

- for AMSR-E precipitation shown in Fig. 1F (red curve) have ~40 degrees of freedom.
18. P. J. Webster, J. L. Keller, *J. Atmos. Sci.* **32**, 1283–1301 (1975).
 19. J. W. Kidson, *Mon. Weather Rev.* **114**, 1654–1663 (1986).
 20. G. K. Vallis, *Atmospheric and Oceanic Fluid Dynamics* (Cambridge Univ. Press, Cambridge, 2006).
 21. R. S. Lindzen, B. F. Farrell, *J. Atmos. Sci.* **37**, 1648–1654 (1980).
 22. B. J. Hoskins, P. J. Valdes, *J. Atmos. Sci.* **47**, 1854–1864 (1990).
 23. P. J. Kushner, I. M. Held, *Geophys. Res. Lett.* **25**, 4213–4216 (1998).
 24. H. Nakamura, A. Shimpo, *J. Clim.* **17**, 1828–1844 (2004).
 25. D. W. J. Thompson, T. Birner, *J. Atmos. Sci.* **69**, 1811–1823 (2012).
 26. The baroclinicity is defined as the meridional slope of the isentropic surfaces relative to Earth's surface. It thus includes two components: the meridional gradient in temperature and the vertical gradient in temperature (i.e., the static stability).
 27. D. G. Andrews, M. E. McIntyre, *J. Atmos. Sci.* **33**, 2031–2048 (1976).
 28. The Geophysical Fluid Dynamics Laboratory (GFDL) CM3 is a coupled chemistry climate model. The analyses shown here are based on a simulation performed for the Coupled Model Intercomparison Project (CMIP5), phase 5. See (33). We analyzed one ensemble of the historical simulation over years 1970–2004. The aquaplanet simulation is identical to that used in (34). The model is integrated at T85 resolution for 21 years, with the first 2 years discarded for spin-up. All model parameters are set to the values from the control experiment of Frierson *et al.* (34). The idealized dry GCM is the GFDL spectral dry dynamical core of (35). The model does not include radiation or moisture and uses Newtonian relaxation to a zonally symmetric equilibrium temperature profile and Rayleigh damping of the low-level winds to simulate boundary layer friction. The model is integrated at T42 (42-wave triangular truncation) resolution for 6000 days with 2000 days for spin-up. Both the aquaplanet and the dry dynamical core do not have topography and are run under perpetual equinoctial conditions with no diurnal cycle.
 29. C.-G. Rossby, H. C. Willett, *Science* **108**, 643–652 (1948).
 30. J. Namias, *J. Meteorol.* **7**, 130–139 (1950).
 31. J. M. Wallace, H.-H. Hsu, *Tellus* **37A**, 478–486 (1985).
 32. D. P. Dee *et al.*, *Q. J. R. Meteorol. Soc.* **137**, 553–597 (2011).
 33. L. J. Donner *et al.*, *J. Clim.* **24**, 3484–3519 (2011).
 34. D. M. W. Frierson, I. M. Held, P. Zurita-Gotor, *J. Atmos. Sci.* **63**, 2548–2566 (2006).
 35. I. M. Held, M. J. Suarez, *Bull. Am. Meteorol. Soc.* **75**, 1825–1830 (1994).

Acknowledgments: We thank R. Garreaud and J. M. Wallace for helpful discussion of the results, R. Barnes for insight into the analytic model, S. Wills for assistance with the AMSR-E data, and three anonymous reviewers for helpful comments on the manuscript. D.W.J.T. is supported by the NSF Climate Dynamics program.

Supplementary Materials

www.sciencemag.org/content/343/6171/641/suppl/DC1
Supplementary Text
Figs. S1 and S2
Reference

25 October 2013; accepted 7 January 2014
10.1126/science.1247660

A Promiscuous Intermediate Underlies the Evolution of LEAFY DNA Binding Specificity

Camille Sayou,^{1,2,3,4*} Marie Monniaux,^{1,2,3,4*} Max H. Nanao,^{5,6*†} Edwige Moyroud,^{1,2,3,4*‡} Samuel F. Brockington,⁷ Emmanuel Thévenon,^{1,2,3,4} Hicham Chahtane,^{1,2,3,4} Norman Warthmann,^{8§} Michael Melkonian,⁹ Yong Zhang,¹⁰ Gane Ka-Shu Wong,^{10,11} Detlef Weigel,⁸ François Parcy,^{1,2,3,4,12†} Renaud Dumas^{1,2,3,4}

Transcription factors (TFs) are key players in evolution. Changes affecting their function can yield novel life forms but may also have deleterious effects. Consequently, gene duplication events that release one gene copy from selective pressure are thought to be the common mechanism by which TFs acquire new activities. Here, we show that LEAFY, a major regulator of flower development and cell division in land plants, underwent changes to its DNA binding specificity, even though plant genomes generally contain a single copy of the *LEAFY* gene. We examined how these changes occurred at the structural level and identify an intermediate LEAFY form in hornworts that appears to adopt all different specificities. This promiscuous intermediate could have smoothed the evolutionary transitions, thereby allowing LEAFY to evolve new binding specificities while remaining a single-copy gene.

The rewiring of transcriptional networks is an important source of evolutionary novelty (1–3). Variation often occurs through changes in cis-regulatory elements, which are DNA sequences that contain binding sites for transcription factors (TFs) regulating nearby genes (3, 4). There is less evidence for regulatory changes affecting the protein-coding sequence of TFs. Such changes are expected to be under highly stringent selection because they could impair the expression of many downstream targets. Gene duplication provides a solution to this dilemma, as additional TF gene copies may acquire new functions, provided that the aggregate copies fulfill the function of the original TF (5). Indeed, TF DNA binding specificity has been shown to diversify within multigene families (6, 7). In some cases, however, TF coding genes remain as single-copy

genes because of phenomena such as paralog interference (8), which can impede neofunctionalization. When essential TFs are maintained as single-copy genes, the extent to which they can evolve is not clear. To address this question, we examined the *LEAFY* (*LFY*) gene as an evolutionary model.

Except in gymnosperms, in which two paralogs (*LEAFY* and *NEEDLY*) are usually present (Fig. 1A), *LFY* exists mostly as a single-copy gene in land plants (9). *LFY* plays essential roles as a key regulator of floral identity in angiosperms, as well as in cell division in the moss *Physcomitrella patens* (10). *LFY* encodes a TF that binds DNA through a highly conserved dimeric DNA binding domain (DBD) (11). Despite this conservation, PpLFY1, a *LFY* homolog from the moss *P. patens*, is unable to bind the DNA sequence recognized

by *LFY* from *Arabidopsis thaliana* (AtLFY) (9), suggesting that *LFY* DNA binding specificity might have changed during land plant evolution.

We mined the transcriptomes from algal species, whose origin predates the divergence of mosses and tracheophytes, and found *LFY* homologs in six species of streptophyte green algae (Fig. 1A and fig. S1) (see also supplementary materials and methods). Thus, *LFY* is not specific to land plants. Despite this extended ancestry, the *LFY*-DBD sequence, including the amino acids in direct contact with DNA, remains highly conserved (Fig. 1B and fig. S1). We used high-throughput SELEX (systematic evolution of ligands by exponential enrichment) (12) experiments to systematically analyze the DNA binding specificity of *LFY* proteins from each group of plants. After

¹CNRS, Laboratoire de Physiologie Cellulaire et Végétale (LPCV), UMR 5168, 38054 Grenoble, France. ²Université Grenoble Alpes, LPCV, F-38054 Grenoble, France. ³Commissariat à l'Énergie Atomique et aux Énergies Alternatives, Direction des Sciences du Vivant, Institut de Recherches en Technologies et Sciences pour le Vivant, LPCV, F-38054 Grenoble, France. ⁴Institut National de la Recherche Agronomique, LPCV, F-38054 Grenoble, France. ⁵European Molecular Biology Laboratory (EMBL), 6 Rue Jules Horowitz, BP 181, 38042 Grenoble, France. ⁶Unit of Virus Host-Cell Interactions, Université Grenoble Alpes–EMBL–CNRS, UMI 3265, 6 Rue Jules Horowitz, 38042 Grenoble Cedex 9, France. ⁷Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK. ⁸Department of Molecular Biology, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany. ⁹Botanisches Institut, Lehrstuhl I, Universität zu Köln, Biozentrum Köln, Zùlpicher Strasse 47b, 50674 Köln, Germany. ¹⁰Beijing Genomics Institute (BGI)–Shenzhen, Beishan Industrial Zone, Yantian District, Shenzhen 518083, China. ¹¹Department of Biological Sciences, Department of Medicine, University of Alberta, Edmonton, Alberta T6G 2E9, Canada. ¹²Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, University of British Columbia, Vancouver, British Columbia V5Z 4H4, Canada.

*These authors contributed equally to this work.

†Corresponding author. E-mail: francois.parcy@cea.fr (F.P.); mnanao@embl.fr (M.H.N.)

‡Present address: Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK.

§Present address: Research School of Biology, The Australian National University, Acton, ACT 0200, Australia.

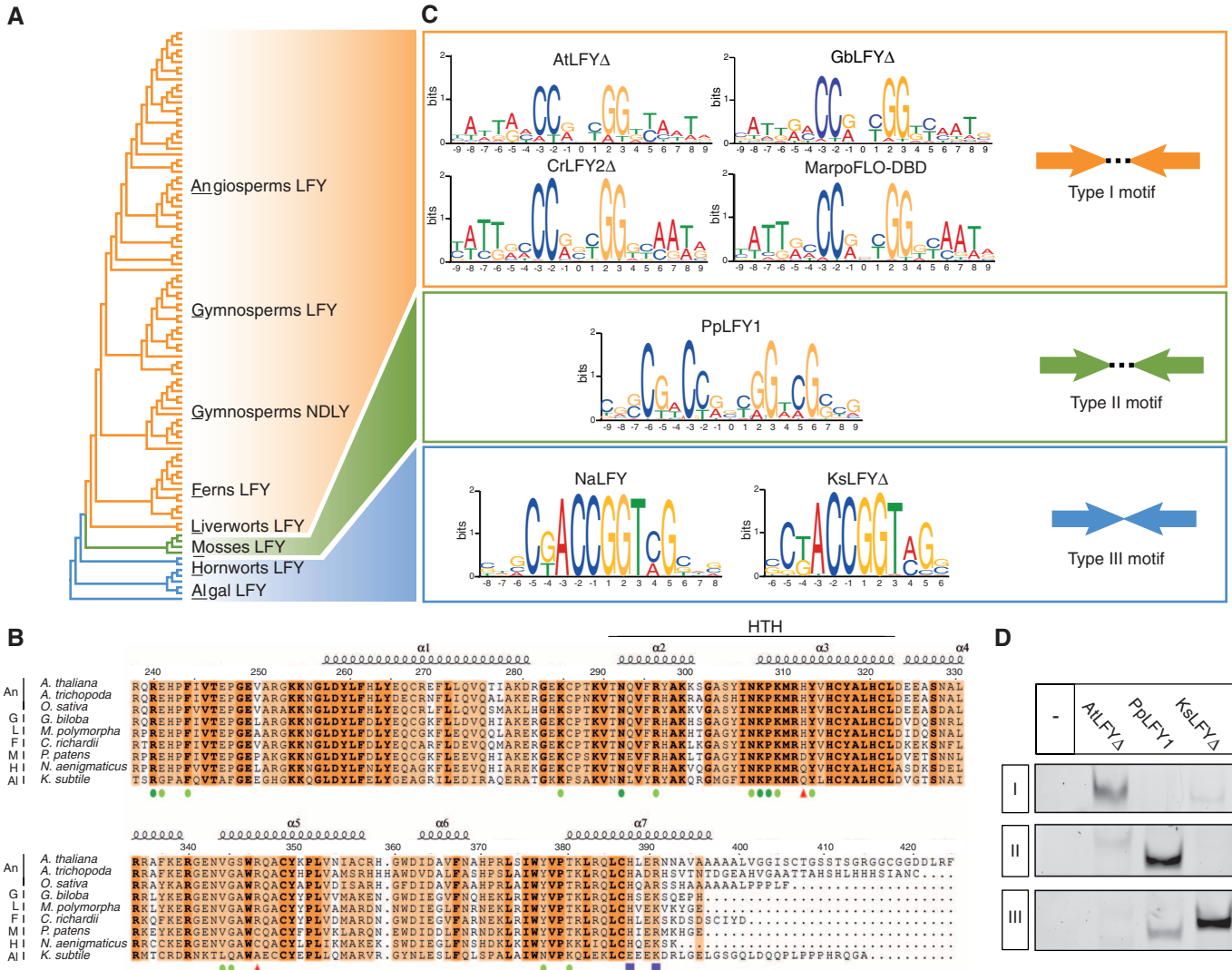


Fig. 1. Evolution of LFY DNA binding specificity. (A) Simplified *LEAFY* phylogeny (detailed in fig. S5). DNA binding specificities are color coded: type I, orange; type II, green; or type III, blue. (B) Alignment of LFY-DBDs. Amino acid numbering and secondary structure annotation (α , alpha helices; HTH, helix-turn-helix domain) are based on AtLFY from *A. thaliana*. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. Dark green dots, DNA base contacts; light green dots, phosphate backbone contacts; red triangles, resi-

dues involved in the PpLFY1-specific DNA contacts; purple rectangles, residues involved in the interaction between DBD monomers. (C) SELEX motifs for AtLFY Δ , GbLFY Δ (*Ginkgo biloba*), CrLFY2 Δ (*Ceratopteris richardii*), MarpoFLO-DBD (*Marchantia polymorpha*), PpLFY1 (*P. patens*), NaLFY (*N. aenigmaticus*), and KsLFY Δ (*K. subtilis*) are shown. Δ denotes proteins starting at amino acid 40 (on the basis of the AtLFY sequence). Cartoons at right depict binding site organization: half-site (arrows) with or without a 3-bp spacer. (D) EMSA with AtLFY Δ , PpLFY1, and KsLFY Δ proteins (10 nM) and the three types (I, II, III) of DNA probes. Only the protein-DNA complexes are shown.

optimizing alignments (13), we found that the SELEX motifs fell into three groups (Fig. 1C and fig. S2), suggesting that LFY changed specificity at least twice.

Most LFY proteins from land plants (angiosperms, gymnosperms, ferns, and liverworts) bind the same DNA motif (type I) as AtLFY (13). PpLFY1, however, binds to a different motif (type II), despite possessing the same 15 DNA binding amino acids as AtLFY (Fig. 1B). These SELEX results explain why all embryophyte LFY homologs, except PpLFY1, display AtLFY-like activity when expressed in *A. thaliana* (9). Motifs I and II share a similar overall organization, consisting of two 8-base pair (bp) inverted half-sites separated

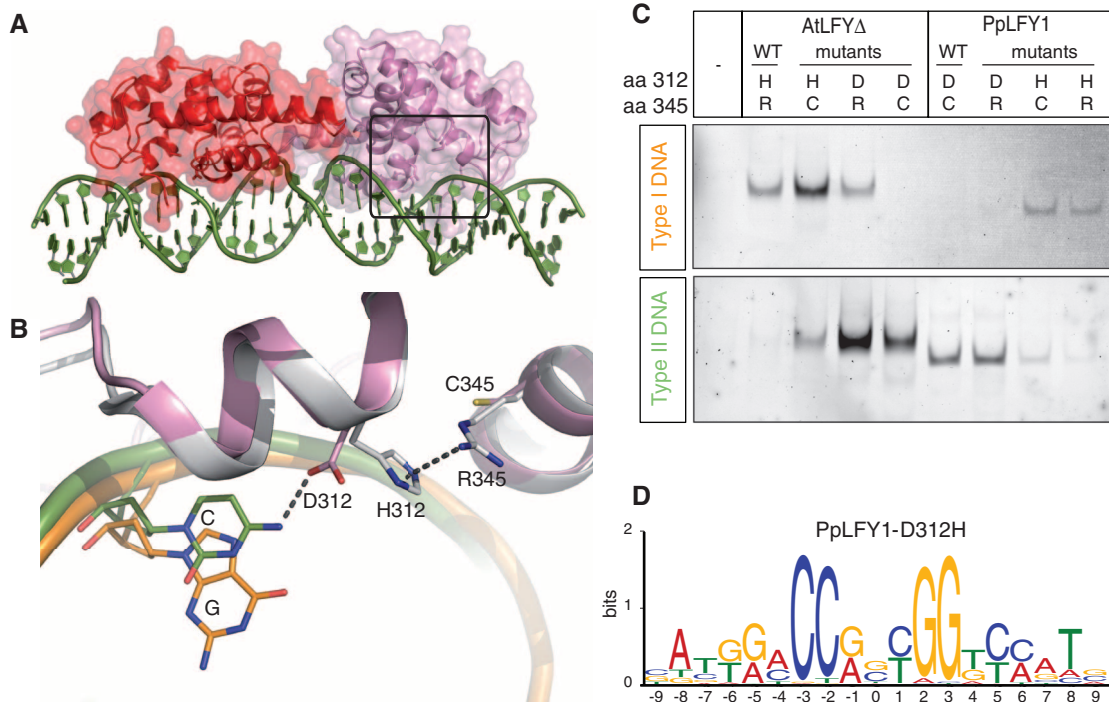
by three nucleotides, but their peripheral positions differ. The newly identified hornwort and algal LFY proteins bind to a third motif (type III) that resembles motif II, but without the central 3-bp spacer (Fig. 1C). With AtLFY, PpLFY1, and KsLFY (from *Klebsormidium subtilis*) as representative proteins of the three specificities, we confirmed that each protein displays a strong preference for one motif type (Fig. 1D, fig. S3, and table S1).

Given the broad conservation of the LFY-DBD sequence, we asked how these different specificities could be explained molecularly. We solved the crystal structure of PpLFY1-DBD bound to a motif II DNA (Fig. 2A and table S2) and compared it to the previously determined AtLFY-

DBD dimer-type I DNA complex (11). The two ternary complexes are highly similar (root mean square deviation of protein backbone atoms of 0.6 Å). However, PpLFY1-DBD makes additional contact with DNA: Aspartic acid 312 (D312) interacts with the cytosine base (C) at position 6 of the DNA binding motif, which is the nucleotide most different between motifs I and II obtained by SELEX (Figs. 1C and 2B). In AtLFY, position 312 is occupied by a histidine residue (H312), which is pulled away from the DNA by an arginine (R345), a conformation that precludes direct H312-DNA contact. In contrast, in PpLFY1, a cysteine residue (C345) replaces R345, which does not affect the positioning of D312, thus al-

Fig. 2. Structural basis for type II specificity.

(A) Crystal structure of PpLFY1-DBD (red and pink) bound to DNA (green). The boxed area is detailed in **(B)** after applying a 70° rotation. **(B)** Superimposition of AtLFY-DBD (gray)—DNA (orange) and PpLFY1-DBD (pink)—DNA (green) complexes. Specificity determinant residues and bases are represented as sticks. For amino acids: H, histidine; R, arginine; D, aspartate; C, cysteine; for DNA bases: C, cytosine; G, guanine. **(C)** Effect of specific mutations on the DNA binding specificity of AtLFY Δ and PpLFY1 in EMSA. Note that the H312-C345 combination allows binding to both motifs I and II. All proteins are at 25 nM, and only the protein-DNA complexes are shown. WT, wild type; aa, amino acid. **(D)** SELEX motif of the PpLFY1-D312H protein, bearing a strong resemblance to motif I.



lowing it to contact the cytosine base. To test the importance of positions 312 and 345, we swapped these residues between PpLFY1 and AtLFY (Fig. 2, C and D). This was sufficient to convert specificity from type I to type II and vice versa, confirming the key role of these two positions. This result is consistent with an *in vivo* study showing that a PpLFY1-D312H (D312H, Asp³¹²→His³¹²) mutant can bind a type I sequence and partially complement a *lfy* mutation in *A. thaliana* plants (9).

We next investigated binding to motif III. Motif III half-sites are similar to those of motif II (Fig. 1C), owing to the presence of a glutamine (Q) at position 312 in type III LFYs: Q is known to interact with multiple bases (14) (fig. S4), and the small residues present at position 345 (cysteine, alanine, or serine) allow Q312 to freely interact with position 6. Critically, motif III differs from motif II by the lack of the central 3-bp spacer (Fig. 1C). Modeling a LFY-DBD–motif III ternary complex by removing the 3-bp spacer in the type II DNA sequence (Fig. 3A) revealed that the interaction between helices $\alpha 1$ and $\alpha 7$, which stabilizes dimeric AtLFY- and PpLFY1-DBD positioning (11), could no longer exist for motif III.

Consistent with this observation, interacting regions of helices $\alpha 1$ and $\alpha 7$ [including the key amino acid H387 on $\alpha 7$ (11)] are highly conserved from bryophytes to angiosperms (type II and I), but are variable in algae (type III) (Fig. 1B and fig. S1). To test the importance of the $\alpha 1$ - $\alpha 7$ interaction in binding to 3-bp-spaced half-sites, we mutated PpLFY1 H387 and R390 residues (which make most $\alpha 1$ - $\alpha 7$ contacts). This was sufficient

to shift the DNA binding preference of PpLFY1 from type II to type III (Fig. 3B). These observations suggest that LFY-DBD preferentially binds to 3-bp-spaced half-sites (motifs I and II) when the $\alpha 1$ - $\alpha 7$ interaction surface is present and to motif III in the absence of this surface. Nevertheless, both the pseudosymmetry of motif III (fig. S2) and the size of LFY-DNA complexes (fig. S4) suggest that LFY binds motif III as a dimer, possibly through an alternative dimerization surface. These analyses pinpoint the molecular basis of DNA specificity changes to three amino acid sites: Positions 312 and 345 determine the half-site sequence, and position 387 determines the dimerization mode.

However, if, as shown in *P. patens* and angiosperms, LFY plays a key role throughout plant evolution, how could these changes have been tolerated? Because once arisen, they would have instantaneously modified the expression of the entire set of LFY target genes. Our LFY phylogeny (fig. S5) yields two insights: (i) Although we cannot completely rule out the occurrence of transient ancient duplications, all known duplication events occurred subsequent to changes in the binding specificity of the protein; therefore the LFY gene probably evolved new DNA binding modes independently of changes in copy number. (ii) The hornwort LFY lineage diverges from a phylogenetic node that lies between the type III and type I-II binding specificities. On closer examination, we realized that NaLFY from the hornwort *Nothoceros aenigmaticus* had type III specificity according to the SELEX experiment, despite having the H387 dimerization residue

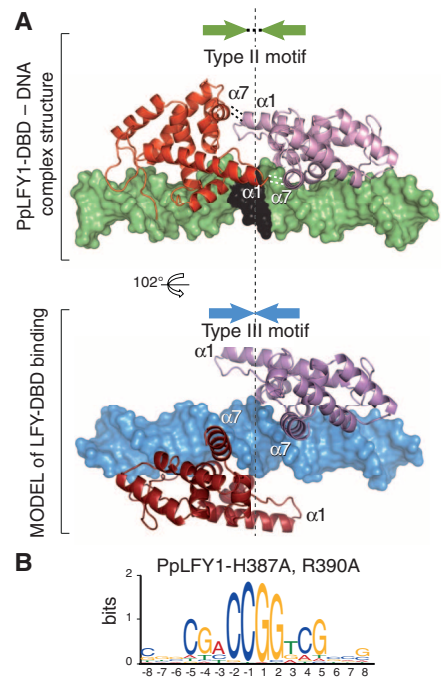


Fig. 3. Structural model for type III specificity. **(A)** (Top) PpLFY1-DBD dimer (in red and pink) bound to DNA (in green, except the black 3-bp spacer). Interactions between monomers (involving α helices $\alpha 1$ and $\alpha 7$) are shown with dashed lines. (Bottom) Modeled type III binding with DNA shown in blue. The dashed vertical line denotes the center of the pseudopalindromic DNA sequence. **(B)** SELEX motif of PpLFY1-H387A, R390A, showing a strong resemblance to motif III.

typical for type I and II specificities (Fig. 1, B and C). Using electrophoretic mobility shift assay (EMSA) experiments, we assayed NaLFY and NaLFY-DBD DNA binding and found that their dimers (fig. S6) could bind all three types of DNA motifs (Fig. 4, and figs. S3 and S7). We also established that NaLFY binding to motifs I and II was allowed by the presence of a functional $\alpha 1-\alpha 7$ interaction surface (Fig. 4). The SELEX experiment most likely identified only motif III because of its slightly more efficient binding to NaLFY (fig. S3 and table S1).

Our amino acid reconstruction analyses across the LFY phylogeny identify the phylogenetic location of the three specificity transitions that occurred during LFY evolution (Fig. 4 and fig. S8). Initially, the ancestral algal LFY bound motif III as a dimer (with Q312 and C345 half-site determinants). Subsequently, the evolution of the $\alpha 1-\alpha 7$ interaction surface generated a promiscuous LFY intermediate with two modes of DBD dimerization and a versatile glutamine residue at position 312, which bound all three types of DNA motifs. Mutations affecting positions 312 and 345 then completed the transition to type I or II

specificities. Although this precise path cannot be unambiguously determined by reconstruction alone (Fig. 4 and fig. S8), the biochemical data reveal that two LFY states (Q312-C345 and H312-C345) bind to both motifs I and II (Figs. 2C and 4). Our scenario, using either of these two states as an intermediate, provides an evolutionary route through a promiscuous platform that avoids deleterious transitions. Furthermore, this scenario is equally parsimonious in the context of all alternative organismal phylogenetic hypotheses (fig. S9). Whether these transitions were accompanied by a complete change in target gene sets or whether some cis elements coevolved with DNA binding specificity (15) is unknown. Scanning the *P. patens* genome for PpLFY1 binding sites does not suggest any global conservation of targets but does identify several MADS-box genes potentially bound by LFY in both *Arabidopsis* and *P. patens* (table S3).

A highly conserved and essential TF evolved radical shifts in DNA binding specificity by a mechanism that does not require gene duplication. Detailed structural characterization of the different modes of DNA binding across the tran-

sition to land plants enabled us to capture LFY in a state of increased promiscuity that has persisted in *N. aenigmaticus*. This promiscuous intermediate probably facilitated the evolutionary transition between specificities, as previously shown for the evolution of metabolic enzymes or nuclear receptors (16–18). Although we have focused on the more intractable problem of evolution in single-copy TFs, it is plausible that the mechanisms we describe could also contribute to the evolution of TFs encoded by multigene families.

References and Notes

1. S. B. Carroll, *Cell* **101**, 577–580 (2000).
2. I. S. Peter, E. H. Davidson, *Cell* **144**, 970–985 (2011).
3. B. Prud'homme, N. Gompel, S. B. Carroll, *Proc. Natl. Acad. Sci. U.S.A.* **104** (suppl. 1), 8605–8612 (2007).
4. G. A. Wray, *Nat. Rev. Genet.* **8**, 206–216 (2007).
5. H. E. Hoekstra, J. A. Coyne, *Evolution* **61**, 995–1016 (2007).
6. G. Badis *et al.*, *Science* **324**, 1720–1723 (2009).
7. M. F. Berger *et al.*, *Cell* **133**, 1266–1276 (2008).
8. C. R. Baker, V. Hanson-Smith, A. D. Johnson, *Science* **342**, 104–108 (2013).
9. A. Maizel *et al.*, *Science* **308**, 260–263 (2005).
10. T. Tanahashi, N. Sumikawa, M. Kato, M. Hasebe, *Development* **132**, 1727–1736 (2005).
11. C. Hamès *et al.*, *EMBO J.* **27**, 2628–2637 (2008).
12. Y. Zhao, D. Granas, G. D. Stormo, *PLOS Comput. Biol.* **5**, e1000590 (2009).
13. E. Moyroud *et al.*, *Plant Cell* **23**, 1293–1306 (2011).
14. N. M. Luscombe, R. A. Laskowski, J. M. Thornton, *Nucleic Acids Res.* **29**, 2860–2874 (2001).
15. C. R. Baker, B. B. Turch, A. D. Johnson, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7493–7498 (2011).
16. O. Khersonsky, C. Roodveldt, D. S. Tawfik, *Curr. Opin. Chem. Biol.* **10**, 498–508 (2006).
17. A. Aharoni *et al.*, *Nat. Genet.* **37**, 73–76 (2005).
18. G. N. Eick, J. K. Colucci, M. J. Harms, E. A. Ortlund, J. W. Thornton, *PLOS Genet.* **8**, e1003072 (2012).

Acknowledgments: We thank M. Schmid for help with sequencing; E. Masson, J. Kyoizuka, M. Hasebe, C. Scutt, C. Finet, and J. C. Villarreal for sharing materials; A. Maizel and S. Rensing for discussion; and A. Maizel, R. Worsley-Hunt, A. Mathelier, M. Blazquez, O. Nilsson, C. Zubieta, and J. Chen for critical reading of the manuscript. This work was supported by funds from the Max Planck Society (D.W.), the Agence Nationale de la Recherche (grant Charmful SVSE2–2011) and Coopérations et Mobilités Internationales Rhône-Alpes (F.P.), and the FP7 P-CUBE number 227764 (R.D.). The 1000 Plants (1KP) initiative, led by G.K.-S.W., is funded by Alberta Ministry of Enterprise and Advanced Education, Alberta Innovates Technology Futures, Innovates Centre of Research Excellence, Musea Ventures, and BGI-Shenzhen. GenBank accession numbers are as follows: *AmboLFY*, KF193872; *NaLFY*, KF269532; *KsLFY*, KF269535; *CsLFY*, KF269533; *CvLFY*, KF269536; and *CyLFY*, KF269534. The Protein Data Bank identification number is 4BHK for the reported crystal structure. We declare no competing financial interests.

Supplementary Materials

www.sciencemag.org/content/343/6171/645/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S9
Tables S1 to S5
References (19–42)
Database S1

7 November 2013; accepted 23 December 2013
Published online 16 January 2014;
10.1126/science.1248229

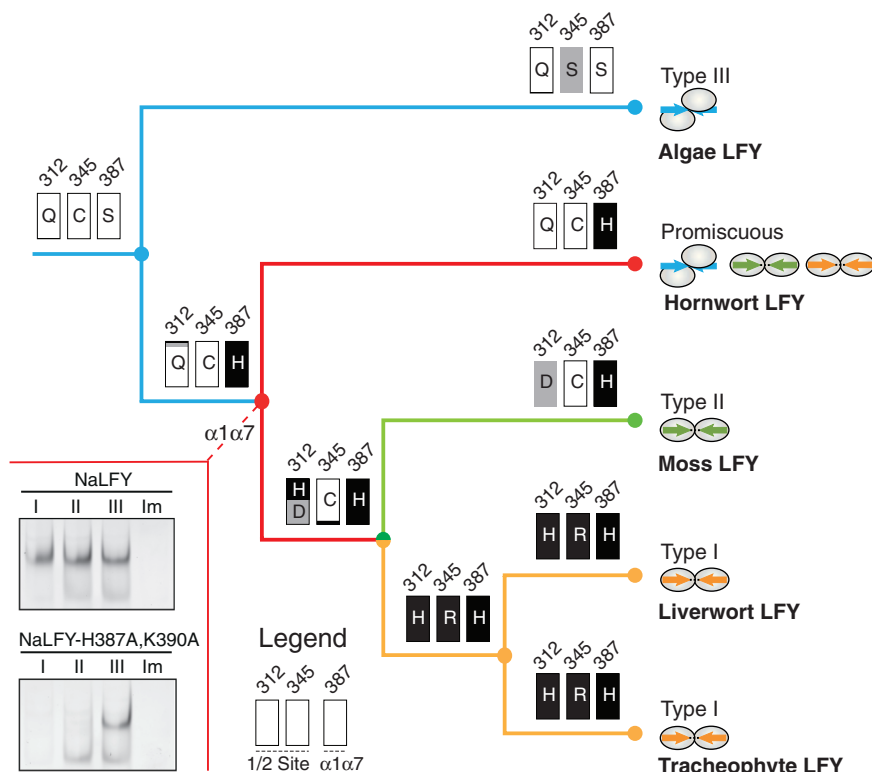


Fig. 4. Proposed evolution of LFY DNA binding specificity in green plants. The Bayesian estimation of the posterior probability of ancestral states for amino acid positions 312, 345, and 387 is depicted at the major phylogenetic nodes. Probabilities for different residues at a given position and node are indicated by the relative size of stacked boxes. The analysis shows that the ancestral LFY most likely possessed a type III specificity and that the promiscuous form arose when land plants emerged. DNA binding specificity is color-coded: type I, orange; type II, green; type III, blue; relaxed specificity, red. $\alpha 1\alpha 7$ refers to the $\alpha 1-\alpha 7$ dimerization interface. (Inset) NaLFY interacts with all three types of DNA binding motifs in EMSA (see also fig. S7), but not with the type I mutated probe (Im). The H387A and K390A mutations reduced the binding to type I or II motifs, but not to type III. Both proteins are at 1 μ M; only the protein-DNA complexes are shown.