

LETTERS

Escape from adaptive conflict after duplication in an anthocyanin pathway gene

David L. Des Marais¹ & Mark D. Rausher¹

Gene duplications have been recognized as an important source of evolutionary innovation and adaptation since at least Haldane¹, and their varying fates may partly explain the vast disparity in observed genome sizes². The expected fates of most gene duplications involve primarily non-adaptive substitutions leading to either non-functionalization of one duplicate copy or subfunctionalization³, neither of which yields novel function. A significant evolutionary problem is thus elucidating the mechanisms of adaptive evolutionary change leading to evolutionary novelty. Currently, the most widely recognized adaptive process involving gene duplication is neo-functionalization (NEO-F), in which one copy undergoes directional selection to perform a novel function after duplication⁴. An alternative, but understudied, adaptive fate that has been proposed is escape from adaptive conflict (EAC), in which a single-copy gene is selected to perform a novel function while maintaining its ancestral function^{5,6}. This gene is constrained from improving either novel or ancestral function because of detrimental pleiotropic effects on the other function. After duplication, one copy is free to improve novel function, whereas the other is selected to improve ancestral function. Here we first present two criteria that can be used to distinguish NEO-F from EAC. Using both tests for positive selection and assays of enzyme function, we then demonstrate that adaptive evolutionary change in a duplicated gene of the anthocyanin biosynthetic pathway in morning glories (*Ipomoea*) is best interpreted as EAC. Finally, we argue that this phenomenon likely occurs more often than has been previously believed and may thus represent an important mechanism in generating evolutionary novelty.

Novel gene function associated with gene duplication can arise in either of two ways. According to the NEO-F model, novel function arises after gene duplication, with one copy maintaining ancestral function, whereas the second copy is selected to perform a new function. By contrast, under EAC, novel function arises first in the single-copy ancestral gene, which results in a reduction in that copy's ability to perform the original function. After duplication, each copy is free to specialize on either the original or the novel function and improve those functions. Although both of these processes have been recognized for at least two decades as alternative possibilities^{4,6}, we are aware of only one previous attempt to distinguish between them⁷.

Two criteria may be used to distinguish between NEO-F and EAC. First, under EAC adaptive change occurs in both duplicate copies, whereas under NEO-F only one copy undergoes adaptive change because purifying selection acts to maintain ancestral function in one copy. Second, under EAC ancestral function is improved, whereas under NEO-F it is not. Here we apply both of these criteria to distinguish between NEO-F and EAC as explanations for adaptive change in duplicated copies of the anthocyanin biosynthetic pathway gene dihydroflavonol-4-reductase (DFR).

Plant dihydroflavonol-4-reductases (EC number 1.1.1.219) function most conspicuously in the reduction of several flavonoid

precursors of anthocyanin pigments and their related phytoalexins, though the full scope of DFR function is unknown. The widely distributed 3-hydroxyanthocyanidins pelargonidin, cyanidin and delphinidin are primarily responsible for red, purple and blue flowers in angiosperms and are the downstream products of DFR activity on dihydrokaempferol (DHK), dihydroquercetin (DHQ) and dihydromyricetin (DHM), respectively⁸. DFR from several taxa is able to reduce the flavonols naringenin and eriodyctiol to produce the flavan-4-ols, which are themselves precursors of phlobaphenes and of two rare anthocyanidins, apigeninidin and luteolinidin^{9,10}. In the common morning glory, *Ipomoea purpurea*, DFR is present as a small, tandemly arrayed three-gene family spanning approximately 17 kilobases (ref. 11). All three copies are expressed in *I. purpurea* (although *DFR-A* and *DFR-C* are expressed in overlapping but fewer tissues relative to *DFR-B*) and have the conserved intron/exon structure observed across eudicots.

Phylogenetic analysis of *DFR* in the *Ipomoea* gene family and closely related single-copy taxa shows that each *DFR* copy identified in *I. purpurea*, A, B and C, forms a clade (Fig. 1). Each clade of orthologues, the entire *DFR* gene family, *DFR* from all Convolvulaceae and *DFR* from all Solanales constitute well-supported clades. Sister to the three-copy species clade is a large clade of morning glories characterized by a single copy of *DFR* and represented here by *Evolvulus glomeratus*; most Solanaceae species are also single-copy and our results suggest that there was a lineage specific duplication in *Petunia*. The gene tree also reveals two separate *DFR* duplication events in the Convolvulaceae: the first gave rise to *DFR-B* and another lineage that experienced a second duplication event creating the *DFR-A* and *DFR-C* copies.

Codon-based models of sequence evolution¹² indicate that single-copy *DFRs* have historically experienced purifying selection ($dN/dS(\omega) = 0.13$; Fig. 1), as did the base of the clade containing the Convolvulaceae *DFR* gene family ($dN/dS = 0.104$). After the first gene duplication, the lineage subtending *DFR-B* has a ratio of replacement to synonymous substitutions statistically indistinguishable from the single copy *DFRs* ($dN/dS = 0.183$). By contrast, this ratio has increased on the lineage subtending the *DFR-A/C* clade ($dN/dS = \text{infinite}$). A likelihood model infers that 18 replacement and no synonymous substitutions occurred at the base of the A/C clade. The probability of this pattern occurring under neutrality is $P = 0.004$ (see Supplementary Methods), indicating the action of repeated positive selection along this branch after the first duplication. Non-synonymous substitution was also significantly elevated above background on the branch subtending the *DFR-A* clade ($dN/dS = 0.842$, $P = 0.009$) and marginally so on the branch subtending the *DFR-C* clade ($dN/dS = 0.29$, $P = 0.06$), although a branch-sites test fails to detect positive selection on either of these ($P = 0.21$ and $P > 0.9$, respectively). Changes immediately after the second duplication thus hint at positive selection, but do not definitively reveal it.

¹Department of Biology and University Program in Genetics and Genomics, Box 90338, Duke University, Durham, North Carolina 27708-0338, USA.

Subsequent non-synonymous substitution on copies in the B and the A/C clades varies slightly from background rates but still bears the signature of purifying selection. A clade–sites test for the *DFR-A/C* clade as a whole compared with *DFR-B* as a whole reveals that most codon sites in both clades remain under strong purifying selection (Fig. 1). It thus appears that there was a burst of adaptation in the A and C copies very soon after the first, and possibly after the second, duplication. After this burst, selection once again became primarily purifying.

We chose *Solanum lycopersicon* and *E. glomeratus* as representative ‘pre-duplication’ species and *I. purpurea* and *Convolvulus arvensis* as representative ‘post-duplication’ species to determine whether enzyme function changed after duplication. Protein products of each *DFR* copy from these four taxa were assayed for enzymatic activity on three common *DFR* substrates (dihydrokaempferol, dihydroquercetin and dihydromyricetin) and two substrates less commonly reduced by plant *DFR* enzymes (naringenin and eriodyctiol).

Three major patterns arise from this analysis. First, the pre-duplication copies exhibit moderate activity on DHK, DHQ and DHM (Fig. 2c–e), very low activity on eriodyctiol (Fig. 2b) and minimal activity on naringenin (Fig. 2a). Second, the post-duplication A and C copies exhibit essentially no activity on any of the five substrates (Fig. 2). Moreover, *DFR-A* and *-C* activities on eriodyctiol, DHK, DHQ and DHM are significantly reduced compared with activity in the pre-duplication copies (see Supplementary Methods for statistical analysis). Third, activity on all five substrates is higher for the post-duplication *DFR-B* copies than for the pre-duplication copies, all but DHQ significantly so even after a Bonferroni correction for multiple comparisons (see Supplementary Methods). The increased *DFR-B* activities on the molecularly smaller naringenin (4.6-fold), eriodyctiol

(6.1-fold) and DHK (7.9-fold) are substantially greater than the increases on DHQ (1.4-fold) and DHM (1.4-fold).

These results indicate that adaptive change in *DFR* immediately subsequent to the first duplication is best explained as escape from adaptive conflict. Adaptive change apparently occurred in both copies after the first duplication, consistent with expectations under EAC. Adaptive change in the A/C copy is indicated by the signature of repeated positive selection in the lineage subtending the A/C clade. Although such a signature of positive selection was not detected on the B copy, enzyme functional analysis indicates that adaptive substitution likely occurred. In particular, after duplication, enzyme activity on ancestral flavonoid substrates increased substantially. Because most mutations are believed to have detrimental effects on enzyme stability or activity¹³, we believe that neutral evolution in the B-lineage can be rejected in favour of positive selection that is undetected at the sequence level, possibly because it involved few substitutions.

Further evidence for adaptive change along the B-lineage is provided by an analysis of the substitutions that occurred at its base. The crystal structure of *DFR* from *Vitis vinifera* shows that Asn 133 lines

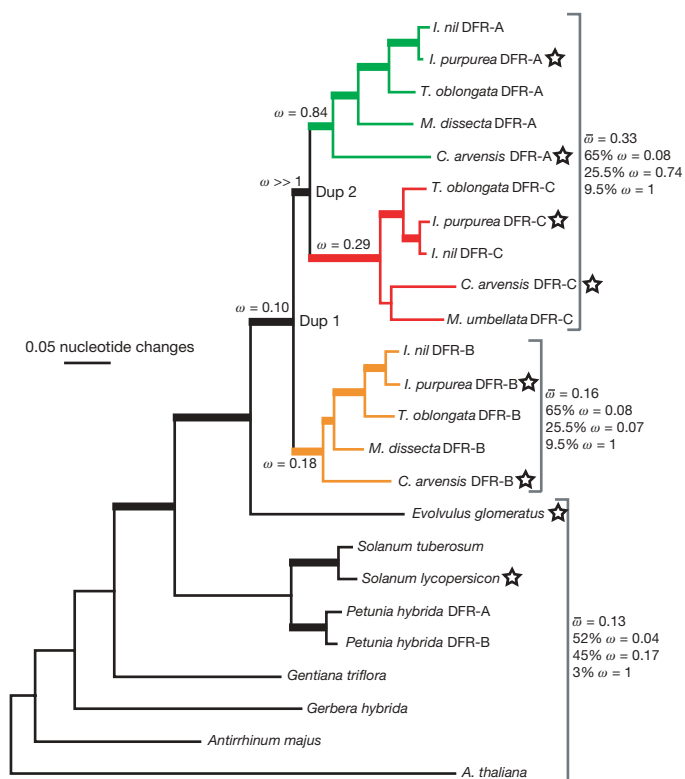


Figure 1 | Gene tree of sampled *DFR* copies from Convolvulaceae and outgroup taxa. Topology is identical under bayesian, maximum likelihood and maximum parsimony criteria. Thick branches indicate clades with strong statistical support (bayesian posterior probability greater than 0.95). Starred genes are those included in enzymatic studies. Results of selection tests are reported for tested branches, as are the proportion of sites under particular selective regimes for the *DFR-B*, *DFR-A/C* and outgroup clades. ω is the ratio of non-synonymous to synonymous nucleotide substitutions.

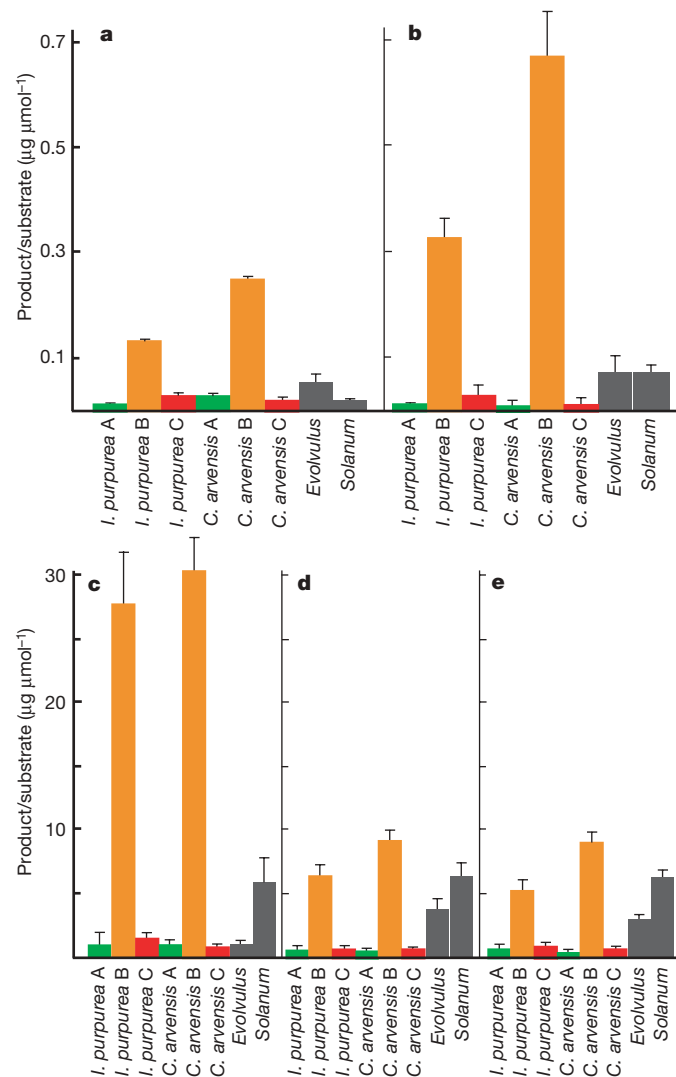


Figure 2 | Results of enzyme activity assays. Substrates are naringenin (a), eriodyctiol (b), dihydrokaempferol (c), dihydroquercetin (d) and dihydromyricetin (e). These correspond to products apigeninidin and luteolinidin (3-deoxyanthocyanidins, top chart) and pelargonidin and delphinidin (bottom chart). Results are in micrograms of product per micromole of substrate. Colours reflect coloured orthologue clades as in Fig. 1. Bars are one standard error; $n = 4$ for each assay.

the substrate-binding pocket and forms a hydrogen bond with the 4' hydroxyl (and, if present, the 3' hydroxyl) of the substrate B-ring¹⁴. DFR from maize¹⁰ and *Pyrus*⁹ has activity on naringenin, eriodyctiol, DHK and DHQ (DHM activity was not assayed in either of these studies). Both of these DFR variants have asparagine in the 133 position; in fact, nearly all known DFR genes from angiosperms have Asn133. Our phylogenetic reconstruction reveals that an Asn133 Asp substitution occurred before the divergence of the Gentianales and Solanales (Supplementary Fig. 1); this aspartate is retained in our two assayed single-copy species and correlates with relatively poorer flavonoid function. DFR-A copies from different species have different non-Asn amino acids at this site, and all DFR-C copies have the hydrophobic isoleucine; substituting the hydrophobic leucine at residue 133 has been shown to eliminate flavonoid activity in *Gerbera* DFR¹⁵. After gene duplication, the Asn133 variant re-evolved in the DFR-B lineage, coinciding with the improvement of DFR-B activity on the five substrates examined. Given this pattern, it is tempting to suggest that the original substitution of Asp for Asn at the base of the Gentianales and Solanales reflects the acquisition by the single-copy DFR of the unknown second function, and that the reverse transition in the DFR-B lineage reflects loss of this function in this lineage. Moreover, the failure to regain Asn in DFR-A and DFR-C is consistent with the loss of the ability of these copies to metabolize dihydroflavonols.

The improvement of ancestral function in the DFR-B lineage satisfies the second criterion for the operation of EAC. Apparently, the ability of DFR to reduce dihydroflavonols was constrained in the ancestral single-copy gene, and this constraint was released when the first duplication occurred.

Ideally, if EAC explains evolutionary change in duplicated copies of DFR, one should be able to demonstrate that the different copies perform different functions and that both of these functions were performed by the ancestral single-copy DFR, albeit at reduced efficiency. Unfortunately, we are unable to confirm this expectation because we have not yet identified what function(s) the DFR-A and -C copies perform. Although it is clear that these copies have lost the ability to perform one ancestral function of DFR (reduction of dihydroflavonols), we do not have direct evidence that they perform an ancestral rather than a novel function. Nevertheless, the improvement of flavonoid function in the DFR-B copies points to this function being constrained in the ancestral single-copy gene. Escape from adaptive conflict is thus clearly a feature of DFR evolution.

A potential alternative explanation for our results is that the activity spectrum of DFR-B reflects the ancestral state of DFR, and that the loss of ability to metabolize naringenin and eriodyctiol, as well as reduced function on DHK, DHQ and DHM, in *E. glomeratus* and *S. lycopersicon* reflects independent losses of activity. However, an ancestral state reconstruction of substrate specificity does not support this interpretation: an analysis using all species for which activity on all five flavonoid substrates were tested indicates that although ability to use all five substrates is deeply ancestral in the angiosperms, the ability to use naringenin and eriodyctiol was significantly reduced on the lineage leading to *Solanum*, *Evolvulus* and *Ipomoea* (Supplementary Fig. 2). Loss of function on these two substrates occurred before the *DFR* duplication in *Ipomoea*. This pattern is consistent with the EAC hypothesis: the reconstruction demonstrates first a reduction of ability to metabolize dihydroflavonols, naringenin and eriodyctiol, corresponding to when the single-copy DFR gained an additional (currently unknown) function; and then an increase in the ability to metabolize these compounds after duplication, corresponding to release from adaptive constraint.

Our analysis illustrates the larger issue that many published examples of NEO-F may actually represent EAC. There are some cases of NEO-F that by the two criteria used in our analysis are better explained by NEO-F than by EAC^{16,17}. However, there are also many purported cases of NEO-F that have not characterized ancestral function before duplication, that have not determined whether adaptive

evolution occurred in one or in both duplicate copies subsequent to duplication, and that have not determined whether either copy exhibits improvement of known ancestral function (for example, the origin of vertebrate MHC¹⁸ and plant R-gene alleles¹⁹). In these cases, EAC cannot be excluded. Consequently, it remains unclear whether the absence of reported cases of EAC indicates a true rarity of its occurrence, or just that much of what is considered NEO-F is actually EAC.

Unfortunately, the literature is silent on the expected relative frequencies of these two processes. Although some models predict that the relative importance of NEO-F and subfunctionalization depends on factors such as population size^{20,21}, these results cannot necessarily be extrapolated to either a comparison of NEO-F and EAC or of EAC and subfunctionalization. Although EAC resembles subfunctionalization in that it results in the partitioning of ancestral functions, subfunctionalization is envisioned as involving only neutral or slightly deleterious mutations²², whereas EAC involves primarily adaptive substitutions. Similarly, although NEO-F involves adaptive mutations in one copy of a duplicate, EAC involves adaptive mutations in both copies. Neither of these features of EAC is incorporated into any existing models of the fate of duplicate genes. Intuitively, we suspect that the probability of preservation of duplicate copies is higher for EAC than for NEO-F, given that the duplicated gene has already acquired multiple functions, because mutations in either copy can start the EAC process. However, the overall occurrence of EAC compared with NEO-F will also depend upon the fraction of duplicated genes that have acquired new functions before duplication, a quantity for which no information exists.

That the *DFR-B* copy also evolved the capability of reducing naringenin and eriodyctiol is surprising because the presumed downstream products of these reactions are rare in angiosperms and unreported from the Solanales²³. It is therefore doubtful that this new capacity is actually used in post-duplication *Ipomoea* species. Nevertheless, 3-deoxyanthocyanins and their polymers have been reported in some crops to have insecticidal and anti-fungal properties²⁴, which suggests that the ability of DFR-B to synthesize their precursors may constitute an exaptation for the evolution of such resistance. Under this interpretation, improvement of DFR-B activity on the more common substrates DHK, DHQ and DHM may have been the adaptive change whereas improvement in flavonol reduction was a byproduct. Interestingly, this result indicates that the evolution of novel function may sometimes be a byproduct of the evolution in enhanced ancestral activity.

METHODS SUMMARY

We used PCR to isolate *DFR* copies from species sampled across the Convolvulaceae and Solanaceae in an effort to identify when on the known species phylogeny the duplication events occurred. Gene trees were reconstructed using maximum parsimony, maximum likelihood and bayesian criteria, and branch support was assessed by bootstrapping and by examining posterior probabilities of nodes. To determine whether changes in selection pressures followed inferred duplication events, we used codon-based models of substitution as implemented in PAML. Specifically, we performed lineage and clade-based tests of each *DFR* copy's subtending lineage and entire clade, as well as branch-sites tests and clade-sites tests.

We subcloned each *DFR* copy identified in *S. lycopersicon*, *E. glomeratus*, *C. arvensis* and *I. purpurea* into bacterial overexpression vectors and then expressed these constructs in *Escherichia coli* BL-21 cell lines. After protein extraction, each enzyme was assayed for substrate specificity on five compounds: DHK, DHQ, DHM, naringenin and eriodyctiol. The products of these reactions were purified and then quantified spectrophotometrically.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 15 November 2007; accepted 14 May 2008.

Published online 25 June 2008.

1. Haldane, J. B. S. The part played by recurrent mutation in evolution. *Am. Nat.* **67**, 5–19 (1933).

2. Hahn, M. W., De Bie, T., Stajich, J. E., Nguyen, C. & Cristiani, N. Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome Res.* **15**, 1153–1160 (2005).
3. Lynch, M. & Conery, J. S. The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151–1155 (2000).
4. Ohno, S. *Evolution by Gene Duplication* (Springer, Berlin, 1970).
5. Hughes, A. L. The evolution of functionally novel proteins after gene duplication. *Proc. R. Soc. Lond. B* **256**, 119–124 (1994).
6. Piatigorsky, J. & Wistow, G. The recruitment of crystallins: new functions precede gene duplication. *Science* **252**, 1078–1079 (1990).
7. Hittinger, C. T. & Carroll, S. B. Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* **449**, 677–681 (2007).
8. Schwartz-Sommer, Z. *et al.* Influence of transposable elements on the structure and function of the *A1* gene of *Zea mays*. *EMBO J.* **6**, 287–294 (1987).
9. Fischer, T. C., Halbwirth, H., Meisel, B., Stich, K. & Forkman, G. Molecular cloning, substrate specificity of the functionally expressed dihydroflavonol 4-reductases from *Malus domestica* and *Pyrus communis* cultivars and the consequences for flavonoid metabolism. *Arch. Biochem. Biophys.* **412**, 223–230 (2003).
10. Halbwirth, H., Martens, S., Wienand, U., Forkmann, G. & Stich, K. Biochemical formation of anthocyanins in silk tissue of *Zea mays*. *Plant Sci.* **164**, 489–495 (2003).
11. Inagaki, Y. *et al.* Genomic organization of the genes encoding dihydroflavonol 4-reductase for flower pigmentation in the Japanese and common morning glories. *Gene* **226**, 181–188 (1999).
12. Yang, Z. PAML, a program for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556 (1997).
13. DePristo, M. A., Weinreich, D. M. & Hartl, D. L. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nature Rev. Genet.* **6**, 678–687 (2005).
14. Petit, P. *et al.* Crystal structure of grape dihydroflavonol 4-reductase, a key enzyme in flavonoid biosynthesis. *J. Mol. Biol.* **368**, 1345–1357 (2007).
15. Johnson, E. T. *et al.* Alteration of a single amino acid changes the substrate specificity of dihydroflavonol 4-reductase. *Plant J.* **25**, 325–333 (2001).
16. Benderoth, M. *et al.* Positive selection driving diversification in plant secondary metabolism. *Proc. Natl Acad. Sci. USA* **103**, 9118–9123 (2006).
17. Zhang, J., Zhang, Y.-P. & Rosenberg, H. F. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nature Genet.* **30**, 411–415 (2002).
18. Li, W.-H. *Molecular Evolution* (Sinauer, Sunderland, Massachusetts, 1997).
19. Michmore, R. W. & Meyers, B. C. Clusters of resistance in plants evolve by divergent selection and a birth-and-death process. *Genome Res.* **8**, 1113–1130 (1998).
20. Lynch, M. & Force, A. The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**, 459–473 (2000).
21. Lynch, M., O'Hely, M., Walsh, B. & Force, A. The probability of preservation of a newly arisen gene duplicate. *Genetics* **159**, 1789–1804 (2001).
22. Force, A. *et al.* Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531–1545 (1999).
23. Iwashina, T. The structure and distribution of the flavonoids in plants. *J. Plant Res.* **113**, 287–299 (2000).
24. Nicholson, R. L., Kollipara, S. S., Vincent, J. R., Lyons, P. C. & Cadena-Gomez, G. Phytoalexin synthesis by the sorghum mesocotyl in response to infection by pathogenic and nonpathogenic fungi. *Proc. Natl Acad. Sci. USA* **84**, 5520–5524 (1987).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Grotewold and B. Winkel for technical advice, and R. Hopkins for implementation of statistical models. This work was supported by National Science Foundation funding. D.L.D. was supported in part by a National Institutes of Health training grant to Duke University's University Program in Genetics and Genomics.

Author Contributions D.L.D. and M.D.R. designed the research; D.L.D. performed the research and the analyses; D.L.D. and M.D.R. wrote the paper.

Author Information The new DNA sequences reported here are deposited in GenBank under accession numbers EU189072–EU189082. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.L.D. (dld3@duke.edu) or M.D.R. (mrausher@duke.edu).

METHODS

Plant materials and extractions. We used the phylogeny of Stefanovic *et al.*²⁵ to target our sampling to a set of species that would help determine the timing of evolutionary transitions from a single-copy *DFR* ancestor to the three-copy family observed in *I. purpurea* and *I. nil*¹¹. We focused on five species representing the phylogenetic diversity of the Convolvulaceae: *I. purpurea*, *C. arvensis*, *E. glomeratus*, *Merremia dissecta* and *Merremia umbellata*, and *S. lycopersicon* from the sister family Solanaceae. All plants were grown from seed. DNA and RNA were extracted from freshly harvested buds, leaves and stem tissue using the DNeasy Plant Mini kit and RNeasy kit (Qiagen), respectively, according to the manufacturer's protocol.

Cloning and sequencing of *DFR* copies. To isolate *DFR* copies from previously unstudied taxa, we designed primers based on a consensus of known *DFR* copies from the Convolvulaceae. Initially, forward primer DFR59F: 5'-GCGTCACCGGAGCTGCTGG-3' was paired with DFR405R: 5'-GAGGAAGTGAACCAGC-3' or the GeneRacer oligo dT in 3' rapid amplification of complementary DNA (cDNA) ends (RACE) reactions using either genomic DNA or cDNA synthesized from total RNA as template. Gel-extracted PCR products were cloned into TOPO-TA 2.1 vectors (Invitrogen). Individual clones were sequenced using BigDye version 3.0 (Applied Biosystems) and visualized on ABI 3730 sequencers. To amplify the variable 5' ends of the *DFR* copies, we designed nested gene-specific primers and carried out 5' RACE reactions using the 5' RACE System (Invitrogen) according to the manufacturer's specifications. To confirm all copies from a given taxon had been identified, we also used multiple pairs of degenerate primers in PCRs using independently synthesized cDNA pools as template (primer sequences available on request). Gene-specific primers were then used to amplify the full-length cDNAs of each paralog, which were cloned into TOPO-TA 2.1 vectors and transformed into Top10 cells (Invitrogen) for storage at -80°C .

Phylogenetic analysis of gene-tree topology, changes in selection and timing of duplication. We combined new sequences collected in the present study with *Ipomoea* sequences available in GenBank and several outgroups to make a single alignment. Because the amino- and carboxy-terminal regions were highly variable, we were unable to discern positional homology with reasonable certainty. These regions (consensus nucleotide positions 13–66 and 1054–1320) were excluded from all subsequent analyses. Alignments are available on request.

The optimal tree and estimates of branch-support were determined using a bayesian Markov chain Monte Carlo (MCMC) approach in the program MrBayes 3.1.2 (ref. 26). In bayesian analyses, we used two models. In the first model, all sites were included in one partition assumed to evolve according to the general time reversible (GTR) with gamma (+G) model. In the second model, one partition contained all first and second codon positions and a second partition contained all third positions, treating each partition independently with its own GTR+G parameters. MCMC analyses were started with flat priors and run for 2 million generations, a burn-in was discarded and consensus trees were determined using MrBayes. At the conclusion of each MCMC run, we estimated that the chains had converged upon a stable set of parameters by calculating the potential scale reduction factor using MrBayes. Optimal topologies were also estimated using maximum parsimony and maximum likelihood criteria (see Supplementary Methods).

We determined if changes in selection followed the origin of duplicate copies of *DFR* using the codeml program in the PAML 3.15 package¹² to perform lineage and clade-specific analyses of K_a/K_s ratios (ω). We first tested a series of nested models that allowed particular branches to have ω independent of

background ratios: the lineage subtending all duplicated *DFR* copies, the lineage subtending the *DFR-B* clade, the lineage subtending the *DFR-A/C* clade, the lineage subtending the *DFR-A* clade and the lineage subtending the *DFR-C* clade. We also tested each of these branches for a signature of positive selection using the branch-sites test of Zhang *et al.*²⁷, and each of the clades arising from these branches using the clade-sites test of Yang *et al.*²⁸. We applied a sequential Bonferroni correction to account for the multiple comparisons made in these analyses and report only results that were significant after this correction (see Supplementary Methods).

In vitro assays of enzyme function. Full-length cDNA transcripts of each *DFR* copy from *I. purpurea*, *C. arvensis*, *E. glomeratus* and *S. lycopersicon* were amplified with Invitrogen Pfx polymerase and cloned directly into pENTR/D-TOPO vectors and then recombined into pDEST17 vectors using Clonase 2.1 (Invitrogen) according to the manufacturer's specifications. Plasmids containing inserts in the correct orientation and coding frame were transformed into BL21 Star (DE3) *E. coli* for overexpression. After overnight growth in 5 ml LB-Amp₁₀₀, we inoculated 200 ml of each culture and grew the cells at 37°C and 200 r.p.m. until mid-log phase ($\text{OD}_{600} = 0.4$). To induce expression, isopropyl- β -D-thiogalactopyranoside was added to a final concentration of 0.5 mM and the cultures were grown for 3 h at 37°C . Cultures were then spun down and the pellet resuspended in lysis buffer (50 mM HEPES, 50 mM NaCl, 1 mM EDTA, 1 mM PMSF, 100 μM TPCK, 100 μM TLCK), lysed for 30 min on ice with 1 mg ml⁻¹ lysozyme and pelleted again leaving the protein-containing supernatant.

Substrates and standards for the enzyme assays were obtained from commercial manufacturers: dihydrokaempferol (TransMit), dihydroquercetin (Alexis Biochemical), dihydromyricetin (Apin Chemicals), eriodictiol (Indofine), naringenin (Sigma), pelargonidin (Indofine), delphinidin (Polyphenols Laboratories AS), cyanidin (Indofine), apigeninidin (Apin) and luteolinidin (Apin). Enzyme extract (250 μl) was added to reaction mixtures for *in vitro* functional enzyme assays as described previously²⁹. For each gene construct, two replicates each of 30 min and 3 h time points were taken. The immediate product of these assays are unstable leucoanthocyanidins (DHK, DHQ and DHM as substrates) or flavan-4-ols (naringenin and eriodictiol as substrates). The former were converted to stable anthocyanidins as described²⁹ and the latter as described³⁰. Products and their appropriate standards were identified spectrophotometrically and by thin-layer chromatography and then quantified spectrophotometrically on a Shimadzu UV-2401PC. Two negative control assays were also run: one assayed overexpressed *I. purpurea* chalcone synthase enzyme, the other substituted lysis buffer for enzyme.

25. Stefanovic, S., Krueger, L. & Olmstead, R. G. Monophyly of the convolvulaceae and circumscription of their major lineages based on DNA sequences of multiple chloroplast loci. *Am. J. Bot.* **89**, 1510–1522 (2002).
26. Ronquist, F. & Huelsenbeck, J. MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
27. Zhang, J., Nielsen, R. & Yang, Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* **22**, 2472–2479 (2005).
28. Yang, Z., Wong, W. S. & Nielsen, R. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **22**, 1107–1118 (2005).
29. Stafford, H. A. & Lester, H. H. Enzymatic and nonenzymatic reduction of (+)-dihydroquercetin to its 3,4-diol. *Plant Physiol.* **70**, 695–698 (1982).
30. Stafford, H. A. Flavonoids and related phenolic compounds produced in the first internode of *Sorghum vulgare* Pers. in darkness and in light. *Plant Physiol.* **40**, 130–138 (1965).