

Title:

The evolutionary paths towards complexity: a metabolic perspective

Author:

Jing-Ke Weng*

Affiliation:

Whitehead Institute for Biomedical Research and Department of Biology,
Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, MA
02142, USA

Keywords:

Specialized metabolism, divergent evolution, parallel evolution, convergent
evolution, catalytic promiscuity, enzyme evolvability, pathway assembly

*Correspondence author:

Jing-Ke Weng

Whitehead Institute for Biomedical Research and Department of Biology,
Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, MA
02142, USA

E-mail: wengj@wi.mit.edu

Summary

As sessile organisms, land plants have exploited their metabolic systems to produce a panoply of structurally and functionally diverse natural chemicals and polymers to adapt to challenging ecosystems. Many of these core and specialized metabolites confer chemical shields against a multitude of abiotic stresses, while others play important roles in plants' interactions with their biotic environments. Plant specialized metabolites can be viewed as complex traits in the sense that the biosynthesis of these molecules typically require multistep metabolic pathways comprised of numerous specific enzymes belonging to diverse protein fold families. Resolving the evolutionary trajectories underlying the emergence of these specialized metabolic pathways will impact a fundamental question in biology - how do complex traits evolve in a Darwinian fashion? Here, I discuss several general patterns observed in rapidly evolving specialized metabolic systems in plants, and surmise mechanistic features at enzyme, pathway and organismal levels that rationalize the remarkable malleability of these systems through stepwise evolution. Future studies, focused on fine sampling of metabolic enzymes and pathways in phylogenetically related plant species, or employing directed evolution strategies in synthetic systems, will significantly broaden our perspective on how biological complexity arises at the metabolic level.

Introduction

Complex traits are phenotypic features emerged from the interplay of multiple genetic components (Lander & Schork, 1994). Many complex traits are also described as irreducible, because the absence of any component of such a trait would abolish its overall function (Weber, 1999). Although complex traits are omnipresent in life at different biological scales, with examples ranging from the butterfly wing patterning to the assembly of multiprotein complexes (Beldade & Brakefield, 2002; Alber *et al.*, 2007), little is known about the stepwise trajectories through which complex traits arises in nature. Charles Darwin pondered this question in the *Origin of Species* using the evolution of the complex human eye as an example writing, “To suppose that the eye ... could have been formed by natural selection, seems, I freely confess, absurd in the highest possible degree (Darwin, 1859).” Despite a lack of molecular details available at the time, he offered a rational hypothesis to explain this seemingly impossible puzzle, “If numerous gradations from a perfect and complex eye to one very imperfect and simple, each grade being useful to its possessor, can be shown to exist, ... then the difficulty of believing that a perfect and complex eye could be formed by natural selection, though insuperable by our imagination, can hardly be considered real (Darwin, 1859).” One and a half centuries later, as biologists have now elucidated many molecular underpinnings of eyes as well as those related but less complex light sensitive structures in multiple animal lineages, Darwin’s hypothesis concerning the evolution of the complex eye through stepwise selection is being corroborated in principle, and further enriched with

more detailed mechanisms (Lamb, 2011). Indeed, since Darwin, elucidation of plausible evolutionary paths towards complexity is growing in appreciation in modern biology, which is actively pursued in a wide selection of biological systems (Lenski *et al.*, 2003; Gompel *et al.*, 2005; Capra *et al.*, 2012; Finnigan *et al.*, 2012).

Metabolism is a defining property of cellular life, often depicted as a complex network of chemical transformations mediated by a multitude of enzymes (Weng & Noel, 2012b). Living organisms harness energy and chemical substances from their exterior environments, while synthesizing and degrading a plethora of metabolites to fulfill discrete physiological needs for the host survival and fitness in ever-changing environments. Metabolic systems can be viewed as complex traits, since individual metabolic enzymes in isolation are rarely pertinent to *in vivo* physiology until part of an organized metabolic pathway. Extant metabolic pathways in general contain a set of enzymes catalyzing sequential reactions in a highly concerted manner (Weng & Noel, 2012b). Given that the end metabolites of metabolic pathways ultimately confer selective advantages to the host, it is indeed difficult to understand how individual incipient enzymes arrive at specific activities over defined periods of time through step-wise evolution - Descent with Modification - prior to assembly of a more complex and integrated pathway. Without reference to the historical records encoded in organismal phylogeny and function, this quandary touches on the same question raised by Darwin concerning the evolution of the complex eye.

Metabolism provides an attractive platform to study evolutionary processes leading towards biological complexity. Metabolic systems exhibit tremendous depth and richness in their phylogenetic distribution throughout all domains of life. On the one hand, primary metabolism, encompassing pathways immediately required for the survival of the host, is conserved to a great extent in all living organisms, subject to modifications often in organisms living under extreme conditions (Mullins *et al.*, 2008); on the other hand, specialized metabolism, referring to pathways that yield chemicals dispensable for survival at normal conditions but contributing critically to the population fitness of the host in ecological niches, is often distributed in a taxonomically restricted manner. This feature is especially important for comparative studies of extant metabolic systems across lineages spanning a dynamic range of evolutionary time scales, through which useful information could be extracted to help reconstitute plausible trajectories underlying the occurrences of discrete metabolic traits.

Plants are renowned for their ability to produce an enormous array of chemicals as unique adaptive strategies well suited for their sessile life style in challenging terrestrial environments (Weng *et al.*, 2012b). The total number of these so called “specialized” metabolites present in the plant kingdom remains elusive as more and more species are examined molecularly. Nonetheless, through our limited knowledge of a small fraction of this remarkable chemodiversity, we continue to add to our knowledge base concerning the amazing ability of plants to select and exploit a diverse collection of peculiar physicochemical properties of natural chemicals, including but not limited to

color, flavor, fragrance, toxicity, stickiness, hydrophobicity, physical rigidity and redox potential, as means to overcome the multitude biotic and abiotic challenges facing plants throughout their life cycle. In the past few decades, the employment of molecular genetics, biochemistry and structural biology for studying plant metabolism is yielding a growing molecular appreciation for the mechanisms of how specialized metabolites are biosynthesized in plants. Recent advances in the development of genomic resources for a widening collection of taxonomically diverse plant species across the green plant lineage has further facilitated genome mining for identifying new metabolic pathways, and affords growing opportunities for comparative studies of phylogenetically related metabolic enzymes and pathways.

To address the conundrum of the evolutionary origin of multi-step metabolic pathways, several hypotheses have been put forward historically, including the retrograde model (Horowitz, 1945), the patch hypothesis (Jensen, 1976), the screening hypothesis (Jones *et al.*, 1991), and more recent hypotheses deduced from studying adaptive microbial metabolic pathways for detoxifying xenobiotics (Copley, 2009). Here, in the context of metabolic enzyme and pathway evolution, I summarize a set of general observations regarding to the rapid expansion of specialized metabolic systems in plants, and provide a number of examples to illustrate probable evolutionary trajectories underlying certain specialized metabolic traits. I then discuss the mechanistic basis for enzyme catalytic promiscuity and its recently recognized role in metabolic evolvability. Finally, I propose a generalized model to explain how complex

metabolic traits could arise sometimes in a saltatory fashion through the assembly of promiscuous enzymes into new pathways, followed by functional refinement at the catalytic, spatial and temporal levels.

General evolutionary patterns in plant specialized metabolism

System-level phylogenetic examination of specialized metabolic enzymes and pathways across major phyla of the plant kingdom reveals several general patterns, all consistent with the stepwise evolutionary processes underlying the emergence and evolution of the metabolic traits observed in extant plants. First, diverse plant specialized metabolic pathways branch from core primary metabolism at different nodes (Fig. 1a). For example, enormously rich phenylpropanoid metabolism widely present in plants starts with deamination of the aromatic amino acid phenylalanine through phenylalanine ammonium lyase (PAL) (Vogt, 2010). The diverse family of secondary plant terpenes (isoprenoids) begin with the primary metabolites isopentenyl pyrophosphate and dimethylallyl pyrophosphate (Chen *et al.*, 2011). Caffeine and related purine alkaloids, sparsely found in 13 orders of flowering plants, begin with core purine nucleotides (Ashihara *et al.*, 2008). This observation suggests that the emergence of new specialized pathways probably involved emergent catalytic activities towards certain primary metabolites, which yield new compounds that enhance host fitness in particular environments.

Second, in general, the taxonomic distribution of plant specialized metabolic traits correlates with the gradual evolutionary development of

specialized tissue types, organs, and/or life styles observed in land plants as they underwent extensive divergence over the last 500 million years (Fig. 1b). Several pathways, such as the biosynthesis of core phenylpropanoids, cuticles, sporopollenins, abscisic acid and flavonoids, are absent in extant charophytic algae most closely related to land plants, but ubiquitously present in all extant land plants (Weng & Chapple, 2010). The primary functions of these compounds are for protection against UV radiation and desiccation, representing major abiotic stresses facing those early land plants when migrating from aquatic habitats to terrestrial environments. When vascular plants arose, the ancestral core phenylpropanoid pathway was further elaborated to produce lignin, a phenolic polymer that provides physical rigidity to water-conducting xylem cells in vasculature and enables vascular plants to stand upright (Weng & Chapple, 2010; Weng *et al.*, 2010b). The evolution of trichomes in euphyllophytes coincides with the occurrence of a diverse array of metabolites enriched in these specialized surface structures, wherein most of these compounds are involved in chemical defense against herbivores (Dai *et al.*, 2010). The emergence of seed plants approximately 300 million years ago also led to the precipitation of a number of metabolic features related to seed physiology, such as the accumulation of condensed tannins in the seed coat and the rapid breakdown of starch during seed germination (Bradford & Nonogaki, 2007). Moreover, the rise of flowering plants over the past 200 million years led to an explosion of chemodiversity in volatile compounds, emitted by floral tissues to attract co-evolving pollinating insects (Pichersky *et al.*, 2006). Similarly, many land plant

lineages are also known to form symbiotic relationships with root microbiomes, wherein certain metabolites are secreted from root into the rhizosphere as chemical signals in to mediate specific root-microbial interactions (e.g. the trihydroxychalcone-derived isoflavonoids produced in legumes induces nodulation by rhizobia) (Walker *et al.*, 2003; Bulgarelli *et al.*, 2013). In plant specialized metabolism, new pathways continuously build on existing pathways, resulting in a relatively conserved set of earlier-evolved pathways extended by variable lineage-specific peripheral pathway branches in the extant land plants.

Third, the expansion of the specialized metabolism in plants did not involve the emergence of new protein folds, but rather the extensive exploitation of the sequence space in the pre-existing protein folds by natural selection. Many of these fold families are rooted in more ancient primary metabolic systems (Weng *et al.*, 2012b) (Fig. 2, Table 1). Particular catalytic machineries inherited along with the ancestral folds are often conserved during enzyme family expansion, although it is also common that new catalytic chemistry could arise within the fold family by reassembly of new catalytic residues in the active site (Weng & Noel, 2012b). For example, chalcone synthase (CHS), the first committed enzyme in flavonoid biosynthesis in plants, shares the same fold and catalytic machinery as the β -ketoacyl-ACP synthase III (KAS III), a key enzyme of fatty acid biosynthesis in plants and bacteria (Weng & Noel, 2012a). In another remarkable case, chalcone isomerase (CHI), a stereo-specific and catalytically perfected isomerase downstream of CHS in the flavonoid biosynthesis, evolved from a clade of non-catalytic CHI-fold proteins. In the green plant lineage

including simple chlorophyte algae, these non-catalytic CHI-fold proteins play a role in lipid biosynthesis and homeostasis in plants, and are structurally conserved in several other eukaryotic lineages as well as in some bacteria (Ngaki *et al.*, 2012). Different enzyme families, typified by their structural folds, underwent tremendous radiation during land plant evolution due to gene duplication events followed by selective refinement (Xue *et al.*, 2012). New specialized metabolic pathways continuously emerged in a lineage-specific manner by independently recruiting and refining descendants of these radiating families to catalyze sequential chemical reactions (Weng *et al.*, 2012b) (Fig. 2).

Repeated emergence of identical metabolic traits in disparate lineages

As divergent evolution predominantly drives the continuous expansion of chemical complexity in plants (Fig. 3a), identical metabolic traits often arise independently in disparate lineages through parallel or convergent evolution (Barton *et al.*, 2007; Pichersky & Lewinsohn, 2011; Weng & Noel, 2013) (Fig. 3b, c). Since a number of plant species derived from divergent lineages often co-occupy the same ecological niches, repeated evolution of common metabolic traits likely resulted from natural selection driven by similar selective pressures associated with particular environments. Alternatively, as plant specialized pathways are typically less constrained than their cousins in primary metabolism, early neutral drift followed later by selection before gene loss to some extent may also have contributed to the contingent occurrences of identical specialized metabolites in separate lineages.

Here, in the context of enzyme evolution, I use the term parallel evolution to describe independent acquisitions of identical catalytic properties in homologous enzymes belonging to the same fold family, whereas convergent evolution is reserved here to refer to the evolution of identical catalytic functions in nonhomologous enzymes possessing distinct protein folds (Fig. 3). It should be noted that these definitions, using protein fold structure as a major evaluation criterion, are more stringent than the ones previously used in describing repeated evolution in plant metabolic systems (Weng *et al.*, 2010a; Pichersky & Lewinsohn, 2011). These updated definitions of parallel and convergent evolution as used in describing enzyme evolution are in compliance with other sub-disciplines of evolution biology (Zuckerlandl *et al.*, 1965; Barton *et al.*, 2007). Moreover, the separation of convergent and parallel evolution, where previously parallel was considered subordinate to convergent, clarify existing, blurry and imperfect boundaries between the two terms currently in use.

In an apparent example of parallel evolution, syringyl (S) lignin, a fundamental building block of plant cell walls, occurs in two major plant lineages, lycophytes and angiosperms, which diverged from each other >400 Mya (Towers & Gibbs, 1953; Weng *et al.*, 2008). In angiosperms, S lignin biosynthesis requires two enzymes, ferulate 5-hydroxylase (F5H) and caffeic acid O-methyltransferase (COMT), forming a metabolic branch diverting flux from the biosynthesis of guaiacyl (G) lignin, a lignin type common to all the vascular plants (Weng & Chapple, 2010). It was later discovered that the lycophyte *Selaginella moellendorffii* independently evolved a bifunctional phenylpropanoid 3,5-

hydroxylase, *SmF5H*, and a companion bifunctional phenylpropanoid OMT, *SmCOMT*. Together, *SmF5H* and *SmCOMT* mediate a new S lignin pathway directly derived from *p*-hydroxyphenyl (H) monolignols, bypassing four steps of the canonical lignin biosynthesis defined in angiosperms (Weng *et al.*, 2008; Weng *et al.*, 2010a; Weng *et al.*, 2011). At the enzyme level, F5Hs and COMTs of angiosperms and *Selaginella* belong to the cytochrome P450 family and the S-adenosyl-L-Met (SAM)-dependent OMT family respectively, and have apparently arrived at their homologous activities in the two lineages via parallel evolution (Fig. 3b). Nonetheless, due to major differences in the substrate specificities between the homologous enzymes of angiosperm and *Selaginella* origins, the exact metabolic routes of S lignin biosynthesis are distinct at the pathway level in the two lineages (Weng *et al.*, 2010a).

Clear cases of convergent evolution within the boundaries of the above separation of convergent and parallel evolution at the protein structure level are relatively rare compared to clear cases of parallel evolution, but have been documented in various biological systems (Zuckermandl *et al.*, 1965). A clear example of such a case lies in the plant flavonoid biosynthetic pathway. Whereas most of the flowering plants examined to date employ type II flavone synthases (FNS II), belonging to the cytochrome P450 family, to catalyze the oxidation of flavanones to the corresponding flavones, members of the Apiceae family evolved a distinct type I flavone synthases (FNS I), belonging to the 2-oxoglutarate-dependent dioxygenase (ODD) family, to mediate the same overall

oxidation reaction but through a distinct chemical mechanism as employed by FNS II (Leonard *et al.*, 2005) (Fig. 3c).

Repeated emergence of similar or identical metabolic traits in distantly related taxa is indeed a common theme during land plant evolution, indicative of the considerable pliability of plant specialized metabolic systems (Pichersky & Lewinsohn, 2011; Weng & Noel, 2013). Nevertheless, our current knowledge of this remarkable phenomenon still remains mostly at the level of comparative phytochemistry focused on end product identity but rarely clear end product function. The ultimate elucidation of the molecular underpinnings of these cases will greatly illuminate how specific enzyme mechanisms, structures and metabolic networks are accessed through alternative evolutionary trajectories in nature.

Catalytic promiscuity and its role in metabolic evolvability

Enzymes are erroneously considered exquisitely precise and efficient molecular catalysts. Natural enzymes are also evolvable, in other words, can migrate through mutational trajectories to arrive at alternative catalytic activities sometimes in unexpected saltatory fashion (Zuckerlandl *et al.*, 1965; O'Maille *et al.*, 2008). How does an enzyme evolve from a given ancestral function to novel functions through descent with modification? Under scrutiny of multitude enzyme behaviors, it has been increasingly recognized that enzymes are not as “perfected” as we assumed, but rather contain varying levels of catalytic promiscuity (Tokuriki & Tawfik, 2009) (Fig. 4a). Enzyme catalytic promiscuity is

intrinsically associated with the dynamic nature of proteins, and indeed encompasses a range of mechanistic processes, namely substrate permissiveness, mechanistic elasticity, and concomitant product diversity (Weng & Noel, 2012b). Upon varying physicochemical conditions or mutations, the latent behaviors of enzymes may be amplified to yield alternative metabolites at sufficient levels, on which natural selection can operate (Aharoni *et al.*, 2005) (Fig. 4a). Gene duplication events, in many cases, yield multiple copies of isozymes, subject to subsequent differentiation in expression regulation or subcellular localization without changes in the catalytic function. In some other cases, gene duplication events can also disconnect the newly derived allele from its ancestral function that is often under evolutionary restraints, and allow divergence of catalytic function to promptly occur through additional mutations (Fig. 4a).

In primary metabolism, the essential functions of these core pathways impose constant selective pressures over long historical time, ultimately driving the participating enzymes towards catalytic efficiency, and as a result, much lower levels of catalytic promiscuity (Bar-Even *et al.*, 2011). However, in specialized metabolism, as enzyme functions tend to meander more frequently in response to fluctuating environments and/or have not been under sufficiently strong selection to reach optimum efficiency, catalytic promiscuity is more easily observed in these secondary enzymes (Tokuriki *et al.*, 2012; Weng *et al.*, 2012b). For example, terpene synthases (TPSs) are a class of plant specialized enzymes highly recognized for their catalytic promiscuity. One TPS can be responsible for

the formation of a bouquet of hydrocarbon compounds *in vivo* (Tholl *et al.*, 2005), wherein the product distribution fingerprints of a given TPS are acutely sensitive to environmental conditions and mutations (O'Maille *et al.*, 2008). The elevated level of catalytic promiscuity in specialized metabolic enzymes is probably one of the fundamental factors that contribute to the remarkable evolvability seen in these metabolic systems found throughout nature.

A clear example of metabolic evolution through exploitation of ancestral enzyme promiscuity was recently demonstrated in *Arabidopsis thaliana*, highlighting the emergence of a new set of α -pyrone-bearing metabolites, arabidopyrones (Weng *et al.*, 2012a). The first step of arabidopyrone biosynthesis is catalyzed by a P450 enzyme, CYP84A4, which evolved from a gene duplicate of the ancestral CYP84A1, a phenylpropanoid 5-hydroxylase involved in S lignin biosynthesis. CYP84A4 has apparently adopted one of the ancestral latent activities of its progenitor, and neofunctionalized into a specific phenylpropanoid 3-hydroxylase. This evolutionary development led to the synthesis of a catechol-substituted substrate, which can be further metabolized by a conserved extradiol ring-cleavage dioxygenase together with additional downstream redox enzymes to produce arabidopyrones. This study indicates that the system-level plasticity of the ancestral metabolic network allows the impact of a newly evolved catalytic activity in one enzyme to project through multiple pre-existing promiscuous catalytic steps to give rise to more profound metabolic outcomes.

Enzyme catalytic promiscuity seems to be a key component in how complex specialized metabolic pathways can rapidly evolve through Darwinian evolution (Fig. 4b). In the crowded cellular milieu, thousands of different types of enzymes are working simultaneously transforming a multitude of metabolites at a given time point. Due to the intrinsic catalytic promiscuity of enzymes, especially those involved in specialized metabolism, there is always a level of “metabolic noise” underlying so-called normal metabolic behavior, yielding low-level metabolites that are selectively neutral and may also vary greatly in their identities and concentrations throughout cellular metabolism. Contingency in evolution, such as genetic variations, oscillating ecological conditions or sudden changes in environments, can elicit rapid selection of promiscuous activities and products, resulting in the recruitment of ancillary enzymes to shape an emergent metabolic pathway for enhanced host and population fitness. Natural selection continually samples low-level promiscuously-derived metabolites, and when conditions are ripe, captures previously neutral metabolites that now confer selective advantages. When these selective alleles persist, they rapidly spread throughout a population, progressing from pathways of stochastic origins to fixed metabolic traits subject at times to further improvement through additional rounds of positive selection.

Perspectives

According to a recent estimation, out of the four billion species that ever lived on the Earth over the past 3.5 billion years, only 1% of them have survived today

(Barnosky *et al.*, 2011). As biologists trying to understand the evolutionary trajectories that connect related but discrete biological features, we are indeed left with only a very small fraction of the actual historical phylogenetic sampling explored throughout global life. Many highly complex traits, e.g. the human eye, are widely present in extant organisms. However, it is very common to find huge phylogenetic gaps that will never be known with certainty even with the most sophisticated means of "resurrecting" so-called ancient proteins, particularly as the phylogenies used for such calculations represent a very sparse sampling of extant organisms (Thornton *et al.*, 2003). In short, most evolutionarily intermediates - nodes on a phylogenetic tree - even those providing historical fitness advantages in organismal populations, were swept away in parallel by distinct lineages. This key aspect of descent with modification hinders our understanding of the detailed processes underlying molecular evolution, a problem also impinging on the emerging field of synthetic biology (Smock & Gierasch, 2005). Examination of metabolic systems using a molecular evolutionary perspective restrained by our understanding of catalytic mechanisms also reveals a set of optimized complex pathways, e.g. the TCA cycle, with little clue as to how these pathways initially emerged. However, plant specialized metabolism, which has expanded exponentially over the past 500 million years, provides a phylogenetically rich system for studying evolutionary processes leading to metabolic complexity on varying biological, chemical, architectural and time scales.

At the enzyme level, one can explore homologous enzymes isolated from closely related species to identify recently acquired mutations under positive selection in certain lineages that alter enzymatic activities, representing an early record of divergent evolution. By harnessing the great sequence diversity of a number of enzyme families using bioinformatics methods, e.g. statistical coupling analysis (SCA) (Halabi *et al.*, 2009), one can infer and further test the biophysical restraints imposed on a given protein fold family that shape evolvability. At the pathway level, the progenitor enzymes of a clearly characterized evolutionarily new pathway (Matsuno *et al.*, 2009; Weng *et al.*, 2012a) can be reassembled *in vitro* or introduced into a suitable set of hosts to monitor system-level metabolic promiscuity under differing extant conditions. Directed evolution can then follow to reveal mutational effects and their contribution to catalytic promiscuity or specificity, flux, etc., as well as to partially recapitulate mutational trajectories that mimic the natural evolutionary processes involving modification of multiple pathway components simultaneously. At the organismal level, it will be of great interest to investigate whether the rapid enzyme evolution in plants is facilitated by additional molecular machineries encoded by the plant genomes. For example, heat shock proteins may assist in stabilizing inherently unstable proteins folds carrying destabilizing mutations long enough for favorable activities to be selected for (Queitsch *et al.*, 2002; O'Maille *et al.*, 2008).

Acknowledgements

I thank my postdoctoral advisor Joseph P. Noel for his support and helpful comments in preparation of this manuscript. I also acknowledge the constructive comments from the reviewers. J.K.W. was supported by the Howard Hughes Medical Institute and a postdoctoral fellowship from the Pioneer Foundation.

Table 1 Major enzymes families known to be involved in specialized metabolism in extant land plants and their presumed cousins in primary metabolism. The number of genes in the genome of the green algae *Chlamydomonas reinhardtii* and the flowering plant *Arabidopsis thaliana* encoding enzymes belonging to each enzyme family are listed for comparison.

Major enzyme families involved in plant specialized metabolism	Gene number		Cousins in primary metabolism
	<i>C. reinhardtii</i>	<i>A. thaliana</i>	
Cytochromes P450	39	245	Sterol 14-demethylase
Glycosyltransferase family 1	1	115	Lipopolysaccharide glycosyltransferase
Class III peroxidase	0	75	Cytochrome c peroxidase
BAHD acyltransferase	0	61	Carnitine acetyltransferase
SCPL acyltransferase	7	51	Serine carboxypeptidase
Terpene synthase	0	32	Farnesyl pyrophosphate synthase
CCR-like NAD(P)H-dependent reductase	5	31	3- β -Hydroxysteroid dehydrogenase
Flavin-dependent monooxygenase	10	29	trimethylamine oxidase
SABBATH methyltransferase	0	24	Catechol- <i>O</i> -methyltransferase
CAD-like alcohol dehydrogenase	5	19	Alcohol dehydrogenase
GH3 acyl-adenylase	0	19	Long-chain fatty acyl-CoA synthetase
COMT-like methyltransferase	1	17	Catechol- <i>O</i> -methyltransferase
Strictosidine synthase-like	1	15	Gluconolactonase
Acyl-CoA ligase	4	13	Long-chain fatty acyl-CoA synthetase
CCoAOMT-like methyltransferase	1	8	Catechol- <i>O</i> -methyltransferase
Type III polyketide synthase	0	4	β -ketoacyl-ACP synthase III

Figure legends

Figure 1. Evolutionary progression of plant specialized metabolism. **(a)** Most of the plant specialized metabolic pathways are extended from the core primary metabolic network at various nodes. The colored networks represent the major primary metabolic pathways, modified from an image by the Kyoto Encyclopedia of Genes and Genomes (KEGG, <http://www.genome.jp/kegg/>). As examples, three plant specialized pathways leading towards the biosynthesis of phenylpropanoids, terpenoids, and caffeine are derived from their corresponding primary metabolite precursors. PHE, phenylalanine; IPP, isopentenyl pyrophosphate; DMIPP, dimethylallyl pyrophosphate; X5P, xanthosine 5'-phosphate. **(b)** The taxonomical distribution of plant specialized metabolic pathways echoes the gradual diversification of developmental programs and life styles during land plant evolution. A number of such examples are illustrated at the major branching points of the green plant phylogeny. L, land plants; V, vascular plants; E, euphyllophytes; S, seed plants, and F, flowering plants. Images of representative species belonging to each major lineage are placed on top of the tree branches.

Figure 2. Disparate plant specialized metabolic pathways assembled from descendents of radiating enzyme families with their ancestry deeply rooted in primary metabolism. For example, acyltransferase, P450, and OMT-fold enzyme families are typical plant specialized metabolic enzyme families that share the similar protein folds with their cousins in primary metabolism yet have undergone

extensive expansion during land plants evolution. CCoAOMT, caffeoyl CoA 3-O-methyltransferase, HCT, hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyl transferase; PHE, phenylalanine; SHT, spermidine hydroxycinnamoyl transferase; RAS, rosmarinic acid synthase; TSM, Tapetum-specific methyltransferase.

Figure 3. Divergent, parallel and convergent evolution of enzyme functions in plant specialized metabolism. **(a)** ChOMT and IOMT are two O-methyltransferases involved in different stages of flavonoid biosynthetic pathway in *Medicago sativa*. As an example for divergent evolution, ChOMT (PDB: 1FP1) and IOMT (PDB: 1FP2) share the same overall fold, but differ significantly in their active-site residues, resulting in distinct substrate specificities (Zubieta *et al.*, 2001). ChOMT, Chalcone O-methyltransferase; IOMT, isoflavone O-methyltransferase. **(b)** The lycophyte *Selaginella* and angiosperms have independently recruited paralogous O-methyltransferases for syringyl lignin biosynthesis through parallel evolution. Structural comparison between LpCOMT (PDB: 3P9I) and SmCOMT (modeled over LpCOMT) suggests that the similar active-site steric features for accommodating the same substrate in the two enzymes are contributed from distinct amino acids at corresponding positions. COMT, caffeic acid O-methyltransferase; Lp, *Lolium perenne*; Sm, *Selaginella moellendorffii*. **(c)** As an apparent case of convergent evolution, members of the Apiceae family have recruited a unique FNS I, belonging to the ODD family, to convert flavanones to flavones, in addition to FNS II, an enzyme more widely

spread in flowering plants belonging to the P450 family. FNS I, type I flavone synthase; FNS II, type II flavone synthase; ODD, 2-oxoglutarate-dependent dioxygenase.

Figure 4. Enzyme catalytic promiscuity and its role in metabolic evolution. **(a)** Catalytic promiscuity is deeply rooted in the dynamic nature of proteins. The level of latent activities of an enzyme may vary drastically upon environmental conditions or mutations. Gene duplications followed by neo- or sub-functionalization are common routes towards metabolic diversity in plant metabolic evolution. The colored patches illustrated in the middle of the enzyme structures represent different steric/electrostatic/ hydrophobic features of the enzyme active site subject to changes upon protein dynamics or mutations. Hypothetical percentage values denote the proportion of given conformations of an enzyme associated with disparate catalytic activities in the entire enzyme population in solution. **(b)** A mechanistic model explaining the rapid emergence of complex metabolic pathways in plant specialized metabolism. The combinatorial effect of catalytic promiscuity of individual enzymes in the cellular milieu elevates to system-level metabolic promiscuity, yielding incipient pathways through stochastic assembly of components from the preexisting pathways (highlighted in the dotted circle). If such pathways provided selective advantages to the host organism, they would become fixed and further improved through positive selection.

References

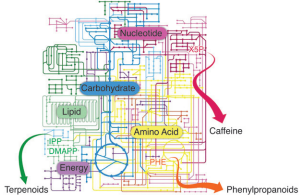
- Aharoni A, Gaidukov L, Khersonsky O, Mc QGS, Roodveldt C, Tawfik DS. 2005.** The 'evolvability' of promiscuous protein functions. *Nature Genetics* **37**: 73-76.
- Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, Devos D, Suprpto A, Karni-Schmidt O, Williams R, Chait BT, Rout MP, Sali A. 2007.** Determining the architectures of macromolecular assemblies. *Nature* **450**: 683-694.
- Ashihara H, Sano H, Crozier A. 2008.** Caffeine and related purine alkaloids: Biosynthesis, catabolism, function and genetic engineering. *Phytochemistry* **69**: 841-856.
- Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, Milo R. 2011.** The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **50**: 4402-4410.
- Barnosky AD, Matzke N, Tomiya S, Wogan GO, Swartz B, Quental TB, Marshall C, McGuire JL, Lindsey EL, Maguire KC, Mersey B, Ferrer EA. 2011.** Has the earth's sixth mass extinction already arrived? *Nature* **471**: 51-57.
- Barton NH, Briggs DEG, Eisen JA, Goldstein DB, Patel NH. 2007.** *Evolution*. Cold Spring Harbor, New York, U. S. A.: Cold Spring Harbor Laboratory Press.
- Beldade P, Brakefield PM. 2002.** The genetics and evo-devo of butterfly wing patterns. *Nat Rev Genet* **3**: 442-452.
- Bradford K, Nonogaki H. 2007.** *Annual plant reviews, seed development, dormancy and germination*. Oxford, Oxfordshire, U. K.: John Wiley & Sons.
- Bulgarelli D, Schlaeppi K, Spaepen S, Ver Loren van Themaat E, Schulze-Lefert P. 2013.** Structure and functions of the bacterial microbiota of plants. *Annu Rev Plant Biol* **64**: 807-838.
- Capra EJ, Perchuk BS, Skerker JM, Laub MT. 2012.** Adaptive mutations that prevent crosstalk enable the expansion of paralogous signaling protein families. *Cell* **150**: 222-232.
- Chen F, Tholl D, Bohlmann J, Pichersky E. 2011.** The family of terpene synthases in plants: A mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant Journal* **66**: 212-229.
- Copley SD. 2009.** Evolution of efficient pathways for degradation of anthropogenic chemicals. *Nat Chem Biol* **5**: 559-566.
- Dai X, Wang G, Yang DS, Tang Y, Broun P, Marks MD, Sumner LW, Dixon RA, Zhao PX. 2010.** Trichome: A comparative omics database for plant trichomes. *Plant Physiology* **152**: 44-54.
- Darwin C. 1859.** *On the origin of species by means of natural selection, or, the preservation of favoured races in the struggle for life*. London: J. Murray.
- Finnigan GC, Hanson-Smith V, Stevens TH, Thornton JW. 2012.** Evolution of increased complexity in a molecular machine. *Nature* **481**: 360-364.
- Gompel N, Prud'homme B, Wittkopp PJ, Kassner VA, Carroll SB. 2005.** Chance caught on the wing: Cis-regulatory evolution and the origin of pigment patterns in drosophila. *Nature* **433**: 481-487.

- Halabi N, Rivoire O, Leibler S, Ranganathan R. 2009.** Protein sectors: Evolutionary units of three-dimensional structure. *Cell* **138**: 774-786.
- Horowitz NH. 1945.** On the evolution of biochemical syntheses. *Proceedings of the National Academy of Sciences of the United States of America* **31**: 153-157.
- Jensen RA. 1976.** Enzyme recruitment in evolution of new function. *Annual Review of Microbiology* **30**: 409-425.
- Jones CG, Firn RD, Malcolm SB. 1991.** On the evolution of plant secondary chemical diversity. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **333**: 273-280.
- Lamb TD. 2011.** Evolution of the eye. Scientists now have a clear vision of how our notoriously complex eye came to be. *Scientific American* **305**: 64-69.
- Lander ES, Schork NJ. 1994.** Genetic dissection of complex traits. *Science* **265**: 2037-2048.
- Lenski RE, Ofria C, Pennock RT, Adami C. 2003.** The evolutionary origin of complex features. *Nature* **423**: 139-144.
- Leonard E, Yan Y, Lim KH, Koffas MA. 2005.** Investigation of two distinct flavone synthases for plant-specific flavone biosynthesis in *Saccharomyces cerevisiae*. *Applied and Environmental Microbiology* **71**: 8241-8248.
- Matsuno M, Compagnon V, Schoch GA, Schmitt M, Debayle D, Bassard JE, Pollet B, Hehn A, Heintz D, Ullmann P, Lapierre C, Bernier F, Ehlting J, Werck-Reichhart D. 2009.** Evolution of a novel phenolic pathway for pollen development. *Science* **325**: 1688-1692.
- Mullins EA, Francois JA, Kappock TJ. 2008.** A specialized citric acid cycle requiring succinyl-coenzyme A (CoA):Acetate CoA-transferase (Aarc) confers acetic acid resistance on the acidophile *Acetobacter acetii*. *Journal of Bacteriology* **190**: 4933-4940.
- Ngaki MN, Louie GV, Philippe RN, Manning G, Pojer F, Bowman ME, Li L, Larsen E, Wurtele ES, Noel JP. 2012.** Evolution of the chalcone-isomerase fold from fatty-acid binding to stereospecific catalysis. *Nature* **485**: 530-533.
- O'Maille PE, Malone A, Dellas N, Andes Hess B, Jr., Smentek L, Sheehan I, Greenhagen BT, Chappell J, Manning G, Noel JP. 2008.** Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases. *Nat Chem Biol* **4**: 617-623.
- Pichersky E, Lewinsohn E. 2011.** Convergent evolution in plant specialized metabolism. *Annu Rev Plant Biol* **62**: 549-566.
- Pichersky E, Noel JP, Dudareva N. 2006.** Biosynthesis of plant volatiles: Nature's diversity and ingenuity. *Science* **311**: 808-811.
- Queitsch C, Sangster TA, Lindquist S. 2002.** Hsp90 as a capacitor of phenotypic variation. *Nature* **417**: 618-624.
- Smock RG, Gierasch LM. 2005.** Finding the fittest fold: Using the evolutionary record to design new proteins. *Cell* **122**: 832-834.
- Tholl D, Chen F, Petri J, Gershenzon J, Pichersky E. 2005.** Two sesquiterpene synthases are responsible for the complex mixture of sesquiterpenes emitted from *Arabidopsis* flowers. *Plant Journal* **42**: 757-771.
- Thornton JW, Need E, Crews D. 2003.** Resurrecting the ancestral steroid receptor: Ancient origin of estrogen signaling. *Science* **301**: 1714-1717.

- Tokuriki N, Jackson CJ, Afriat-Jurnou L, Wyganowski KT, Tang R, Tawfik DS. 2012.** Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nat Commun* **3**: 1257.
- Tokuriki N, Tawfik DS. 2009.** Protein dynamism and evolvability. *Science* **324**: 203-207.
- Towers GHN, Gibbs RD. 1953.** Lignin chemistry and the taxonomy of higher plants. *Nature* **172**: 25-26.
- Vogt T. 2010.** Phenylpropanoid biosynthesis. *Mol Plant* **3**: 2-20.
- Walker TS, Bais HP, Grotewold E, Vivanco JM. 2003.** Root exudation and rhizosphere biology. *Plant Physiology* **132**: 44-51.
- Weber BH. 1999.** Irreducible complexity and the problem of biochemical emergence. *Biology & Philosophy* **14**: 593-605.
- Weng JK, Akiyama T, Bonawitz ND, Li X, Ralph J, Chapple C. 2010a.** Convergent evolution of syringyl lignin biosynthesis via distinct pathways in the lycophyte selaginella and flowering plants. *Plant Cell* **22**: 1033-1045.
- Weng JK, Akiyama T, Ralph J, Chapple C. 2011.** Independent recruitment of an o-methyltransferase for syringyl lignin biosynthesis in selaginella moellendorffii. *Plant Cell* **23**: 2708-2724.
- Weng JK, Chapple C. 2010.** The origin and evolution of lignin biosynthesis. *New Phytologist* **187**: 273-285.
- Weng JK, Li X, Stout J, Chapple C. 2008.** Independent origins of syringyl lignin in vascular plants. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 7887-7892.
- Weng JK, Li Y, Mo H, Chapple C. 2012a.** Assembly of an evolutionarily new pathway for α -pyrone biosynthesis in arabidopsis. *Science* **337**: 960-964.
- Weng JK, Mo H, Chapple C. 2010b.** Over-expression of f5h in comt-deficient arabidopsis leads to enrichment of an unusual lignin and disruption of pollen wall formation. *Plant Journal* **64**: 898-911.
- Weng JK, Noel JP. 2012a.** Structure-function analyses of plant type iii polyketide synthases. *Methods in Enzymology* **515**: 317-335.
- Weng JK, Noel JP. 2012b.** The remarkable pliability and promiscuity of specialized metabolism. *Cold Spring Harbor Symposia on Quantitative Biology*: doi:10.1101/sqb.2012.1177.014787.
- Weng JK, Noel JP. 2013.** Chemodiversity in selaginella: A reference system for parallel and convergent metabolic evolution in terrestrial plants. *Front Plant Sci* **4**: 119.
- Weng JK, Philippe RN, Noel JP. 2012b.** The rise of chemodiversity in plants. *Science* **336**: 1667-1670.
- Xue Z, Duan L, Liu D, Guo J, Ge S, Dicks J, P OM, Osbourn A, Qi X. 2012.** Divergent evolution of oxidosqualene cyclases in plants. *New Phytologist* **193**: 1022-1038.
- Zubieta C, He XZ, Dixon RA, Noel JP. 2001.** Structures of two natural product methyltransferases reveal the basis for substrate specificity in plant o-methyltransferases. *Nature Structural Biology* **8**: 271-279.

Zuckerkindl E, Pauling L, Bryson V, Vogel HJ 1965. Evolutionary divergence and convergence in proteins. *Evolving genes and proteins*. New York: Academic Press, 97-166.

(a)



(b)

