

However, substantial progress has been made in both ameliorating important elemental limits to crop production and improving food safety by utilizing a molecular understanding of elemental accumulation (23, 24).

As an example of the former, analysis of wheat lines with an introgressed QTL for sodium tolerance identified an HKT1 allele that can extract  $\text{Na}^+$  from xylem sap and thereby prevent sodium translocation to the shoots (25). Although a difference in leaf sodium was evident in all environments tested, yield gains were only evident where concentrations of sodium in the soil were highest. This confirms that, in the absence of meter-scale environmental information, heritable ionic phenotypes are more informative for the adaptation of crop genotypes to high-sodium environment than yield and other complex traits that integrate many biological processes.

The shared transport of compounds containing arsenic and chemically similar molecules containing the nutrients silicon and phosphorus underlies both a global food safety crisis and its solution. In rice, silicic acid and phosphate transporters can also move arsenite and  $\text{AsO}_4^{3-}$ , respectively. Particularly in regions of Southeast Asia with high arsenic concentrations in groundwater, the promiscuity of these transporters is responsible for acute toxicity, disease, and shortened life spans because of dietary intake of arsenic via consumption of rice. Even in the United States, arsenic intake from rice increases breast cancer

risk (26). Because we have a molecular understanding of arsenic uptake, breeding (27) and transgenic modification of crops with a transporter that sequesters arsenic in the root (28) have the potential to improve food safety and the health of hundreds of millions of people.

The prediction of tolerance to sodium stress by element accumulation measurements demonstrates that ionomics can accelerate crop improvement. This is complicated by the many agroecological challenges that limit yield and our insufficient understanding of the trade-offs that result from adaptation to particular soil conditions. Fortunately, the problem of local adaptation has been solved by evolution many times over. Ionomics and genetic association studies in model organisms and crops will directly identify alleles that promote element uptake or exclusion by plants. Using precise quantitative phenotyping of the ionome to characterize variation in plant-soil interactions, we are on the cusp of adding a new dimension to our understanding of why and how particular plants occupy their positions in the landscape and adapting agriculture to marginal soils.

#### References and Notes

1. D. E. Salt, I. Baxter, B. Lahner, *Annu. Rev. Plant Biol.* **59**, 709 (2008).
2. I. Baxter, *Briefings Funct. Genomics* **9**, 149 (2010).
3. I. Baxter *et al.*, *PLoS Genet.* **6**, e1001193 (2010).
4. M. W. Horton *et al.*, *Nat. Genet.* **44**, 212 (2012).
5. C. M. Palmer, M. L. Gueriot, *Nat. Chem. Biol.* **5**, 333 (2009).

6. F. J. Zhao, J. F. Ma, A. A. Mehar, S. P. McGrath, *New Phytol.* **181**, 777 (2009).
7. N. Mitani-Ueno, N. Yamaji, F. J. Zhao, J. F. Ma, *J. Exp. Bot.* **62**, 4391 (2011).
8. M. J. Haydon *et al.*, *Plant Cell* **24**, 724 (2012).
9. I. R. Baxter *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12081 (2008).
10. E. Buescher *et al.*, *PLoS ONE* **5**, e11081 (2010).
11. A. Ghandilyan *et al.*, *J. Exp. Bot.* **60**, 1409 (2009).
12. M. A. Klein, M. A. Grusak, *Genome* **52**, 677 (2009).
13. R. P. Sankaran, T. Huguet, M. A. Grusak, *Theor. Appl. Genet.* **119**, 241 (2009).
14. I. Baxter *et al.*, *PLoS Genet.* **5**, e1000492 (2009).
15. D. Y. Chao *et al.*, *Plant Cell* **23**, 1061 (2011).
16. R. Caspary, *Jahrb. wissensch. Botanik* **4**, 101 (1865).
17. R. Munns, M. Tester, *Annu. Rev. Plant Biol.* **59**, 651 (2008).
18. A. Rus *et al.*, *PLoS Genet.* **2**, e210 (2006).
19. K. U. Brady, A. R. Kruckeberg, H. D. Bradshaw Jr., *Annu. Rev. Ecol. Evol. Syst.* **36**, 243 (2005).
20. J. M. Yost, T. Barry, K. M. Kay, N. Rajakaruna, *Am. J. Bot.* **99**, 890 (2012).
21. S. Wood, K. Sebastian, S. Scherr, *Pilot Analysis of Global Ecosystems: Agroecosystems* (International Food Policy Research Institute and the World Resources Institute, Washington, DC, 2000), pp. 45–54.
22. C. P. Vance, *Plant Physiol.* **127**, 390 (2001).
23. P. R. Ryan *et al.*, *J. Exp. Bot.* **62**, 9 (2011).
24. T. Sutton *et al.*, *Science* **318**, 1446 (2007).
25. R. Munns *et al.*, *Nat. Biotechnol.* **30**, 360 (2012).
26. D. Gilbert-Diamond *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20656 (2011).
27. G. J. Norton *et al.*, *New Phytol.* **193**, 650 (2012).
28. G. Duan, T. Kamiya, S. Ishikawa, T. Arao, T. Fujiwara, *Plant Cell Physiol.* **53**, 154 (2012).

**Acknowledgments:** The authors acknowledge NSF (IOS-1126950), U.S. Department of Energy (DE-EE0003046), and USDA-ARS intramural funds.

10.1126/science.1219992

## REVIEW

# Achieving Diversity in the Face of Constraints: Lessons from Metabolism

Ron Milo<sup>1</sup> and Robert L. Last<sup>2\*</sup>

Metabolic engineering of plants can reduce the cost and environmental impact of agriculture while providing for the needs of a growing population. Although our understanding of plant metabolism continues to increase at a rapid pace, relatively few plant metabolic engineering projects with commercial potential have emerged, in part because of a lack of principles for the rational manipulation of plant phenotype. One underexplored approach to identifying such design principles derives from analysis of the dominant constraints on plant fitness, and the evolutionary innovations in response to those constraints, that gave rise to the enormous diversity of natural plant metabolic pathways.

Metabolism meets two seemingly conflicting needs: responding dynamically to developmental and environmental changes while maintaining the homeostasis required by a living cell, organ, or whole organism.

<sup>1</sup>Department of Plant Sciences, Weizmann Institute of Science, Rehovot 76100, Israel. <sup>2</sup>Department of Biochemistry and Molecular Biology and Department of Plant Biology, Michigan State University, East Lansing, MI 48824, USA.

\*To whom correspondence should be addressed. E-mail: lastr@msu.edu

This challenge is especially acute for plants, which are sessile organisms that endure constantly changing environmental conditions over life spans ranging from weeks to hundreds of years. For example, carbon fixation and allocation in leaves responds dynamically to unpredictable changes in environment, with time scales ranging from minutes to months. Consistent with a need for rapid response, the turnover time of most key metabolites of central carbon metabolism is on the order of 1 s (1, 2).

Plant metabolic phenotypes are the result of hundreds of millions of years of evolutionary history, during which some ancestral metabolic networks were restructured to meet the demands of changing environments while others remained close to their evolutionary ancient forms. For example, changes in temperature and aridity led to dozens of independently evolved variants of C4 metabolism for carbon fixation, even as the core process of the Calvin-Benson-Bassham pathway—which uses ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) for carbon fixation—remained conserved (3–5). A current challenge in metabolism is to understand the physicochemical constraints on the structure and function of the metabolic network, and thereby gain insight into how evolution worked within these restrictions to shape the characteristics of extant plants.

## Beyond Tinkering: The Utility of Design Principles for Plant Metabolic Engineering

Metabolic engineering promises opportunities to increase yield in agriculture and produce chemicals at lower economic and environmental cost. Despite progress, the rate of success in moving from concept to agricultural production or microbial fermentor has fallen short of expectations. For example, tens to hundreds of millions of dollars were spent in the public and private sector in

# Plant Metabolism

efforts to increase RuBisCO carboxylase activity and thus improve photosynthetic productivity of C3 plants. These studies yielded information about the structure and function of the enzyme but did not achieve the desired improvement in its kinetic properties (6). RuBisCO's kinetic properties seem already to be optimized by evolution (7, 8). Thus, efforts in metabolic engineering have been limited by existing physicochemical constraints.

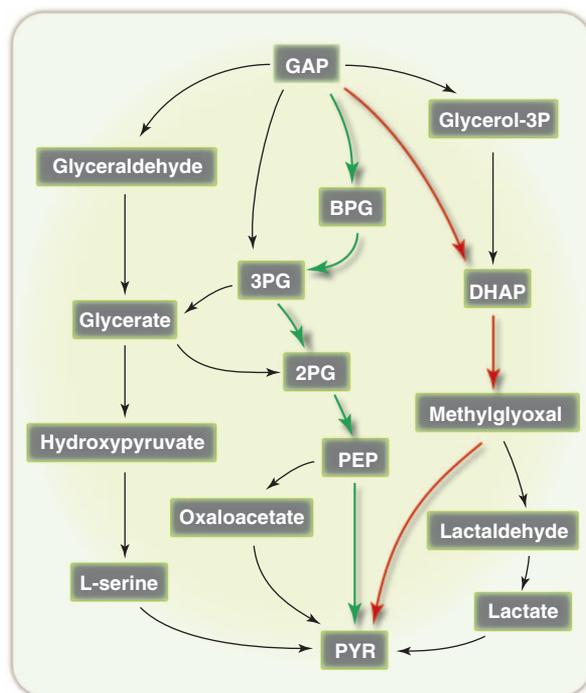
Metabolic engineering of a medically important plant metabolite, the anti-malarial artemisinin (9), is said to have required an investment of more than \$25 million and 150 person-years (10). One reason for the cost and complexity of the project is that it required the balanced expression of a large number of biosynthetic enzymes (11) to ensure a high production rate of the final product without leading to accumulation of deleterious reactive or toxic metabolic intermediates. Future metabolic engineering and synthetic biology projects demand better tools to predict which enzymes are key targets for adjusting expression and throughput.

To improve the efficiency and fulfill the promise of metabolic engineering, we must better understand the design and regulation principles governing metabolism. Understanding how evolution has led to designs that work well for the seemingly conflicting needs of stability and dynamic physiological responses might provide important clues.

## Identification of Dominant Constraints Through Optimality Models

Using the concept of optimality (12), a fitness function is defined—for example, the production of a metabolite in the minimum number of steps—and the landscape of possible solutions is analyzed. The goal is to find the solution with the highest fitness given defined constraints. A constraint might be avoidance of highly reactive intermediates. The metabolic pathway derived is compared to the natural solution; a close correspondence would suggest that the constraints and fitness function analyzed might indeed be relevant. In contrast, if no correspondence is seen, the analysis would indicate that this is not the dominant constraint or fitness objective shaping the system. Although most work on optimality models was done in microbes, this approach is being extended to plant metabolism.

The space of possible paths for transforming an initial compound into a product is theoretically immense. Even with the limitation of using known enzymes with their canonical set of transformations encapsulated in the Enzyme Commission classes (oxidoreductases, transferases, hydrolases, lyases, isomerases, and ligases), there is still a combina-



**Fig. 1.** The large number of alternative possibilities for metabolic transformations. Results from an analysis of the possible pathways for transforming glyceraldehyde-3-phosphate (GAP) into pyruvate (PYR), using the set of characterized enzymes from all organisms, are shown. The number of possibilities quickly increases as the number of steps becomes larger. The shortest theoretical path is denoted in red. The path used by Embden-Meyerhof-Parnas glycolysis is shown in green. 2PG, 2-phosphoglycerate; 3PG, 3-phosphoglycerate; PEP, phosphoenolpyruvate; DHAP, dihydroxyacetone phosphate.

torial explosion of possible paths. However, the set of pathways found in nature, while showing biochemical diversity, is much smaller.

As an example, Fig. 1 illustrates the conversion of the key glycolytic metabolite glyceraldehyde-3-phosphate (GAP) to pyruvate. This can be achieved in numerous ways with known enzymes; there are seven options that each make use of five enzymes (the number of steps in Embden-Meyerhof-Parnas glycolysis, shown in green in Fig. 1) and more than 70 when allowing seven enzymatic steps. Some of these options are shown in Fig. 1, with the shortest path, which uses the highly reactive intermediate methylglyoxal, highlighted in red. The different paths vary in terms of their adenosine triphosphate (ATP) yield; number of steps; reactivity, stability, and toxicity of the metabolic intermediates (e.g., methylglyoxal); and thermodynamic feasibility.

These are only some of the possible constraints for a successful metabolic pathway (13). Uncovering the importance of these constraints is central to understanding the structure of metabolism and a key element in designing novel biosynthetic pathways. Examples of additional constraints are flux kinetics, investment in enzymes (protein cost), affinity and specificity, and

whether intermediates are confined within subcellular compartments.

Analysis of the pentose phosphate pathway (PPP) (14) is an informative example of the application of optimality models to metabolism. In this analysis, the question was posed: How can five six-carbon sugars be converted into six five-carbon sugars with the minimal number of enzymatic steps? By choosing to constrain the possible reaction mechanisms to those catalyzed by known enzymes, the authors showed that the solution with the minimal number of steps is identical to the one arrived at by evolution. This analysis was recently extended to the central carbon metabolic network of *Escherichia coli* (15), in which the observed reaction network was found to connect the 13 precursor metabolites required to build biomass through the shortest paths. It remains to be seen whether optimality models can inform our understanding of pathways beyond central carbon metabolism, and whether these lessons can be tested and used in metabolic engineering.

In a very different approach, a flux balance analysis framework was used to predict the maximal possible yield of biomass per oxygen uptake rate on different carbon sources (16). This parameter was measured experimentally, leading to the conclusion that *E. coli* has close to the optimum predicted value for growth on glucose but is suboptimal for growth on glycerol. Lab evolution by propagation on glycerol for ~700 generations resulted in progressively improved growth, finally reaching the predicted optimum on glycerol.

The *E. coli* lac operon-encoded lactose utilization system is also being used to study factors contributing to optimality. One study investigated the impacts of altered expression of *lacZ*, the gene encoding the  $\beta$ -galactosidase enzyme, under varying concentrations of lactose (17). The protein cost is a quantification of the decrease in growth rate as a result of the expression of *lacZ* product; this could arise, for example, from ribosomes not being available for producing other proteins or because solvent capacity limits the total capacity of a cell to harbor proteins. The benefit measured is the increase in growth rate due to enhanced capacity to use lactose. Computationally maximizing the net difference between the cost and benefit gave predictions of optimal expression levels that varied according to the lactose concentration in the environment. These predictions were tested experimentally: After several hundred generations, *lacZ* expression levels adapted by mutation to achieve values close to those predicted. A recently published cost/benefit analysis of the lac operon presented evidence that activity, rather than expression, of the *lacY* lactose permease is the major physiological cost to the cell (18).

In addition to considering constraints and optimality principles of pathways and gene expression, it is informative to look at constraints on optimizing the building blocks of metabolism, enzymes. Some enzymes are considered optimal because they are able to perform catalysis at rates approaching the limit of diffusion; that is, the maximal turnover rate ( $k_{\text{cat}}$ ) divided by the substrate affinity ( $K_M$ ) is  $\sim 10^8$  to  $10^9 \text{ s}^{-1} \text{ M}^{-1}$ . However, very few metabolic enzymes are even close to this kinetic paragon status (19). A comprehen-

sive analysis of nearly 2000 measured enzymes shows a distribution of turnover rates divided by affinities that peaks at  $\sim 10^5 \text{ s}^{-1} \text{ M}^{-1}$ , three orders of magnitude below the diffusion limit constraint (Fig. 2). In fact, the median maximal turnover rate value over all measured enzymes is  $\sim 10 \text{ s}^{-1}$ —nowhere near the rates of  $10^4$  to  $10^5 \text{ s}^{-1}$  achieved by often-quoted record holders such as carbonic anhydrase and superoxide dismutase.

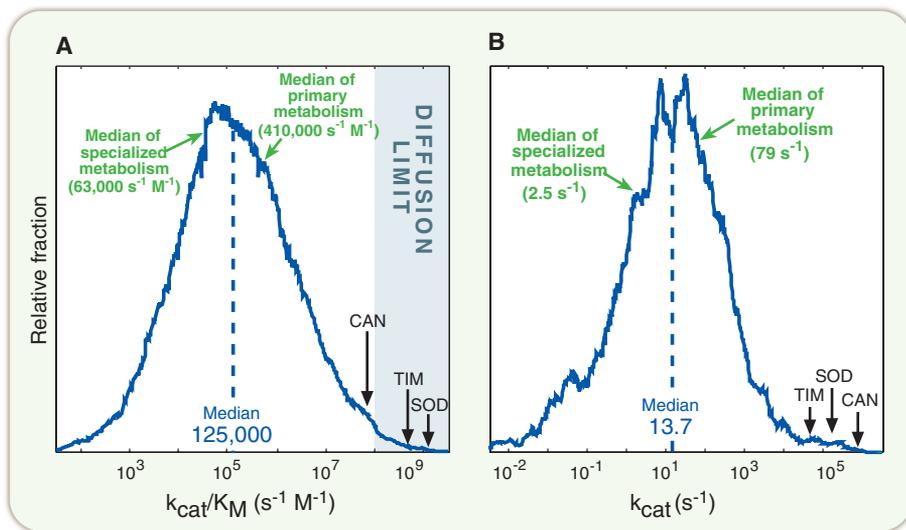
What limits typical enzymes from being better catalysts? The observed distributions suggest

constraints other than the diffusion limit—for example, limitations on the affinity toward metabolites of low molecular weight as a consequence of maintaining specificity and to discriminate between compounds of similar structure. The trends in parameter values also suggest that the required flux might be an important determinant: Enzymes involved in central carbon metabolism are on average  $\sim 30$  times faster than those involved in specialized metabolism (defined below and historically referred to as “secondary” metabolism; Fig. 2) (20). This might result from lower selective pressure for optimizing the kinetic parameters for maximal flux in specialized versus central metabolism.

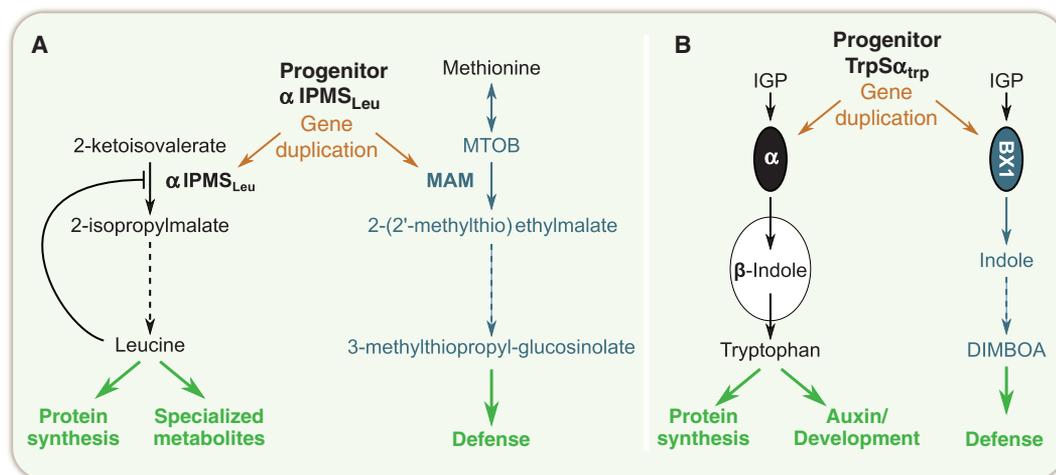
Optimality model approaches aim to identify the forces that shape and constrain biological systems, rather than to ask whether biology is optimal. These studies sharpen our insight into likely important influences in the evolution of metabolic systems, yet the solutions might be different when driving forces vary. For example, even the fundamental process of carbon fixation, typically performed by the Calvin-Benson-Bassham pathway, is carried out in at least six alternative ways in various organisms (21). Similarly, there are a number of variants of glycolysis including the Entner-Doudoroff pathway. Such diversity is often suggested to be related to selective pressures from the organism’s current or past ecological niche. Nowhere is this theme of metabolic diversity in the service of ecological adaptation more evident than in the realm of specialized metabolism.

### Constraints in the Land of Diversity: Evolving the Specialized Metabolism Buffet

In contrast to the products of central metabolism, specialized metabolites are diverse small molecules, each class of which is generally found in a subset of taxa. They are far more varied in structure than central metabolites, numbering in the hundreds of thousands of structures in the plant kingdom (22). Because of this complexity and their typical restriction to specific groups of plants, many enzymes of specialized metabolism are yet to be discovered. There is an increasing appreciation that these compounds serve a wide variety of physiological and ecological roles. For example, various roles for the taxonomically widespread and structurally diverse aromatic specialized metabolites have been documented, including the house-keeping function of protection against solar ultraviolet-B and specific signaling between legume roots and bacteria early in the establishment of symbiotic nitrogen fixation (23, 24). Specialized



**Fig. 2.** Distributions of kinetic parameters compiled from published enzyme data. Only values obtained with natural substrates were included in the distributions. (A) Enzyme specificity constant ( $k_{\text{cat}}/K_M$ ) values;  $N = 1882$ . (B) Enzyme turnover rate ( $k_{\text{cat}}$ ) values;  $N = 1942$ . Enzymes operating within primary and specialized metabolism show markedly different  $k_{\text{cat}}$  and  $k_{\text{cat}}/K_M$  values. Numbers in parentheses represent the median values for each group. Locations of several well-studied and often quoted rapid enzymes are highlighted: CAN, carbonic anhydrase; SOD, superoxide dismutase; TIM, triosephosphate isomerase. [Adapted from (19)]



**Fig. 3.** Evolution of two enzymes of specialized metabolism by gene duplication and neofunctionalization. Enzymes and products of specialized metabolism are shown in blue, primary metabolism in black. Dashed lines indicate multiple enzymatic steps. (A) The committing enzyme of glucosinolate biosynthesis, MAM, evolved by duplication of a progenitor  $\alpha$ -isopropylmalate synthase gene ( $\alpha\text{IPMS}_{\text{Leu}}$ ). MTOB, 4-methylthio-2-oxobutanoate. (B) The maize BX1 indole synthase activity evolved from a progenitor TrpS $\alpha$  subunit. DIMBOA, 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one.

metabolites fulfill a diversity of adaptive roles for plants (25–27) and also serve as a source of many of the most important therapeutic agents in traditional and modern medicine.

We have a better understanding of the range of constraints and possible solutions for primary metabolism than for specialized metabolism. Nonetheless, it is instructive to attempt to infer some major constraints and design principles, inspired by the approaches in microbes and plant central metabolism. We propose that this strategy will lead to a better fundamental understanding of these fascinating and diverse metabolic networks, while improving our ability to engineer stress tolerance and synthesis of products useful to humankind.

One constraint is that the antimicrobial or antiherbivory functions of many characterized plant specialized metabolites are often associated with toxicity and chemical reactivity. How can such highly reactive compounds be safely produced and stored for a time of need? Strategies used by plants for alleviating the constraint of producing these toxic compounds include sequestration in subcellular compartments (such as the vacuole) and production or transport to specialized organs or cell types. For example, the pungent taste of mustard and horseradish is the result of glucosinolates; defensive compounds containing a reactive SCN group are kept in an inactive form via a glucose moiety (28, 29). Tissue disruption by animal feeding unleashes the defensive activity of these compounds when glucosinolates come in contact with the enzyme myrosinase, and the combination rapidly produces nitriles, isothiocyanates, and other reactive and potentially noxious compounds. Thus, the plant avoids the toxic effects of glucosinolate degradation by sequestering the substrate metabolite and the enzyme in different cells or subcellular compartments until consumption by an herbivore. Another strategy used by many plants is storage of compounds in specialized structures. For example, epidermal hairs called glandular trichomes produce and store the metabolites that make basil and mint aromatic, whereas laticifers in opium poppy are the site of morphine accumulation.

Strong themes are emerging regarding the evolutionary and metabolic mechanisms for the synthesis of structurally diverse specialized metabolites. One is the recruitment of primary metabolic enzymes and pathways by gene duplication and acquisition of new enzymatic activities and biochemical regulatory mechanisms (e.g., altered allostery and protein-protein interaction, as discussed below). Another involves balancing the expression of central metabolic pathways with specialized metabolic pathway enzymes. For example, the regulation of amino acid biosynthetic pathways in plants is highly responsive to varied conditions that cause production of specialized metabolites from amino acids or their precursors (30).

Basing the production of specialized metabolites on the diversion of flux from central metabolism provides a source of abundant precursors but requires regulation at the branch point. One approach employed by evolution is duplication of a primary metabolic enzyme gene followed by alteration of enzyme activity and loss of regulation imposed by end-product feedback regulation on one of the resultant enzymes. An example is the evolution of methylthioalkylmalate synthase (MAM) in *Arabidopsis thaliana* and other mustard family plants, which leads to the protective glucosinolates (Fig. 3A) discussed above. This enzyme evolved from the committing enzyme of leucine biosynthesis: In addition to loss of feedback inhibition by leucine, the substrate specificity of the MAM enzyme has changed to catalyze the synthesis of glucosinolates (31). Together, these changes converted an enzyme of amino acid biosynthesis, which stringently regulates leucine production, into one that synthesizes a variety of alkyl glucosinolates. This example illustrates how relatively simple changes in protein structure and function can convert enzymes of central metabolism into those that divert flux for the production of specialized metabolites.

Indole synthase of maize (Bx1; benzoxazine-less1), derived from the tryptophan synthase  $\alpha$  (TrpS $\alpha$ ) subunit, is an example of a gene duplication event leading to evolution of an enzyme with an unexpected activity (Fig. 3B) (32). The ancestral amino acid biosynthetic TrpS $\alpha$  enzyme requires interaction with a partner subunit (tryptophan synthase  $\beta$ ; TrpS $\beta$ ) for catalytic activity in organisms as diverse as bacteria and plants (33). In fact, the TrpS $\alpha$  product, indole, is channeled to the TrpS $\beta$  active site without being released from the complex. In contrast, indole synthase is a variant form of TrpS $\alpha$  that no longer interacts with the partner TrpS $\beta$  and converts the tryptophan pathway intermediate indole-3-glycerolphosphate (IGP) to indole in the committing step of synthesis of insecticidal and fungicidal cyclic hydroxamic acids. Thus, the evolution of chemical defenses in maize and other grasses accomplished a feat of protein engineering by eliminating the requirement for interaction between TrpS $\alpha$  and TrpS $\beta$  to produce indole for use as a precursor in specialized metabolism.

As is the case for carbon fixation and C4 metabolism described above, production of the same or functionally related metabolites is often achieved through different biosynthetic routes (22). Can optimality analysis identify constraints on specialized metabolic pathways by analysis of the alternative routes?

Monoterpenes and sesquiterpenes (hydrocarbons with carbon chain lengths of 10 and 15, respectively) are widespread in nature. Each class is produced by two enzymatic steps from the same five-carbon substrates: isopentenyl diphosphate and dimethylallyl diphosphate. Despite the use of common intermediates, the pathways are found in

different compartments, and as a result, sesquiterpenes are typically derived from the cytosolic mevalonate pathway and monoterpenes from the plastidic deoxyxylulose 5-phosphate pathway. Glandular trichomes of cultivated tomato and a wild relative (*Solanum habrochaites* accession LA1777) produce and store large amounts of defensive monoterpenes or sesquiterpenes by analogous two-step pathways that are different from those of other organisms (34, 35). Both enzymes arose from proteins that do not normally participate in these pathways: *cis*-polyprenyldiphosphate synthase and diterpene synthase. An outcome of this novel approach to specialized metabolism is that the synthesis of trichome sesquiterpenes in the wild tomato takes place in the plastid and uses substrates from the plastidic deoxyxylulose 5-phosphate pathway, rather than the cytosolic mevalonate pathway.

Two testable hypotheses come from comparing the evolutionarily reengineered pathways in trichomes to conventional terpene biosynthesis: (i) Moving sesquiterpene biosynthesis to the plastid allows higher flux production of these defensive compounds; (ii) freeing trichome metabolism from the canonical terpene biosynthetic pathway could reduce constraints on diversification, allowing faster evolution of novel chemistries in response to changing predator and pathogen populations.

Although the discovery of new plant metabolic enzymes and pathways is still labor-intensive, modern technologies allow this process to advance at an increasingly rapid pace. Rational engineering of plant function requires going beyond documenting the parts of the broad metabolic network toward developing a deep understanding of the fundamental principles that govern metabolic regulation. We suggest that there is value in uniting the diverse approaches discussed in this review, including the use of varied computational methods to generate hypotheses. In addition to providing new insights into metabolic networks, such efforts are essential for predictive metabolic and synthetic engineering that will help meet the looming crisis in providing food, energy, and materials for the rapidly growing world population.

## References and Notes

1. M. Stitt, W. Wirtz, H. W. Heldt, *Biochim. Biophys. Acta* **593**, 85 (1980).
2. S. Arrivault *et al.*, *Plant J.* **59**, 826 (2009).
3. R. F. Sage, X. G. Zhu, *J. Exp. Bot.* **62**, 2989 (2011).
4. S. Weissmann, T. P. Brutnell, *Curr. Opin. Biotechnol.* **23**, 298 (2012).
5. S. von Caemmerer, W. P. Quick, R. T. Furbank, *Science* **336**, 1671 (2012).
6. S. M. Whitney, R. L. Houtz, H. Alonso, *Plant Physiol.* **155**, 27 (2011).
7. Y. Savir, E. Noor, R. Milo, T. Tlusty, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 3475 (2010).
8. G. G. Tcherkez, G. D. Farquhar, T. J. Andrews, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 7246 (2006).
9. D.-K. Ro *et al.*, *Nature* **440**, 940 (2006).
10. C. D. Smolke, P. A. Silver, *Cell* **144**, 855 (2011).
11. D. J. Pitera, C. J. Paddon, J. D. Newman, J. D. Keastling, *Metab. Eng.* **9**, 193 (2007).

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. Schillmiller *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 10865 (2009).

**Acknowledgments:** We thank A. Bar Even, A. Flamholz, A. D. Jones, E. Noor, A. Schillmiller, and T. Skaife for input on the manuscript. Research in R.L.L.'s group is funded by NSF grants DBI-1025636 and MCB-1119778. Research in R.M.'s lab is funded by the European Research Council (grant 260392) and by the Israel Science Foundation (grant 750/09). R.M. is the incumbent of the Anna and Maurice Boukstein career development chair.

10.1126/science.1217665

## 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



**Achieving Diversity in the Face of Constraints: Lessons from Metabolism**

Ron Milo and Robert L. Last (June 28, 2012)

*Science* **336** (6089), 1663-1667. [doi: 10.1126/science.1217665]

Editor's Summary

---

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

---

- Article Tools** Visit the online version of this article to access the personalization and article tools:  
<http://science.sciencemag.org/content/336/6089/1663>
- Permissions** Obtain information about reproducing this article:  
<http://www.sciencemag.org/about/permissions.dtl>

*Science* (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2016 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.