The Rise of Chemodiversity in Plants
Jing-Ke Weng *et al.*
Science **336**, 1667 (2012);
DOI: 10.1126/science.1217411

This copy is for your personal, non-commercial use only.

12. G. A. Parker, J. Maynard Smith, *Nature* **348**, 27 (1990).
13. A. Bar-Even, A. Flamholz, E. Noor, R. Milo, *Nat. Chem. Biol.* **8**, 509 (2012).
14. E. Meléndez-Hevia, A. Isidoro, *J. Theor. Biol.* **117**, 251 (1985).
15. E. Noor, E. Eden, R. Milo, U. Alon, *Mol. Cell* **39**, 809 (2010).
16. R. U. Ibarra, J. S. Edwards, B. O. Palsson, *Nature* **420**, 186 (2002).
17. E. Dekel, U. Alon, *Nature* **436**, 588 (2005).
18. M. Eames, T. Kortemme, *Science* **336**, 911 (2012).
19. A. Bar-Even *et al.*, *Biochemistry* **50**, 4402 (2011).
20. J.-K. Weng, R. N. Philippe, J. P. Noel, *Science* **336**, 1667 (2012).
21. A. Bar-Even, E. Noor, R. Milo, *J. Exp. Bot.* **63**, 2325 (2012).
22. E. Pichersky, E. Lewinsohn, *Annu. Rev. Plant Biol.* **62**, 549 (2011).
23. L. G. Landry, C. C. Chapple, R. L. Last, *Plant Physiol.* **109**, 1159 (1995).
24. S. R. Long, *Plant Physiol.* **125**, 69 (2001).
25. N. K. Clay, A. M. Adio, C. Denoux, G. Jander, F. M. Ausubel, *Science* **323**, 95 (2009).
26. A. Weinhold, I. T. Baldwin, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7855 (2011).
27. P. Bednarek *et al.*, *Science* **323**, 101 (2009).
28. R. J. Hopkins, N. M. van Dam, J. J. van Loon, *Annu. Rev. Entomol.* **54**, 57 (2009).
29. B. A. Halkier, J. Gershenzon, *Annu. Rev. Plant Biol.* **57**, 303 (2006).
30. J. Zhao, C. C. Williams, R. L. Last, *Plant Cell* **10**, 359 (1998).
31. J. W. de Kraker, J. Gershenzon, *Plant Cell* **23**, 38 (2011).
32. M. Frey *et al.*, *Science* **277**, 696 (1997).
33. E. R. Radwanski, J. Zhao, R. L. Last, *Mol. Gen. Genet.* **248**, 657 (1995).
34. C. Sallaud *et al.*, *Plant Cell* **21**, 301 (2009).
35. A. L. Schilmiller *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 10865 (2009).

REVIEW

The Rise of Chemodiversity in Plants

Jing-Ke Weng, Ryan N. Philippe, Joseph P. Noel*

Plants possess multifunctional and rapidly evolving specialized metabolic enzymes. Many metabolites do not appear to be immediately required for survival; nonetheless, many may contribute to maintaining population fitness in fluctuating and geographically dispersed environments. Others may serve no contemporary function but are produced inevitably as minor products by single enzymes with varying levels of catalytic promiscuity. The dominance of the terrestrial realm by plants likely mirrored expansion of specialized metabolism originating from primary metabolic pathways. Compared with their evolutionarily constrained counterparts in primary metabolism, specialized metabolic enzymes may be more tolerant to mutations normally considered destabilizing to protein structure and function. If this is true, permissiveness may partially explain the pronounced chemodiversity of terrestrial plants.

Plants produce a repository of structurally diverse chemicals, which are traditionally known as secondary metabolites, because many of them are not directly involved in central metabolism (*1*). The expansion of chemodiversity associated with secondary metabolites mirrors the tremendous adaptability of terrestrial plants. For instance, phytohormones regulate various aspects of plant growth and development in response to environmental cues, whereas phenolics and waxy cuticles act as ultraviolet sunscreens and prevent desiccation. Plant polymers such as lignin, sporopollenin, and rubber provide mechanical support, gamete protection, and wound healing. A variety of compounds, from pigments and flavors to volatile scents and antimicrobials, mediate an array of interspecies interactions that seduce pollinators and seed dispersers or deter pathogens and herbivores. Unlike primary metabolites required for central metabolism, specialized compounds are often biosynthesized in response to environmental cues or as a consequence of growth and development. In short, it is likely that these phytochemicals shape the interdependencies and diversity of plant ecosystems forming the base of the global food chain.

To date, genome comparisons across the green plant lineage suggest that the expansion of plant-specialized metabolism occurred concurrently with the colonization of land by plants approximately 500 million years ago (*2*). Necessary metabolic processes (for instance, the biosynthesis of phenylpropanoids and sporopollenin, likely prerequisites for the colonization of terrestrial habitats) became established during that period (*3*). New metabolic branches continuously arose throughout land-plant evolution, resulting in a contemporary repertoire of specialized metabolites, some of which are shared across various taxonomic groups, whereas others exist only in a single species (*4*, *5*).

The Emergence of Metabolism

Primordial metabolism is postulated to have consisted of chemical intermediates interconnected by a smaller number of multifunctional catalytic proteins, peptides, and/or RNAs (Fig. 1A) (*6*). Since its origin as a fundamental property of the cell, metabolism is generally regarded as having evolved toward increasing order and catalytic efficiency (Fig. 1A). Presently, enzymes belong to a handful of protein families, possess catalytic precision and kinetic speed, employ a limited repertoire of substrates, and produce a correspondingly narrow range of products (*6*).

Although the number of specialized metabolites and the enzymes required for their biosynthesis continues to expand, the number of protein folds associated with these enzymes is relatively restricted. In contrast to primary metabolism, in which selection constrained mutations to maintain the most stable and functional enzyme forms, we hypothesize that specialized metabolic enzymes may have emerged through early gene duplication, followed by mutations that broadened substrate selection and flattened activation barriers of their catalyzed reactions. The resulting mechanistic elasticity allowed single enzymes to catalyze multiple reactions and biosynthesize multiple products (Fig. 1A). This scenario is consistent with directed evolution focused on enzyme promiscuity (*7*, *8*) and the biochemical characterization of mutant libraries derived from phylogenetic relationships in several plant-specialized metabolic enzyme families (*9–11*).

Phylogenetic analyses suggest that catalytic expansion among many plant-specialized metabolic enzyme families arose once, suggesting that the initial event(s) separating primary and specialized metabolism were either very rare or rarely not deleterious and able to be maintained and eventually fixed within the population and/or species. However, following events such as gene duplication, alleles may occasionally function under relaxed selection such that at least one copy is able to accumulate mutations leading to greater mechanistic elasticity and, ultimately, neofunctionalization before the emergence of inactivating mutations. Expanded substrate recognition, flattened catalytic landscapes, and, consequentially, multiple products from a single enzyme are common in specialized metabolism (Fig. 1B). This contemporary observation hints that genetic drift and gene flow across populations can contribute to chemodiversity in the absence of toxicity or compromised organismal fitness due to a subset of minor products. In other cases, selection could favor specific functions or bias the emergence of multifunctional enzymes due to the advantageous use of multiple substrates and/or the formation of a set of ecologically beneficial products from a single enzyme or metabolic pathway.

Once a duplication-derived progenitor emerged, mutations may have loosened the energetic interdependencies of residues within the protein

Howard Hughes Medical Institute, Jack H. Skirball Center for Chemical Biology and Proteomics, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA.

*To whom correspondence should be addressed. E-mail: noel@salk.edu

Plant Metabolism

fold, which were previously fixed in the absence of a paralog (*12*). Even deleterious changes appearing in one paralog may be tolerated and not eliminated by selection, when the other paralog contributes to fitness. In such cases, the evolution of advantageous activities can now be favored in new environments. Neutral or deleterious allelic variations may also be retained due to genetic hitchhiking when the affected genes reside near loci under positive selection. The process of attenuating energetic interdependencies within an ancestral protein fold in subsequent generations may occur over a sufficiently short period of time to prevent nonfunctionalization. This, in turn, reshapes protein stability and dynamics as well as

the enzyme catalytic properties, resulting in divergence of specialized enzymes from their origin in primary metabolism (Fig. 1B).

The chemically constrained catalytic landscapes of specialized enzymes that correlate sequence variation with catalytic properties bear little resemblance to those of primary metabolic enzymes. In primary metabolism, reactions are often catalyzed with high specificity accompanied by low levels of mechanistic elasticity; in short, the opposite of many specialized enzymes (Fig. 1B). Although protein structure is conserved in primary and secondary metabolism, increased catalytic promiscuity likely molded the evolution of specialized enzymes.

Supporting this view, a number of current specialized metabolic enzymes exhibit, on average, a greater ability to accept a broader range of substrates and to employ multiple energetically similar reaction mechanisms than related primary metabolic enzymes (*8*, *13–15*). Moreover, these enzymes seem to traverse functional space more easily than their structurally related cousins in primary metabolism to evolve new and often several metabolic products while retaining a modicum of their original function (*8*, *10*, *11*). Minimally, paralogs sporadically escaped nonfunctionalization to traverse functional space. Furthermore, specialized metabolic enzymes are ~30-fold less active than those of primary metabolism (*16*).
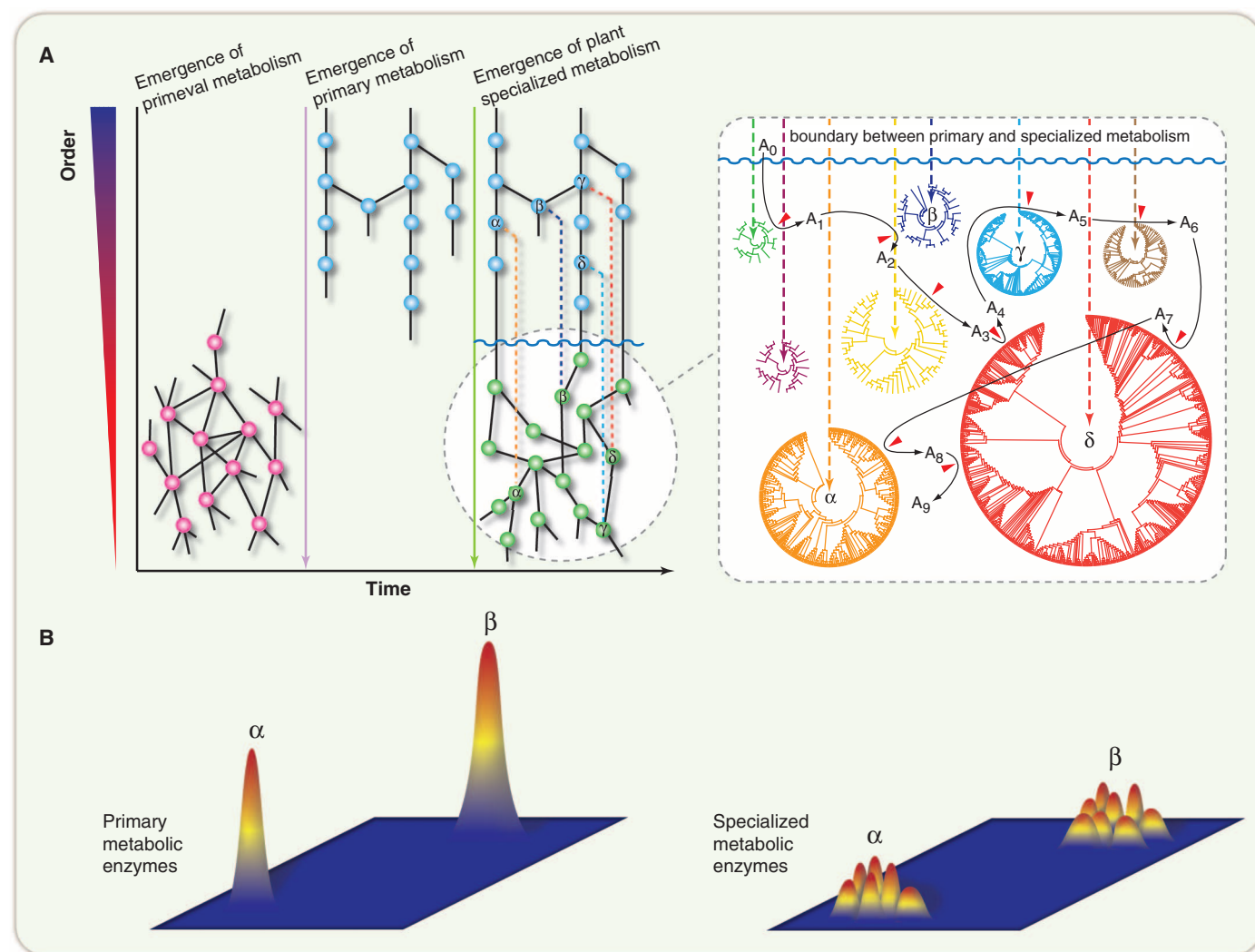
Fig. 1. Patterns of emergence and evolution of primary and specialized metabolism. (**A**) Primary metabolism likely arose from promiscuous primeval metabolic reactions and evolved toward greater catalytic precision and efficiency. Specialized metabolism likely emerged from primary metabolism. Due to early gene-duplication events, the functional constraints acquired by primary metabolic enzymes were released, allowing the mutational exploration of new areas of enzyme chemistry. Enzymes and reactions are represented by nodes (pink, blue, and green spheres) and links (black lines), respectively. The right panel illustrates the stepwise assembly of a specialized metabolic pathway using descendents from enzyme folds rooted in primary metabolism (indicated by circular phylogenetic trees and highlighted with Greek letters). Products of one reaction serve as substrates for another. Red arrowheads indicate the recruitment of single enzymes from protein families. (**B**) Hypothetical catalytic landscapes of primary and specialized metabolic enzymes relating sequence variation (horizontal plane) to the breadth of disparate enzymatic activities of stable protein folds (vertical axis). Catalytic specificity and efficiency for primary metabolic enzymes are maintained by natural selection, constraining their chemical mechanisms. Specialized metabolic enzymes often produce additional products from a single enzyme due to expanded substrate recognition and/or multiple chemical transformations within a single enzyme.

Diminished catalytic efficiency of multifunctional metabolic enzymes probably coincided with greater substrate permissiveness and the occurrence of several mechanistic routes to multiple products with little cost to the fitness of the host population. As long as the enzyme that must produce multiple products by virtue of its chemical mechanism yields at least one conferring a fitness advantage, the enzyme can be retained, barring issues of by-product toxicity. An enzyme does not have to evolve to perfection or absolute product specificity; it merely has to produce enough of the desired compound for the gene to be maintained in the population. As populations experience fluctuating abiotic and biotic ecological changes, one of the minor metabolites may also assume an advantageous function, thus resulting in fixation of the multifunctional paralog.

Molecular Exploitation of New Catalytic Space

Observed features of specialized metabolism include new catalysts emerging from progenitor enzymes catalyzing alternative reactions, or even from noncatalytic proteins. Positing that protein functional promiscuity serves as the starting point for functional innovation through natural selection (7, 11), this property of specialized metabolic enzymes may be key to the rapid expansion of these systems. In some cases, the ancestral promiscuous activity can be inferred using a combination of biochemical and phylogenetic information. For instance, the evolution of rosmarinic acid biosynthesis in Lamiaceae herbs arose from gene duplication of a BAHD acyltransferase, where the progenitor enzyme probably exhibited low but measurable activity against a noncanonical substrate. After a gene-duplication event, one gene copy likely was selected for increased activity toward this substrate, resulting in the emergence of a new metabolic step (Fig. 2A) (17, 18).

Refinement of a generalist ancestral enzyme into a catalytic specialist may also shape a metabolic trait. During anthocyanin biosynthesis in the *Iochroma* genus, dihydroflavonol reductase (DFR) catalyzes reduction of both dihydrokaempferol and its hydroxylated derivative dihydromyricetin. Thus, DFR serves two catalytic roles in parallel pathways resulting in red and blue pigments in flowers, respectively. In *I. gesnerioides*, DFR substrate recognition narrows substantially so that dihydrokaempferol is the preferred substrate, resulting in a derived red flower trait, deviating from the ancestral blue flower trait in the genus (Fig. 2B).

Moreover, enzyme families such as terpene synthases (TPSs) and type III polyketide synthases (PKS IIIs) exhibit a catalytic propensity to biosynthesize a multitude of products from a single enzyme (Fig. 2C) (9, 10). The ability of TPSs and PKS IIIs to produce numerous products correlates with the nature of their bond-forming reactions, shaped by the facile reactivity of their catalytic intermediates (19).

Recurring Patterns of Metabolic Evolution

The phenotypic outcome of an evolving plant-specialized metabolic system relies on the recruitment of multifunctional enzymes from several radiating enzyme families into new pathways (Fig. 1A). Recent technological advances allow us to view the breadth of specialized metabolic
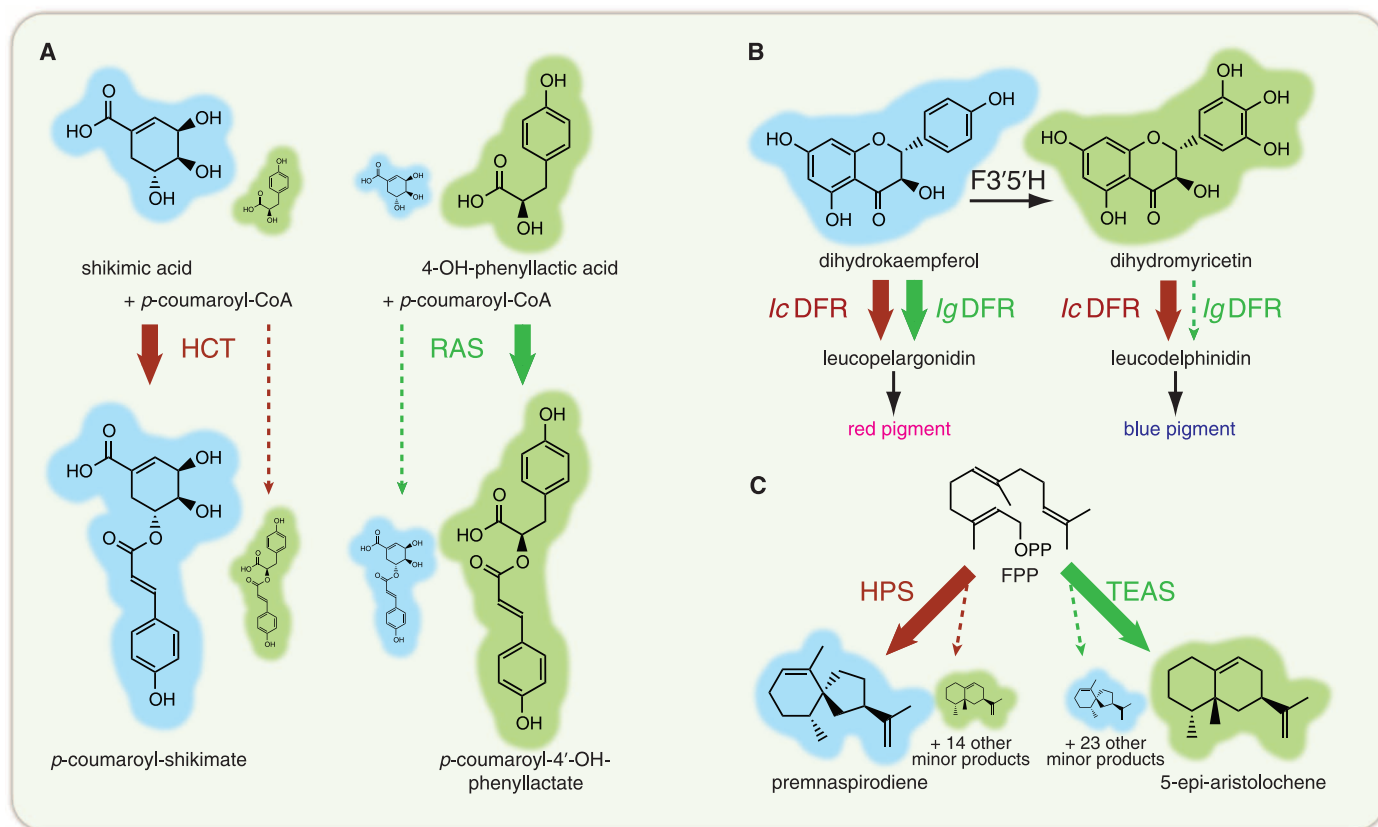
Fig. 2. Enzyme catalytic breadth underlies the expansion of chemodiversity in plant-specialized metabolism. (**A**) The emergence of rosmarinic acid synthase (RAS) in Lamiaceae likely followed substrate permissiveness of its evolutionary progenitor HCT, a more conserved enzyme ubiquitous in land plants. (**B**) By exploiting the broader substrate recognition of ancestral DFR, *I. gesnerioides* evolved a red flower color, deviating from the blue color common in the *Iochroma* genus. F3'5'H, flavonoid 3'5' hydroxylase. (**C**) *Hyocyamus muticus* premnaspirodiene synthase (HPS) and *Nicotiana tabacum* 5-epiaristolochene synthase (TEAS) produce a multitude of products intrinsic to the elevated reactivity of multiple chemical intermediates in the TPS family. In the TEAS/HPS subfamily, this relaxed specificity leads to a diversity of minor products and distinct major products that provide antimicrobial defense in the *Solanaceae*. OPP, pyrophosphate; FPP, farnesyl pyrophosphate.

Plant Metabolism

networks (*20*) and recognize recurring patterns of relaxed substrate recognition and energetically similar chemical mechanisms in individual enzymes affording incorporation into emerging metabolic pathways.

Typically, a handful of metabolites are co-opted by functionally diverse enzymes and serve as chemical "hubs" from which new metabolic paths often emerge in both primary and specialized metabolism. For example, acyl–coenzyme As (acyl-CoAs) serve as substrates for at least three enzyme families in primary and specialized metabolism. These include acyltransferases, NADH/NADPH-dependent reductases (NADH, the reduced form of nicotinamide adenine dinucleotide; NADPH, the reduced form of nicotinamide adenine dinucleotide phosphate), and ketoacyl synthases encompassing specialized PKS IIIs. In plant-specialized metabolism, the acyl-CoA, *p*-coumaroyl-CoA is a hub metabolite used by enzymes drawn from all three of these enzyme families, yielding structurally and functionally diverse phenylpropanoids. Moreover, metabolic pathways branching from these hubs are often taxonomically distributed, suggesting that at least some of these branches emerged after the initial chemical hubs were fixed in most organisms.

In addition to the recruitment of individual enzymes into emerging pathways, enzymes with expanded substrate recognition that act consecutively in a particular pathway can reappear, operating on disparate metabolites. For example, three catalytically sequential enzymes of the lignin biosynthetic pathway—hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyl transferase (HCT), *p*-coumaroyl shikimic acid 3′-hydroxylase, and caffeoyl CoA 3-*O*-methyltransferase—duplicated and underwent neofunctionalization in the Brassicaceae family such that they function in hydroxycinammoyl-spermidine biosynthesis (*18*). Similar recruitment of a segment of the lignin biosynthetic pathway also occurred independently in the Lamiaceae, resulting in rosmarinic acid biosynthesis (*21*).

In some cases, specialized metabolic pathways are encoded as gene clusters in plant genomes as seen in maize (*22*), rice (*23*), *Arabidopsis* (*24*), oat (*25*), and *Selaginella* (*26*), suggesting that the evolution of gene clustering of some metabolic pathways provides a selective advantage. Indeed, the clustering of metabolic genes in plants probably facilitates efficient inheritance, as these genes are less likely to be broken up by recombination. Moreover, the physical proximity of genes can coordinate transcription through additional genomic and epigenetic mechanisms.

Future Directions

Although a few studies have interrogated the minimum set of mutations that dictate the emergence of specific functions in divergent plant-specialized metabolic enzymes (*9, 10*), no particular study has addressed all viable mutational paths in these metabolic systems. This limits our ability to postulate evolutionary scenarios consistent with the stepwise assembly of mechanistically divergent metabolic pathways within the framework of Darwinian evolution and to quantify the incremental emergence of new activities with each mutational step. Could specialized metabolic enzymes and their pathways evolve along a wider set of evolutionary trajectories than their cousins in primary metabolism?

The lineage-specific birth of new metabolic pathways often involves neofunctionalization after gene duplication. Statistical coupling analysis (SCA), which measures covariation between pairs of amino acids on the basis of protein multiple-sequence alignments, can point to probable biophysical traits underpinning the emergence, expansion, and neofunctionalization of specialized metabolic enzyme families (*27*). The interconnected sets of covarying residues often form three-dimensional sector(s) that correlate with specific functionalities of a given protein family (*27*). The outcome of these analyses of primary and specialized metabolic enzymes sharing a common fold may provide biophysical hints to adaptive changes relating to substrate recognition, mechanistic elasticity, and protein structure and dynamics. Ultimately, covariation shapes an ensemble of dynamically accessible enzyme conformations in solution. These motions are undoubtedly linked with varying levels of relaxed catalytic trajectories often separating specialized metabolic enzyme families from their functionally constrained cousins in primary metabolism (*28*).

Ancestral sequence reconstructions, potentially guided by SCA, may offer additional insights into the evolutionary lineages of extant specialized metabolic enzymes (*11, 29*). The reconstructed ancestral sequences are unlikely to represent exact alleles that were fixed in the ancestral population. Nevertheless, if one chooses species with sound phenotypic relationships such as time of divergence, overlapping chemical repertoires, and comparable developmental programs, a collection of calculated sequences should reasonably approximate ancestral sequences. Given that the biochemical functions of many specialized enzymes are typically influenced by a small fraction of their total residues, the ancestral reconstructions may then serve as useful approximations of what might have functionally occurred during the emergence and expansion of particular specialized enzyme families.

Given the widespread occurrence of catalytic promiscuity in specialized metabolism, it is also important to consider that enzymes possessing mechanistic elasticity use varied substrates and produce diverse products to create pools of chemicals that may not be directly selected against. In essence, certain (currently) nonuseful chemicals can be found in a plant due to catalytic linkage, as the enzyme producing a beneficial compound inevitably synthesizes by-products due to the high intrinsic reactivity of chemical intermediates accompanying its catalytic mechanism.

The remarkable chemodiversity in plants and its underlying metabolic diversity are reached via exploration of sequence space restrained by enzyme catalysis, protein stability, emerging and extant metabolic pathways, and, ultimately, organismal fitness. The ability to bridge the fields of evolutionary biology, chemistry, biophysics, and mechanistic enzymology to cooperatively tackle the complexity of specialized metabolism will provide a more informed understanding of the amazing tapestry of plant-specialized metabolites that are so essential to the sessile lifestyle of plants.

References and Notes

1. G. S. Fraenkel, *Science* **129**, 1466 (1959).
2. J. A. Banks *et al.*, *Science* **332**, 960 (2011).
3. J. K. Weng, C. Chapple, *New Phytol.* **187**, 273 (2010).
4. S. E. O'Connor, J. J. Maresh, *Nat. Prod. Rep.* **23**, 532 (2006).
5. R. Croteau, R. E. Ketchum, R. M. Long, R. Kaspera, M. R. Wildung, *Phytochem. Rev.* **5**, 75 (2006).
6. R. Fani, M. Fondi, *Phys. Life Rev.* **6**, 23 (2009).
7. A. Aharoni *et al.*, *Nat. Genet.* **37**, 73 (2005).
8. Y. Yoshikuni, T. E. Ferrin, J. D. Keasling, *Nature* **440**, 1078 (2006).
9. M. B. Austin, M. E. Bowman, J. L. Ferrer, J. Schröder, J. P. Noel, *Chem. Biol.* **11**, 1179 (2004).
10. P. E. O'Maille *et al.*, *Nat. Chem. Biol.* **4**, 617 (2008).
11. R. Huang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 2966 (2012).
12. E. A. Ortlund, J. T. Bridgham, M. R. Redinbo, J. W. Thornton, *Science* **317**, 1544 (2007).
13. E. K. Lim *et al.*, *J. Biol. Chem.* **276**, 4344 (2001).
14. L. Kienow *et al.*, *J. Exp. Bot.* **59**, 403 (2008).
15. J. G. Kopycki *et al.*, *J. Mol. Biol.* **378**, 154 (2008).
16. A. Bar-Even *et al.*, *Biochemistry* **50**, 4402 (2011).
17. C. Landmann *et al.*, *Planta* **234**, 305 (2011).
18. M. Matsuno *et al.*, *Science* **325**, 1688 (2009).
19. M. B. Austin, P. E. O'Maille, J. P. Noel, *Nat. Chem. Biol.* **4**, 217 (2008).
20. A. L. Schilmiller, E. Pichersky, R. L. Last, *Curr. Opin. Plant Biol.* **15**, 338 (2012).
21. M. Petersen *et al.*, *Phytochemistry* **70**, 1663 (2009).
22. M. Frey *et al.*, *Science* **277**, 696 (1997).
23. K. Shimura *et al.*, *J. Biol. Chem.* **282**, 34013 (2007).
24. B. Field, A. E. Osbourn, *Science* **320**, 543 (2008).
25. X. Qi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8233 (2004).
26. J. K. Weng, T. Akiyama, J. Ralph, C. Chapple, *Plant Cell* **23**, 2708 (2011).
27. N. Halabi, O. Rivoire, S. Leibler, R. Ranganathan, *Cell* **138**, 774 (2009).
28. N. Tokuriki, D. S. Tawfik, *Science* **324**, 203 (2009).
29. P. D. Williams, D. D. Pollock, B. P. Blackburne, R. A. Goldstein, *PLOS Comput. Biol.* **2**, e69 (2006).