

Table 2. Estimated proportion (mode and quantiles) of assortative mating for moths of both carbon types for all sites in 2002, 2003, and for both years together.

Year	Proportion of assortative mating					
	C ₃ moths			C ₄ moths		
	mode	q0.025	q0.975	mode	q0.025	q0.975
2002	0.949	0.741	0.988	0.961	0.771	0.990
2003	0.983	0.732	0.998	0.965	0.783	0.991
Both years	0.970	0.832	0.993	0.964	0.857	0.989

each site, we sampled insects once per week over 5 weeks, covering most of the flight period. We caught a total of 363 individuals. Again, the $\delta^{13}\text{C}$ values of the wings and of the 298 spermatophores carried by the females (Fig. 1B) made it possible to determine host-plant type unambiguously.

As in the 2002 sampling, the four sites differed in the relative proportions of C₃ and C₄ moths (Table 1). Unexpectedly, we found no evidence of temporal isolation between races in terms of flight periods. Indeed, the null hypothesis of homogeneity of C₃:C₄ proportions over time was not rejected at sites 7, 8, and 9, and there was no clear trend over time at site 6 despite rejection of the null hypothesis at this site (table S1). Hence, the two host races occurred together for a period of at least 5 weeks, covering almost the entire flight period. At all four sites, only one (C₄) female ($n = 16$ C₃ and $n = 167$ C₄ females) mated with a male that had not developed on the same type of host plant as herself (Table 1). Our 2003 results therefore confirm the very strong assortative mating observed in 2002 (Table 2). In both years, the maximum-likelihood estimates of the proportion of assortative mating in the C₃ and in the C₄ races in our study area were >95%, with 95% credibility intervals of 83.2 to 99.3% and 85.7 to 98.9%, respectively (Table 2). Interestingly, all three apparently hybrid matings observed in this study occurred in sites with C₃:C₄ moth ratios far from 1:1, as expected under assortative mating, rather than in sites with a ratio close to 1:1, as expected under random mating (Table 1).

Our estimate of hybridization frequency (<5%) is of the same order of magnitude as that of gene flow (<1%) based on the pattern of genetic differentiation between the hop-mugwort-E and maize-Z populations (13). Assortative mating, rather than postzygotic isolation, therefore appears to be the main cause of genetic differentiation in the ECB. Our results provide an example of the maintenance of very strong assortative mating even in the absence of spatial or temporal isolation. Measuring the net result of all factors influencing assortative mating in natura was possible here because *O. nubilalis* host races happen to feed on two isotopically different host-plant groups. Beyond this practical feature, our results reveal two additional reasons for which the ECB is a

particularly interesting and suitable model to study assortative mating and other processes involved in sympatric speciation: (i) It displays a relatively high level of assortative mating compared with other host races; and (ii) this assortative mating appears to be strongly driven by factors not directly related to host-plant adaptation, e.g., sex pheromones (20). Both are somewhat rare cases among host races (1–3). Therefore, studies on assortative mating in ECB host races could pave the way to dissecting the relative contribution of various factors involved in reproductive isolation, as well as pre- and postzygotic barriers to gene flow.

References and Notes

1. S. Via, *Trends Ecol. Evol.* **16**, 381 (2001).
2. S. Berlocher, J. Feder, *Annu. Rev. Entomol.* **47**, 773 (2002).
3. M. Drès, J. Mallet, *Philos. Trans. R. Soc. London B Biol. Sci.* **357**, 471 (2002).
4. A. Kondrashov, M. Shpak, *Proc. R. Soc. London B Biol. Sci.* **265**, 2273 (1998).
5. M. Noor, *Heredity* **83**, 503 (1999).
6. W. Rice, *Evol. Ecol.* **1**, 301 (1987).
7. J. Feder, T. Hunt, L. Bush, *Entomol. Exp. Appl.* **69**, 117 (1993).
8. J. Feder et al., *Proc. Natl. Acad. Sci. U.S.A.* **91**, 7990 (1994).
9. I. Emelianov, M. Drès, W. Baltensweiler, J. Mallet, *Evolution* **55**, 2002 (2001).

10. I. Emelianov, F. Simpson, P. Narang, J. Mallet, *J. Evol. Biol.* **16**, 208 (2003).
11. T. P. Craig, J. D. Horner, J. K. Itami, *Evolution* **51**, 1552 (1997).
12. S. Via, *Evolution* **53**, 1446 (1999).
13. C. Martel, A. Réjasse, F. Rousset, M.-T. Bethenod, D. Bourguet, *Heredity* **90**, 141 (2003).
14. P. Parker, A. Snow, M. Schug, G. Booton, P. Fuerst, *Ecology* **79**, 361 (1998).
15. D. Rubenstein, K. Hobson, *Trends Ecol. Evol.* **19**, 256 (2004).
16. S. Ponsard et al., *Can. J. Zool.* **82**, 1177 (2004).
17. Y. Thomas, M.-T. Bethenod, L. Pélozuelo, B. Frérot, D. Bourguet, *Evolution* **57**, 261 (2003).
18. M.-T. Bethenod et al., *Heredity* **94**, 264 (2005).
19. A. Bontemps, D. Bourguet, L. Pélozuelo, M.-T. Bethenod, S. Ponsard, *Proc. R. Soc. London B Biol. Sci.* **271**, 2179 (2004).
20. L. Pélozuelo, C. Malosse, G. Genestier, H. Guenego, B. Frérot, *J. Chem. Ecol.* **30**, 335 (2004).
21. In other areas (e.g., the United States and Italy), populations on maize use both blends (26).
22. W. Showers, G. Reed, J. Robinson, M. Derozari, *Environ. Entomol.* **5**, 1099 (1976).
23. Materials and methods are available as supporting material on Science Online.
24. M. De Niro, S. Epstein, *Geochim. Cosmochim. Acta* **42**, 495 (1978).
25. Solidified droplets of sperm and nutritious substances deposited by males in females' genital ducts and later used by the female.
26. J. A. Klun et al., *Environ. Entomol.* **4**, 891 (1975).
27. We thank A. Estoup, M. Gibernau, R. G. Harrison, M. Kirkpatrick, L. Lapchin, J. Mallet, F. Rousset, and anonymous referees for helpful comments. Supported by the Institut National de la Recherche Agronomique–Action Incitative Programmée “Organismes Génétiquement Modifiés et Environnement,” the Ministère de l'Éducation Nationale, de la Recherche et de la Technologie–Appel d'Offre “Impact des OGM,” and the Institut Français de la Biodiversité–Appel d'Offre “Biodiversité et Changement Global.”

Supporting Online Material

www.sciencemag.org/cgi/content/full/308/5719/258/DC1

Materials and Methods

Table S1

References and Notes

16 November 2004; accepted 3 February 2005
10.1126/science.1107577

The Floral Regulator LEAFY Evolves by Substitutions in the DNA Binding Domain

Alexis Maizel,^{1,2} Maximilian A. Busch,² Takako Tanahashi,³ Josip Perkovic,¹ Masahiro Kato,³ Mitsuyasu Hasebe,⁴ Detlef Weigel^{1*}

The plant-specific transcription factor LEAFY controls general aspects of the life cycle in a basal plant, the moss *Physcomitrella patens*. In contrast, LEAFY has more specialized functions in angiosperms, where it specifically induces floral fate during the reproductive phase. This raises the question of a concomitant change in the biochemical function of LEAFY during the evolution of land plants. We report that the DNA binding domain of LEAFY, although largely conserved, has diverged in activity. On the contrary, other, more rapidly evolving portions of the protein have few effects on LEAFY activity.

LEAFY (LFY) is found in all land plants, which evolved during the past 400 million years. The proteins are remarkably well

conserved with two blocks of similarity, in the N- and C-terminal regions (the N and C domain) (Fig. 1A). All missense mutations

identified in *Arabidopsis* mutant screens map to these two domains, which make up a little more than half of the protein (table S1 and fig. S1). Between nonflowering and flowering plants, there is very little sequence conservation outside the N and C domains. In *Arabidopsis*, the nuclear LFY protein binds sequences in the enhancers of several floral homeotic genes, including *APETALA1* (*API*) (1–4). The missense changes found in mutant alleles very much reduce in vitro DNA binding to the *API* promoter (fig. S1). Deletion analyses identify a minimal DNA binding domain from amino acids 320 to 507 (numbering refers to consensus sequence), which includes the highly conserved C domain (Fig. 1A and fig. S1). Although the N domain is not essential for DNA binding, DNA binding is compromised in a deletion derivative that retains part of the N domain, as well as in a protein with a point mutation in the N domain. These observations suggest that the N domain regulates the activity of the DNA binding domain proper (fig. S1).

There is generally only a single copy of LFY in angiosperms (Fig. 1B). In species with lineage-specific duplications, these do not seem to have diverged in function (5, 6). Nonflowering plants appear to have additional copies, but these are also all closely related and there is no evidence for major subfunctionalization (7, 8). The strong sequence conservation of the DNA binding domain suggested that the molecular function of LFY is conserved as well. To test this assumption, we linked *LFY* cDNAs from 14 species to the *Arabidopsis LFY* promoter and introduced them into a strong *lfy* mutant. The 14 species represent the three main taxa with known LFY homologs: ferns and mosses, gymnosperms, and angiosperms (Fig. 1B).

Angiosperm genes fully complement *lfy* mutant, whereas gymnosperm genes provide only partial rescue (Fig. 2A), with *PRFLL* from pine (9) having more activity in this assay than *WeINDLY* from *Welwitschia* (7) (table S3). The *LFY* and *NDLY* clades represent a gymnosperm-specific duplication event, with the *NDLY* lineage having been lost in angiosperms (10). Among homologs from the most basal group, the fern genes *CrLFY2* and *AilFY4* (8) have some rescue ability, although less than the gymnosperm

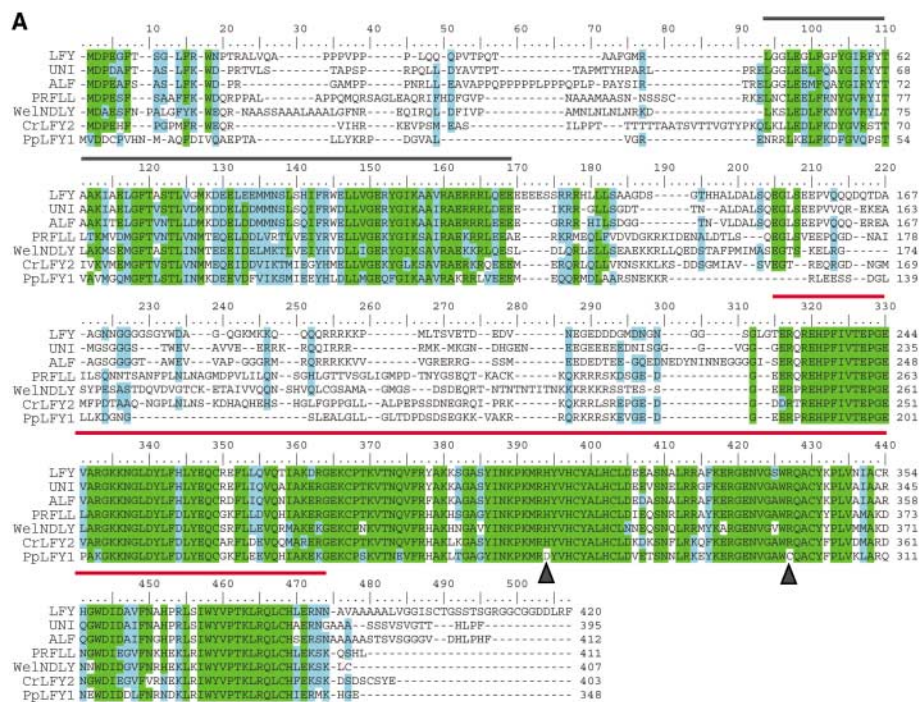
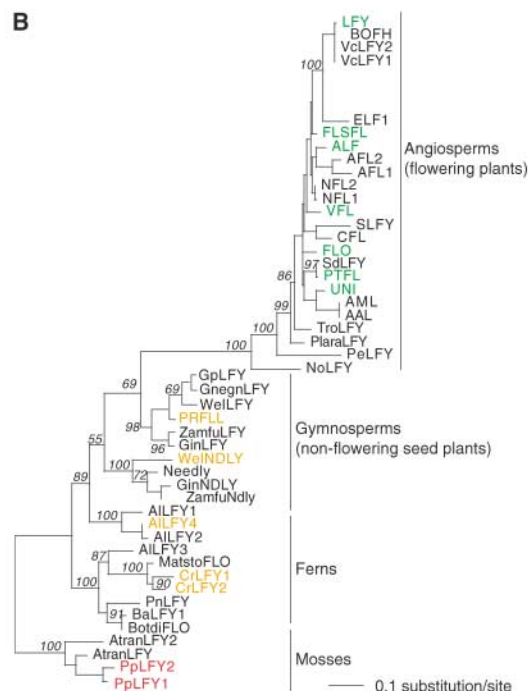


Fig. 1. Phylogenetic relationships of LFY sequences. (A) Aligned amino acid sequences of LFY (*Arabidopsis*), UNI (pea), ALF (petunia), PRFLL (pine), WeINDLY (*Welwitschia*), CrLFY2 (fern *Ceratopteris*), and PpLFY1 (moss *Physcomitrella*) (16). A composite numbering with gaps (dashes) is indicated on top; a gapless count is shown to the right. Unless specified, absolute numbering is used in descriptions. Shaded residues are present in at least 70% of sequences. Green indicates identical residues, and cyan, those with similar biochemical properties. N and C domains are overlined in black and red, respectively. PpLFY1 amino acids that were mutated (Fig. 3) are indicated by arrowheads. (B) Phylogenetic tree of 48 LFY sequences spanning the four major clades of extant land plants. Only bootstrap support above 50% is shown (italic numbers). Color code indicates ability to complement an *Arabidopsis lfy* mutant (green, full; orange, partial; red, no complementation).



genes, whereas the moss genes *PpLFY1* and *PpLFY2* (8) are inactive. This gradient of complementation reflects the phylogenetic distance from angiosperms (Fig. 1B) and suggests that a continuum of discrete and nonneutral changes, rather than a sudden modification, is responsible for changes in function.

We used microarrays to investigate in more detail the different activities of the homologs. Floral development was synchronized by transfer of plants from short days to long days,

which induces *LFY* promoter activity (11, 12). Of the 16 genes responding most strongly to *Arabidopsis LFY*, 15 and 13 are significantly induced by the angiosperm homologs *UNI* and *ALF*, respectively. *WeINDLY*, a gymnosperm representative, induces two targets, whereas *CrLFY2*, a fern gene, induces only one target. None of the *LFY* targets respond to *PpLFY1* from moss (Fig. 2B and fig. S2).

For orthologs of animal HOX proteins, altered activity in cross-species experiments

¹Department of Molecular Biology, Max Planck Institute for Developmental Biology, D-72076 Tübingen, Germany. ²Plant Biology Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037, USA. ³Department of Biological Sciences, Graduate School of Science, University of Tokyo, Tokyo 113-0033, Japan. ⁴National Institute for Basic Biology, Okazaki 444-8585, Japan.

*To whom correspondence should be addressed. E-mail: weigel@weigelworld.org

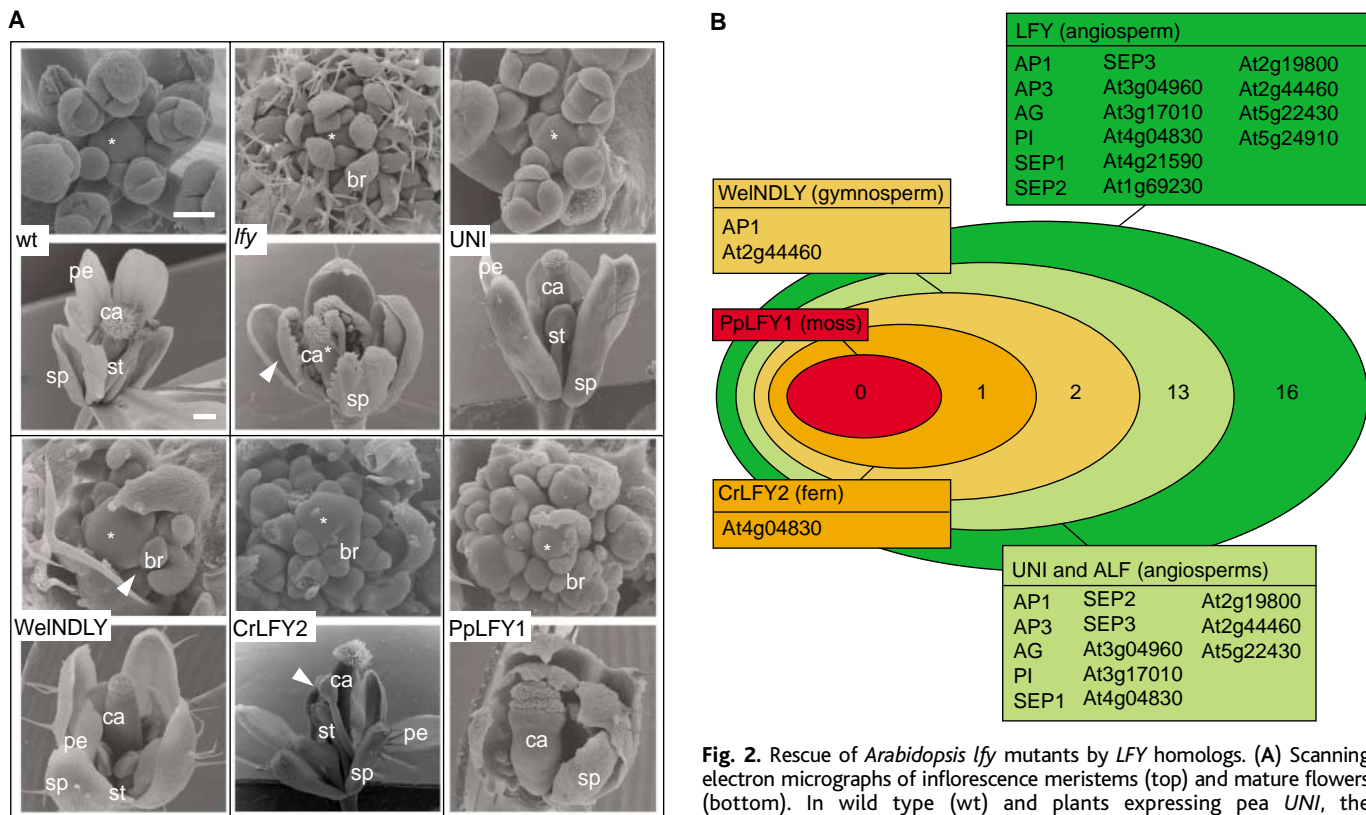


Fig. 2. Rescue of *Arabidopsis lfy* mutants by LFY homologs. (A) Scanning electron micrographs of inflorescence meristems (top) and mature flowers (bottom). In wild type (wt) and plants expressing pea *UNI*, the inflorescence meristem (asterisk) produces floral meristems, which differentiate into flowers composed of sepals (se), petals (pe), stamens (st), and carpels (ca). Plants expressing moss *PpLFY1*, like nontransgenic *lfy* mutants, have bracts (br) subtending secondary inflorescences, which replace early-arising flowers (arrowhead). Later-arising flowers consist of organs with mixed sepal and carpel identity (arrowhead) and unfused carpels (ca*). *Welwitschia WelNDLY* and fern *CrLFY2* partially rescue this phenotype. Later-arising flowers of *lfy WelNDLY* plants have an almost wild-type complement of organs, whereas *CrLFY2* flowers have spirally arranged organs that include petaloid stamens (arrowhead). Scale bar is 200 μ m. (B) Synthetic Venn diagram of genes induced by *Arabidopsis LFY*. Ellipses represent the sets of genes induced by the different homologs (see Fig. 1B for color code).

has been traced back to changes outside the DNA binding domain proposed to affect the transcriptional activation or repression potential (13–15). To determine whether a similar scenario applies to LFY, we fused LFY homologs to the VP16 activation domain and tested whether they could, like an *Arabidopsis LFY*-VP16 fusion, induce expression of yeast reporters under the control of LFY binding sites from the homeotic genes *AP1* or *AGAMOUS (AG)* (1, 2). The ability of the different VP16 fusions to interact with the *AP1* and *AG* sites and activate the yeast reporters parallels their rescue activity in plants. Of the proteins from nonflowering plants, PRFL1 is as effective as *Arabidopsis LFY*, whereas *WelNDLY* and *CrLFY2* are substantially less active. Moss *PpLFY1*, which is inactive in the transgenic plant assay, is also inactive in yeast, indicating that the failure to complement *Arabidopsis lfy* mutants is not simply caused by a change in transcriptional activation potential (Fig. 3A and fig. S3). Thus, in an a-minima model, declining ability to replace *Arabidopsis LFY* in plants is caused by a progressive failure to interact with canonical LFY binding sites.

Because the conserved N and C domains had been implicated in DNA binding, we suspected that the changes in activity are caused by divergence in these two domains, rather than by changes in the surrounding sequences. To test this hypothesis, we swapped the N and C domains between *Arabidopsis LFY* and *CrLFY2* (Fig. 3B). Across the entire sequence, the LFY-*CrLFY2* chimera, in which the N and C domains are derived from fern *CrLFY2*, is more similar to *Arabidopsis LFY* than the *CrLFY2*-LFY chimera (77% versus 67% sequence identity to *Arabidopsis LFY*). Nevertheless, it is the less-similar *CrLFY2*-LFY chimera that provides almost complete rescue when introduced into *Arabidopsis lfy* mutants, whereas the other chimera has very little activity, comparable to *CrLFY2* itself (Fig. 3B). Thus, changes in the highly conserved N and C domains are responsible for most of the functional differences between the proteins from fern and *Arabidopsis*. We confirmed that these differences were caused by differential DNA binding activities with the yeast assay described above (fig. S4). Combination of the N and C halves of *Arabidopsis LFY* and *CrLFY2* showed that

the C domain, which corresponds to the minimal DNA binding domain, is primarily responsible for the divergence in function (fig. S5).

We found two amino acid substitutions, His³⁹⁴→Asp³⁹⁴ (H394D) and Arg⁴²⁷→Cys⁴²⁷ (R427C) (16), which discriminate between the DNA binding domains of proteins at opposite ends of the functional spectrum, *Arabidopsis LFY* and moss *PpLFY1*. We created versions of *PpLFY1* in which these positions were individually changed to the angiosperm sequence and tested them again in the yeast assay. *PpLFY1* (D394H) but not *PpLFY1* (C427R) partially activated transcription (Fig. 3C and fig. S3). Similarly, in transgenic plants *PpLFY1* (D394H) but not *PpLFY1* (C427R) provided partial LFY activity (Fig. 3D). That a one-amino acid change is sufficient to have *PpLFY1* bind a canonical LFY binding site indicates that position 394 is crucial for DNA binding.

The D394H substitution appears to be restricted to true mosses (10) and is not found in the liverworts *Riccia* and *Marchantia*. We consider it unlikely that the aspartate (D) at position 394 in moss *PpLFY1* simply inactivates the protein, because moss

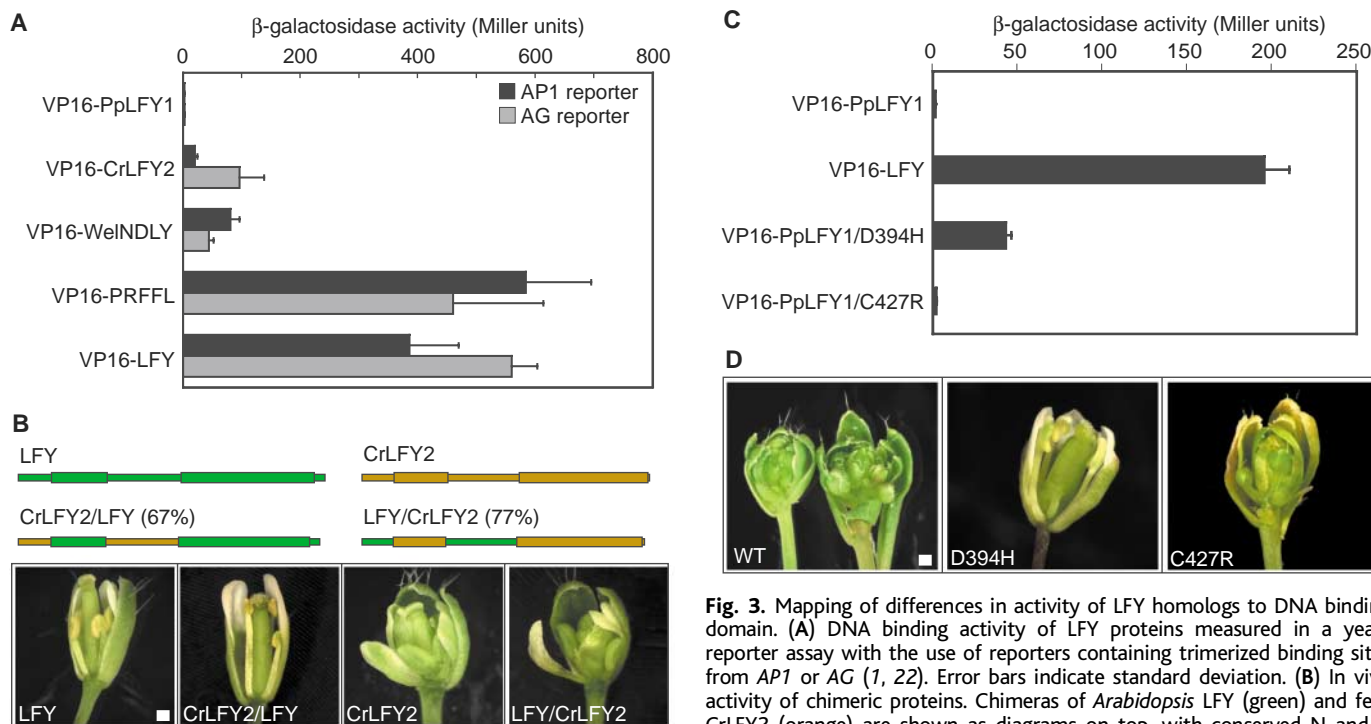


Fig. 3. Mapping of differences in activity of LFY homologs to DNA binding domain. (A) DNA binding activity of LFY proteins measured in a yeast reporter assay with the use of reporters containing trimerized binding sites from AP1 or AG (1, 22). Error bars indicate standard deviation. (B) In vivo activity of chimeric proteins. Chimeras of *Arabidopsis* LFY (green) and fern CrLFY2 (orange) are shown as diagrams on top, with conserved N and C domains shown as boxes; sequence identity with *Arabidopsis* LFY is indicated. Below, partially dissected flowers of *lfy* mutants expressing the different sequences are compared. Scale bar is 200 μm. (C) DNA binding affinity of LFY, PpLFY1, and PpLFY1 mutants assayed in yeast with the AP1 binding site. (D) Images of *lfy* plants expressing wild-type PpLFY1 (WT) or PpLFY1 with D394H and C427R substitutions. Organs with petaloid and stamenoid characters are only seen with the D394H mutant. Scale bar is 200 μm.

plants lacking both *PpLFY* genes have dramatic developmental defects (17). We therefore conclude that PpLFY1 has most likely a different DNA binding specificity than its angiosperm counterpart (10). The gradient of activity among nonflowering plants furthermore suggests a systematic change in the DNA binding specificity of LFY homologs.

In animals, the relationship between molecular evolution of developmental regulators and morphological changes is best understood for HOX and PAX homeodomain proteins. Modification of expression patterns can lead to changes in morphology (18), but changes in the proteins themselves may be important as well. Functional differences in the activity of HOX orthologs from different species map to the repression and activation domains rather than the DNA binding domain (13, 14). In contrast, our results emphasize the LFY DNA binding domain as a source of functional variation across species. Examples of molecular evolution of transcription factors in flies and worms that implicated the DNA binding domains as the source of variation have been interpreted as divergence within families of related multi-copy genes. Gene duplication followed by subfunctionalization (19), as described for the PAX family (20), is unlikely to apply to LFY, because there has been no radiation of the LFY family during the evolution of land plants.

In a broader perspective of functional evolution of LFY among land plants, two scenarios can be considered. LFY might control similar networks of genes in nonflowering and flowering plants, with coevolution of target sequences and LFY DNA binding specificity. If this is the case, one still needs to postulate that these networks have been modified, because the primary target of LFY in angiosperms, *API*, is restricted to flowering plants (21). Alternatively, there may have been a complete change of LFY function between basal taxa and flowering plants, in which an initial, albeit gradual change in biochemical activity was the prerequisite for recruitment and/or intercalation of new targets, such as *API*.

References and Notes

1. F. Parcy, O. Nilsson, M. A. Busch, I. Lee, D. Weigel, *Nature* **395**, 561 (1998).
2. M. A. Busch, K. Bomblies, D. Weigel, *Science* **285**, 585 (1999).
3. R. S. Lamb, T. A. Hill, Q. K. Tan, V. F. Irish, *Development* **129**, 2079 (2002).
4. D. A. William *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 1775 (2004).
5. K. Bomblies *et al.*, *Development* **130**, 2385 (2003).
6. M. Wada, Q. F. Cao, N. Kotoda, J. Soejima, T. Masuda, *Plant Mol. Biol.* **49**, 567 (2002).
7. M. W. Frohlich, D. S. Parker, *Syst. Bot.* **25**, 155 (2000).
8. S. Himi *et al.*, *J. Mol. Evol.* **53**, 387 (2001).
9. E. J. Mellerowicz, K. Horgan, A. Walden, A. Coker, C. Walter, *Planta* **206**, 619 (1998).
10. M. W. Frohlich, G. F. Estabrook, *Mol. Biol. Evol.* **17**, 1914 (2000).
11. M. Schmid *et al.*, *Development* **130**, 6001 (2003).

12. M. A. Blázquez, L. Soowal, I. Lee, D. Weigel, *Development* **124**, 3835 (1997).
13. M. Ronshaugen, N. McGinnis, W. McGinnis, *Nature* **415**, 914 (2002).
14. R. Galant, S. B. Carroll, *Nature* **415**, 910 (2002).
15. B. Gebelein, J. Culi, H. D. Ryoo, W. Zhang, R. S. Mann, *Dev. Cell* **3**, 487 (2002).
16. 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.
17. T. Tanahashi, N. Sumikawa, M. Kato, M. Hasebe, *Development* **132**, 1727 (2005).
18. S. B. Carroll, *Cell* **101**, 577 (2000).
19. M. Lynch, A. Force, *Genetics* **154**, 459 (2000).
20. Z. Kozmik *et al.*, *Dev. Cell* **5**, 773 (2003).
21. A. Litt, V. F. Irish, *Genetics* **165**, 821 (2003).
22. J. U. Lohmann *et al.*, *Cell* **105**, 793 (2001).
23. We thank J.-P. Beltrán, K. Bomblies, E. Coen, E. Dennis, M. Frohlich, R. Koes, J. Kyozyuka, K. Lee, R. Lozano, J.-M. Martínez-Zapater, S. Strauss, M. Wada, and C. Walter for plasmids of LFY homologs and M. Gerberding, M. Frohlich, J. Lempe, I. and J. Lohmann, W. McGinnis, J. Palatnik, M. Schmid, and R. Sommer for discussion. Supported by a Marie Curie Long Term Fellowship to A.M., by a grant from NIH to D.W., by grants from the Ministry of Education, Science, and Technology and the Japan Society for the Promotion of Science to M.H. and M.K., and by the Max Planck Society. D.W. is a director of the Max Planck Institute.

Supporting Online Material
www.sciencemag.org/cgi/content/full/308/5719/260/DC1
 Materials and Methods
 Figs. S1 to S5
 Tables S1 to S4
 References

2 December 2004; accepted 23 February 2005
 10.1126/science.1108229