

Figure 1: Expansion of several TF families in the green lineage.

(A) Simplified phylogenetic tree representing the Streptophyta lineage. The streptophyta algae is composed by six lineages, where the zygnematophyceae was identified as the sister lineage of land plants. The bryophyta lineage, represented here as monophyletic, comprises mosses, liverworts and hornworts. The branching order within the bryophyta is still debated. The tracheophyta lineage is monophyletic and regroups lycophytes, ferns, and the spermatophyta lineage (gymnosperms and angiosperms). The chlorophyta lineage, the sister group of Streptophyta, is also represented. Coloured stars present the lineage of the species used in panel B.

(B) Examples of the TF toolkit present in several plant genomes. The number of genes for each TF family is represented as coloured circles with the legend given below the figure. The gene abundances are taken from [4**,5**,7,8] .

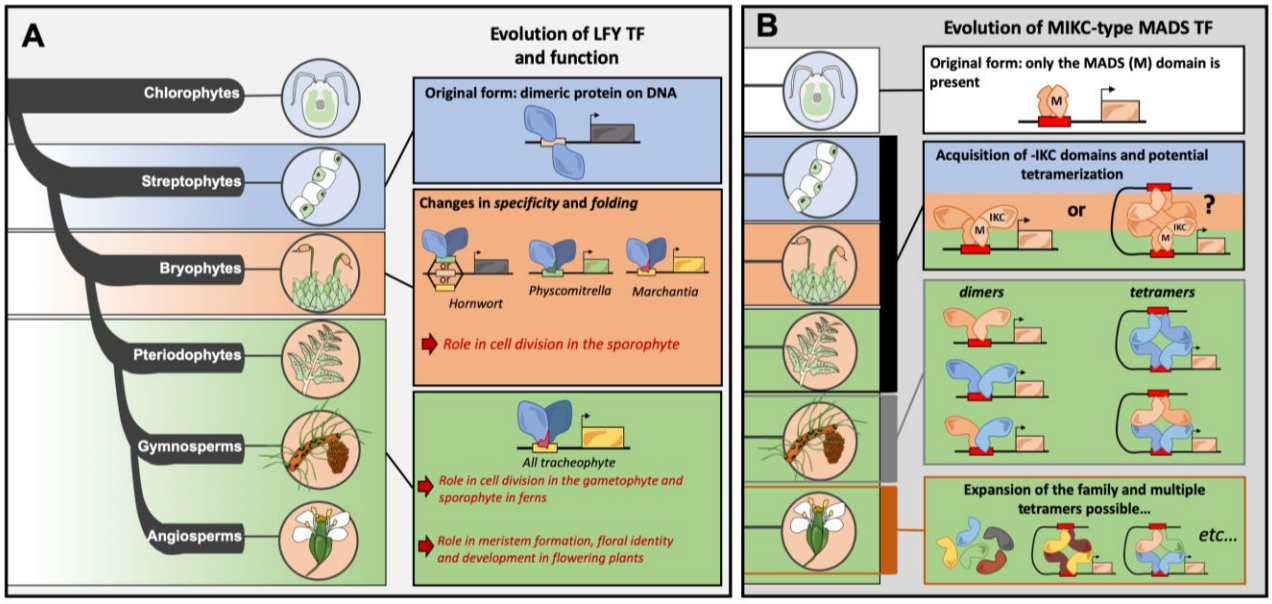

Figure 2: Evolution of TF proteins involved in floral development

(A) Biochemical and functional evolutions of LFY in the green lineage. The simplified phylogeny of Streptophyta, as well as the sister lineage chlorophyta, are represented at left. *LFY* appears first in some streptophyte algae genomes. LFY acts as a dimer on DNA and recognizes a specific *cis*-element depicted schematically as a pink box. In the bryophyta lineage, several modifications affecting the DBD of LFY results in different DNA binding preferences. In *P. patens*, LFY recognizes a different DNA motif show as a green box, whereas in the close ortholog in *M. polymorpha* LFY recognizes a third type

of DNA motif (yellow box). Interestingly, a promiscuous LFY was discovered in hornworts, where LFY binds all three of these DNA motifs. Functional analysis in *P. patens* reveals a key role of LFY in cell division, resulting in the formation of the sporophyte. The LFY specificity primarily found in *M. polymorpha* was kept in all tracheophytes. Functional analyses in ferns revealed the conserved function of LFY in the control of cell division, both in gametophytic and sporophytic tissues. In angiosperms, LFY acts as a master regulator of flower development. Non-floral functions were also reported, such as meristem initiation in Arabidopsis or leaf development in several fabaceae, which may represent its ancestral role reported in ferns and mosses. These functions are depicted schematically.

(B) Biochemical evolution of MICK-type MADS in the green lineage. The simplified phylogeny of streptophyta is represented as pictograms on the left (see panel A for details). MADS TFs appear early in plant evolution and are already present in chlorophyte, the sister lineage of streptophyta. The *cis*-element recognized by MADS TF, the CArG-box, is represented as a red box. MIKC-type MADS TFs appear first in charophytes, where the I, K and C domains are fused to the pre-existing MADS (M) domain. Thus, the origin of the tetramerization property of MIKC-type MADS could date back to the charophyte lineage. In gymnosperms, MIKC-type MADS TFs formed dimers (homodimers and heterodimers) as well as homo and heterotetramers based on biochemical experiments. In angiosperms, the number of different MIKC-type MADS expanded due to several duplications (illustrated by the different protein colours), resulting in a myriad of potential heteromeric combinations. Compared to LFY (panel A), MIKC-type MADS TFs rely less on amino acid substitutions in their DBD. Rather, they acquired a tetramerization interface (K domain) early during evolution and they expanded dramatically in the angiosperm lineage.