



Review

Cite this article: Barkman TJ. 2024 Applications of ancestral sequence reconstruction for understanding the evolution of plant specialized metabolism. *Phil. Trans. R. Soc. B* **379**: 20230348. <https://doi.org/10.1098/rstb.2023.0348>

Received: 6 March 2024

Accepted: 15 April 2024

One contribution of 23 to a theme issue 'The evolution of plant metabolism'.

Subject Areas:

evolution, plant science, molecular biology

Keywords:

enzyme evolution, specialized metabolism, ancestral sequence reconstruction

Author for correspondence:

Todd J. Barkman

e-mail: todd.barkman@wmich.edu

Applications of ancestral sequence reconstruction for understanding the evolution of plant specialized metabolism

Todd J. Barkman

Department of Biological Sciences, Western Michigan University, Kalamazoo, MI 49008, USA

id TJB, 0000-0003-2259-2345

Studies of enzymes in modern-day plants have documented the diversity of metabolic activities retained by species today but only provide limited insight into how those properties evolved. Ancestral sequence reconstruction (ASR) is an approach that provides statistical estimates of ancient plant enzyme sequences which can then be resurrected to test hypotheses about the evolution of catalytic activities and pathway assembly. Here, I review the insights that have been obtained using ASR to study plant metabolism and highlight important methodological aspects. Overall, studies of resurrected plant enzymes show that (i) exaptation is widespread such that even low or undetectable levels of ancestral activity with a substrate can later become the apparent primary activity of descendant enzymes, (ii) intramolecular epistasis may or may not limit evolutionary paths towards catalytic or substrate preference switches, and (iii) ancient pathway flux often differs from modern-day metabolic networks. These and other insights gained from ASR would not have been possible using only modern-day sequences. Future ASR studies characterizing entire ancestral metabolic networks as well as those that link ancient structures with enzymatic properties should continue to provide novel insights into how the chemical diversity of plants evolved.

This article is part of the theme issue 'The evolution of plant metabolism'.

1. Introduction

Specialized metabolites are a diverse set of molecules that serve a wide array of functions in plant biology [1]. Well documented roles include not only signalling for defence against pathogens and herbivores but also the attraction of beneficial mutualists such as pollinators [2]. Hormone modulation, UV protection and osmoregulation are also achieved by specialized metabolite formation [3]. Their biosynthesis is fascinating, with enzymes that can generate multiple products from a single substrate, such as terpene synthases [4], as well as those that can catalyse reactions with multiple substrates to produce an array of different products, such as acyltransferases [5]. There is also growing evidence that the same specialized chemical can be synthesized by unrelated enzymes using different chemistry [6,7] as well as a surprising number of reactions that appear to be non-enzymatic/spontaneous [8,9]. One hallmark of specialized metabolism is the sporadic distribution of many metabolites, which shows that entire pathways can evolve independently and repeatedly using homologous or unrelated sets of enzymes [10]. Some general principles for the evolution of specialized metabolism have emerged, including that many pathways appear to have origins from primary metabolism [3,11]. Recruitment of divergent enzymes into specialized metabolic pathways results from gene duplication of the primary metabolic enzymes, which frequently become organized with others into operon-like

clusters [12]. Various 'omics' techniques have accelerated great progress towards deciphering the modern-day genetic underpinnings of specialized metabolite biosynthesis and how the vast diversity of specialized metabolism has evolved [1,3]. Yet, further insight will likely require the use of more historically explicit approaches like ancestral sequence reconstruction (ASR) combined with experimental characterization.

ASR involves the estimation of extinct protein sequences followed by their resurrection using molecular biology techniques [13–16]. The activities of such resurrected molecules can then be experimentally determined and directly compared with what is known for their modern-day descendants to understand functional change over time. Because numerous reviews have been published in the last 10 years [14,17–20], only a cursory summary of the approach to ASR is provided here. ASR proceeds by the generation of an alignment of homologous sequences from which an optimal tree is estimated, assuming a model of nucleotide/amino acid substitution. The empirical Bayes approach is the most widely used method for ancestral sequence prediction and makes use of the branching pattern, branch lengths, alignment and substitution model to provide some measure of uncertainty associated with the estimated state for each aligned position at each node of the tree [21,22]. Once a sequence estimate is obtained, gene synthesis usually includes codon optimization for the host organism to be used for heterologous protein expression. Simulation studies show that all methods are reasonable but that maximum likelihood (ML) approaches provide the highest accuracy, especially compared with parsimony [23–25]. All methods assume the alignment and tree are known without error [26]. Nonetheless, simulation studies have suggested that errors in the branching structure and branch lengths may have a minimal impact on the estimates of ancestral sequences in many cases [25,27,28], but accuracy also depends on how much phylogenetic signal exists in the dataset [29] and rarely considered phenomena such as recombination [30]. Reconciling gene trees to use species tree-aware methods may also improve accuracy [31]. Estimating positions in an ancestral sequence is more complicated when there are alignment gaps [32–34]. Usually, parsimony is used to estimate whether an ancestral position was present or not, but ML methods can also be applied [22]. Alignment gaps are often particularly abundant at the N- and C-termini, where length variation due to indels is commonly observed. If the N-terminal portion of a protein encodes a signal peptide, which may be poorly conserved among divergent species, ASR may proceed by excluding those regions since they are not necessary for many heterologous protein expression systems. Whereas proteins from primary metabolism may be easily aligned and allow highly confident ASR, not all proteins from a highly diversified specialized metabolite enzyme family are equally amenable to ASR. Sadly, there is no metric that can indicate whether ASR will be reliable for any particular protein alignment, but it is generally true that higher posterior probabilities are associated with sequence estimates from more recent divergences.

2. Interrogation of the plant tree of life using ancestral sequence reconstruction is in its infancy

ASR has found a place in studies of plant proteins, with ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) being a prominent target given its importance for planetary carbon fixation and interesting kinetics with CO₂ and O₂ [35–37]. Additional ASR studies of non-enzymatic proteins include light sensing in ancient cyanobacteriochromes [38], ancestral interactions of MADS box [39] and cyanobacterial photoprotection proteins [40], self-incompatibility ligand–receptor evolution [41] and microRNA effects on plant development [42]. Because ASR is a relatively new technique, there have been only 11 applications to study aspects of ancient plant specialized metabolism (figure 1) [8,43–54]. Most studies have focused on the more recent origins of metabolic innovations that are often restricted to one or a few lineages and are likely less than 10 Ma old, such as terpene derivatives in Lamiales (figure 1a) and anthocyanins in Solanales (figure 1b). Others have targeted the ancient-most origins for metabolites such as methyl salicylate and flavonoids that are shared by most land plants and are more than 100 Ma old (figure 1j,k). Several classes of specialized metabolites have been studied, including terpenoids (figure 1a,c,g,h,i), alkaloids (figure 1c,e,f), flavonoids (figure 1b,k), benzenoids (figure 1j) and nitriles (figure 1d). The resurrected ancient enzymes perform various biochemical modifications of specialized metabolite substrates and include reductases (figure 1a,b), methyltransferases (figure 1e,f,j), cytochrome P450 oxidases (figure 1g,i), isomerases (figure 1k), lyases (figure 1d) and cyclases (figure 1c,h). ASR has found growing application as a tool for protein engineering, and plant metabolic enzymes are no exception. The study of an ancestral CYP725A4 taxadiene hydroxylase (figure 1i) was largely undertaken in order to improve enzymes for industrial purposes. Likewise, an ancestral photodecarboxylase from *Chlorella* and relatives exhibits properties that could be useful for industry [55]. In spite of the progress made, abundant lineages remain unstudied and many of these are well known for producing some of the most iconic specialized plant metabolites, such as glucosinolates, largely restricted to Brassicales, and betalains, found in Caryophyllales. Of course, since the biochemical basis for the production of many metabolites in modern-day species remains unknown, it is hardly a surprise that studies using ASR have yet to be conducted! Below we discuss these ASR studies in terms of the implications of their findings for understanding specialized metabolite evolution.

3. Catalytic proteins may emerge from inactive ancestors to produce specialized metabolites

One study on the origins of plant specialized metabolic enzymes is particularly notable because it showed how a catalytic enzyme arose from a non-catalytic ancestor [49]. Chalcone isomerase (CHI) is a crucial early-pathway enzyme providing flux to various pigments and other flavonoid metabolites in all land plants (figure 1k). CHI is homologous to fatty acid-binding proteins that are non-catalytic in a wide array of organisms. To understand how CHI activity evolved, Kaltenbach *et al.* [49], used ASR to assay the progenitor proteins that gave rise to the catalytic enzyme. They found that an ancient non-catalytic ancestor, presumably involved in fatty acid binding, was duplicated in early land plants, and one of its descendants evolved

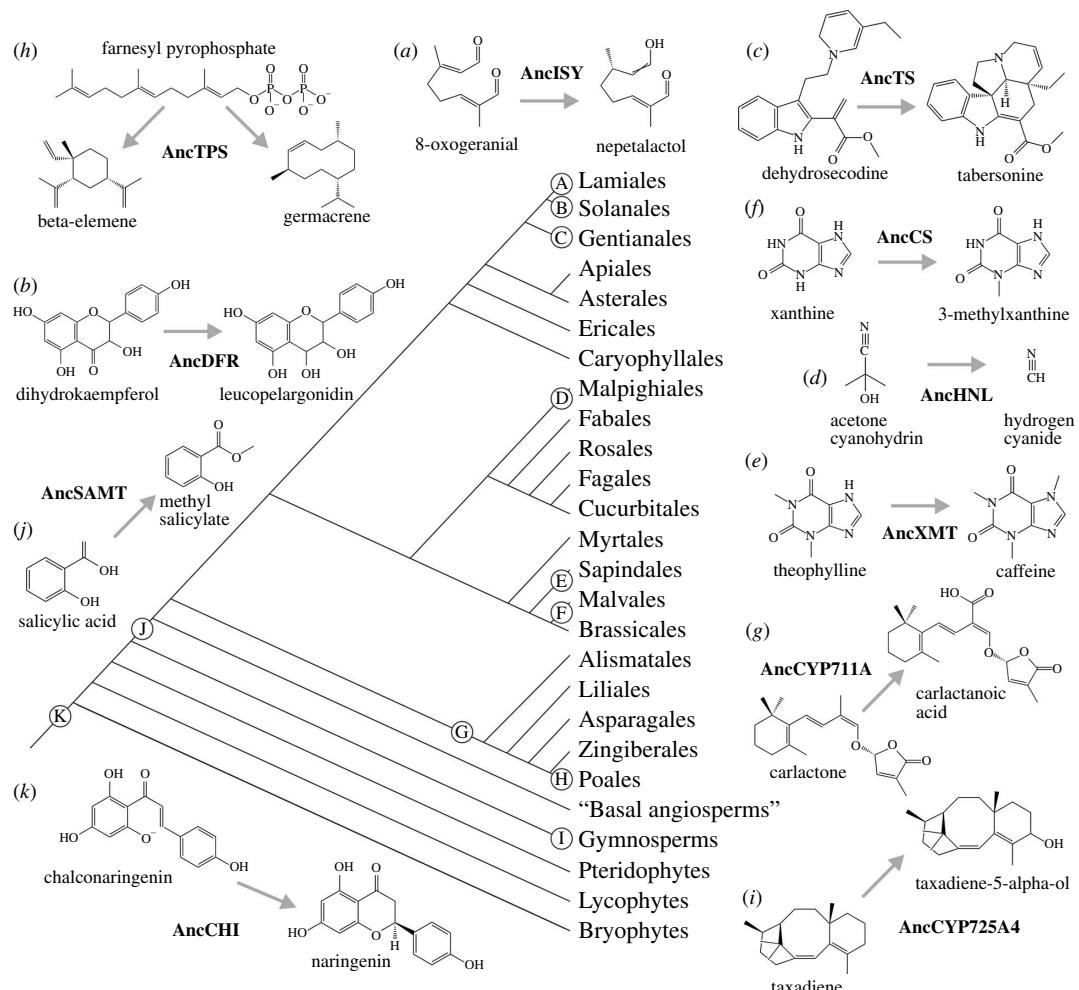


Figure 1. A land plant phylogeny shows the lineage-specific context of published studies using ancestral sequence reconstruction to understand the evolution of plant specialized metabolism. Labelled circles show the primary lineages investigated; however, while some studies investigated multiple nodes in plant phylogeny, those are not shown for the sake of clarity. Biosynthetic reactions are labelled to correspond to branches along which ancient proteins were resurrected. In many cases, additional reactions or substrate stereoisomers were studied but only exemplars are shown.

enzymatic activity to give rise to CHI in modern-day plants. The initial stages of enzymatic emergence appear to have required an increase in the size of the active site to allow for productive chalconaringenin substrate binding because the active site of the non-catalytic ancestor was smaller, such that only non-productive binding would likely occur [49]. Ultimately, it appears that one of the most important catalytic residues for CHI activity was actually found in the ancestral, non-enzymatic fatty acid-binding protein (FABP)-like protein but only later did it become repositioned by other mutations to allow the emergence of enzymatic activity. The evolutionary trajectory towards catalytic activity did not involve strong negative epistatic mutations and, surprisingly, could have evolved via multiple different mutational paths—even some involving mutations that did not occur historically [49]. This study stands as an interesting example of enzyme activity evolving without having a promiscuous ancestor, and this catalytic innovation appears to have been concomitant with early land plant adaptation. It is not clear how many specialized metabolic enzymes have arisen from non-catalytic ancestral proteins, but catalytic promiscuity/plasticity is known in many modern-day enzymes like fatty acid desaturases [56]; therefore, ASR can likely also improve our understanding of the origins and shifts of catalytic properties more broadly.

4. Apparent exaptations of promiscuous ancestral enzymes seem to promote the evolution of novel descendants

Many models of protein functional evolution require gene duplication and posit the existence of multifunctional/promiscuous ancestors [57–61]. The multiple different ancestral activities, which may be thought of as exaptations, can then be partitioned by descendants after gene duplication, thereby contributing to functional diversification [62–65]. The assumption of promiscuous ancestors can be explicitly tested using ASR [66], and several studies of plant specialized metabolism have done just that. One such study system involves the enzymatic generation of hydrogen cyanide (HCN), which has evolved in numerous independent angiosperm lineages and has a potent role in herbivore defence. In the well studied cyanogenic members of Euphorbiaceae (e.g. cassava), the hydroxynitrile lyase (HNL) enzymes that generate HCN (figure 1d) have been identified, and they are most closely related to esterases that work on a variety of substrates including methyl salicylate to form salicylic acid during SAR for pathogen defence. Phylogenetic analysis suggests that HNL evolved from ancestral esterases, which prompts

the question of how a different catalytic reaction could have arisen. To investigate this, Devamani *et al.* [50] and Rauerdink *et al.* [51] used ASR to reconstruct the ancient enzymes that would have given rise to the modern-day HNL [50,51]. In this case, it was shown that, although modern-day esterases are largely specific for ester cleavage, the ancestral enzymes could perform both ester hydrolysis and the elimination reaction to form hydroxynitriles. After gene duplication, one descendant lineage of proteins appears to have specialized as esterases and the other for hydroxynitrile lyase activity. Interestingly, there appears to be an evolutionary incompatibility between specialization for lyase or esterase activity because mutations to improve one activity came at the expense of the other. The structural basis for ancestral promiscuity of ancHNL1 was determined using X-ray crystallography and appears to be due to larger active site volume, which allows different binding orientations of each of the different classes of substrates [67].

In addition to catalytic promiscuity, substrate promiscuity has been reported for numerous ancestral specialized metabolic enzymes. One prominent example was documented from Lamiaceae, in which the origin of nepetalactone iridoids from catnip was studied (figure 1a) [48]. The origin of nepetalactones in catnip is particularly intriguing since iridoids appear to have been historically lost in much of the family, only to re-evolve within *Nepeta*. Iridoid synthase (ISY) is found to catalyse the formation of nepetalactol from 8-oxogeranial (8OH) in *Nepeta* to provide precursors for nepetalactone production. ISY is most closely related to progesterone reductase (PRISE), which contributes to the biosynthesis of cardenolides and other specialized metabolites in many plants. Lichman *et al.* [48] resurrected the ancestral PRISE that appears to have given rise to ISY in modern-day *Nepeta* to produce iridoid terpenoids. The progenitor was promiscuous and showed activity with both progesterone and 8OH. After gene duplication, one daughter enzyme appears to have specialized for progesterone reduction while the other evolved to specialize for activity with 8OH to lead to nepetalactol formation. Thus, ancestral substrate promiscuity promoted the re-evolution of iridoids in the catnip lineage. Substrate promiscuity is also known for several other lineages of enzymes, including ancestral methyltransferases producing volatile methyl esters [43]. In addition, cyclases involved in terpenoid production [8,53] demonstrate promiscuous activities to generate multiple products from a single substrate (figure 1c,h). Because so many modern-day specialized metabolic enzymes are either catalytically or substrate promiscuous [68], this finding for ancestral enzymes is hardly surprising. However, it is important to note that the promiscuous ancestral activities that we assume to be historically neutral exaptations may instead be biochemical adaptations with some selective value. To discern between the two possibilities is difficult because we cannot unambiguously determine whether the *in vitro* measured enzymatic activity actually had fitness value in the context of ancient plant physiology. Enzyme promiscuity has been leveraged in studies of protein engineering, thereby providing further evidence for the importance of this characteristic for generating novelty [69]. While the existence of ancestral promiscuity has been hypothesized to be involved in the generation of novel plant specialized metabolites, ASR has not only verified its role but allowed insights into the mechanisms by which it evolves.

5. Epistasis may or may not limit evolutionary transitions of plant specialized metabolic enzymes

Although epistasis is well known as a classical genetic phenomenon, studies of intramolecular epistasis are relatively new [70]. Positive or negative intramolecular epistatic interactions have been documented whereby one mutation may affect either the sign or magnitude of a phenotype caused by another mutation [71]. In the case of strong negative epistasis, evolutionary paths through an adaptive landscape may be constrained [72–76]. Smith *et al.* [47] investigated the impact of intramolecular epistasis on an anthocyanin biosynthetic colour transition in Solanaceae (figure 1b). Anthocyanins are a subnetwork of the flavonoid pathway and are notable for imparting red, purple and blue coloration to flowers, especially in the Iochrominae, which exhibit myriad colours with many inferred shifts between them. In a particularly integrative and sophisticated analysis, Smith & Rausher [77] mapped changes in floral pigmentation to three loci: flavonoid 3'-hydroxylase (F3'H), flavonoid 3',5'-hydroxylase (F3'5'H) and dihydroflavonol reductase (Dfr). In red-flowered species, both F3'H and F3'5'H were either downregulated or pseudogenized to reduce flux to blue pigment precursors and increase it towards the red-pigment branch of the metabolic network. To study the impact of changes in Dfr (figure 1b), Smith *et al.* [47] resurrected ancestral Dfr from the blue-flowered ancestor and that of the red-flowered ancestor that gave rise to the modern-day red-flowered species *Iochroma gesnerioides* and *Iochroma fuchsiioides*. The blue-flowered ancestral enzyme had a nearly equal preference for all three anthocyanin precursors, dihydrokaempferol (DHK), dihydroquercetin (DHQ) and dihydromyricetin (DHM), whereas the red-flowered ancestral enzyme had more than doubled its relative preference for the red-pigment precursor, DHK. To investigate the mutational basis for this shift in relative enzyme preference, Smith *et al.* [47] mutated three amino acid residues that differ between the two ancestral enzymes and made all possible combinations to evaluate potential epistasis. They found that there was statistically significant positive and negative epistasis among the three sites. The negative epistatic interactions are hypothesized to have restricted the evolutionary paths (order of substitutions) that could have occurred during the specialization for producing red anthocyanins. Whereas this and other studies have documented negative epistatic restrictions on evolutionary transitions [78,79], most studies of plant specialized metabolic enzymes show only minimal such effects.

For instance, negative epistasis was not detected for the origin of CHI activity (figure 1k) from a non-catalytic ancestral protein, thereby leading to the conclusion that evolution would not be limited [49]. More recently, it was shown that epistatic interactions also exist among mutations that underlie the origin of salicylic acid methylation preference by S-adenosyl-L-methionine : salicylic acid carboxyl methyltransferase (SAMT) enzymes [45] (figure 1j). In this case, the evolution of novel methylation characteristics was mostly due to a single mutation. Although statistically significant intramolecular epistasis was detected, the interactions among mutations were of small positive magnitude such that evolutionary paths would not be restricted. The findings of these studies may be generally applicable since they are congruent with studies of modern-day plant specialized metabolic enzymes that have shown substrate preference switches may occur via one or a few substitutions [80]. Of course,

if evolutionary transitions between any two proteins with different activities involve more than six or seven amino acid replacements, the number of mutants to generate and characterize is so vast as to preclude comprehensive analysis of epistasis owing to a lack of high-throughput techniques. However, a fruitful *tour de force* approach for the analysis of hundreds of mutated sesquiterpene synthases has been shown [81]. Future advancements in the study of intramolecular epistasis and its impact on the evolution of plant specialized metabolism will require robust statistical analysis [47,82]. New and existing statistics for epistasis should also be investigated in terms of their sensitivity to violations of assumptions as well as type I and type II error rates. Even in the face of statistical significance from *in vitro* studies, it remains to be seen how meaningful these intramolecular epistatic interactions may be in the context of a long-lived, multicellular plant. Whatever the case, if a particular order of mutations is required for enzymatic specialization, attempts to predict how to best engineer these kinds of enzymes for desired properties may be challenging [83].

6. Modern-day specialized metabolite pathways may evolve by sequential elaboration of ancestral pathway flux or by novel diversions

Modern-day plants exhibit pathways for specialized metabolite synthesis that are often catalysed by homologous enzymes borrowed from other pathways, including those of primary metabolism [12]. These enzymes are often promiscuous and this, coupled with the existence of non-enzymatic substrate conversions in many pathways, leads to the hypothesis that early metabolic networks would have produced an array of metabolites that might not be seen in modern-day plants [60,84]. One simple pathway that has been investigated using ASR is that of caffeine biosynthesis. Caffeine is biosynthetically derived from nucleotide metabolism and is found in flowers, where it can enhance pollinator memory [85], and in leaves, to provide herbivore defence [86]. In order to understand how the three terminal sequential methylation reactions of the pathway to caffeine evolved in the *Citrus*, *Theobroma* and *Paullinia* lineages, ancestral SABATH family methyltransferase enzymes were resurrected [44,46]. Since caffeine biosynthesis appears to employ either caffeine synthase (CS)-type or xanthine methyl transferase (XMT)-type enzymes (figure 1e,f), ASR was used on both paralogous groups of methyltransferases. In these three lineages, the multiple extant enzymes catalysing reactions in the modern-day caffeine pathways are inferred to have arisen from a single ancestral enzyme. In each case, the single ancestral enzyme was capable of performing multiple reactions within the xanthine alkaloid biosynthetic network to produce multiple different products. However, in no case was caffeine produced. Multistep ancient enzymes have been reported in other systems [52,87] and should be hypothesized to exist since earlier plants had less complex genomes to encode fewer enzymes to perform a comparable number of biosynthetic steps. Ultimately, independent gene duplication of the ancestral enzyme in each lineage gave rise to no fewer than two enzymes for the sequential methylation reactions towards caffeine. In each lineage, duplicated descendants became relatively more specialized to carry out different steps towards caffeine production. These results indicate that, at least for the final step of caffeine biosynthesis, this evolved last and built sequentially (or cumulatively) upon the reactions that had been put in place earlier. This model for pathway evolution assumes that the ancestral accumulation of metabolites would be selectively advantageous. Consistent with these assumptions, xanthine alkaloid precursors to caffeine can bind to adenosine receptors in rats and probably other mammals [88].

In a separate study, the origins of monoterpene indole alkaloid (MIA) biosynthesis in Apocynaceae species were investigated (figure 1c) [8]. MIAs are extremely diverse structures involved in plant defence and have a long history in traditional medicine but have now found a place in modern-day pharmaceuticals. The enzymes involved in aspidosperma- and iboga-type MIA biosynthesis in modern-day species appear to be derived from carboxylesterase-like cyclases and are classified as 2-hydroxyisoflavone dehydratase-like (HIDs) cyclase proteins. To understand how ancestral Apocynaceae species might have evolved to synthesize the current suite of MIAs of modern-day species, ancient HID-like enzymes were resurrected and assayed for activity with a central metabolic network precursor, angryline. Several of the ancestors appear to have been multistep enzymes that could channel products to more than one of the alkaloid pathways. In modern-day plants, the pathways towards three different types of MIAs do not sequentially build on ancestral products; rather, with each gene duplication, descendant enzymes evolved to divert flux from the same common ancestral precursor substrate, angryline, into either the aspidosperma- or the iboga-type alkaloid pathways [8]. Diversion of pathway flux after gene duplication has been hypothesized during the evolution of alpha-pyrone biosynthesis [89], making this a potentially more widespread phenomenon in plants. Specialized metabolite pathways are complex and may differ over developmental time as well as in different tissue types and in response to varying environmental stimuli; therefore, it is safe to say we are only at the beginning of our understanding of how pathways are evolutionarily assembled [87].

7. Embracing, and accounting for, uncertainty of ancestral sequence estimates

Of course, advancements towards understanding the evolution of plant specialized metabolism depend critically on the accuracy of ASR methods. However, because of the nature of the sequence estimation problem, there will always be uncertainty and any study using ASR should explicitly investigate the functional implications of it in some way [90]. A simple and frequently used approach to investigate uncertainty is to mutate a single site that has low posterior probability in the optimal sequence estimate and replace it with a suboptimal amino acid. Subsequently, the mutant and optimal protein can be compared in terms of enzymatic activity [43,70]. At the other end of the spectrum is the 'AltAll' approach, which advocates generating a second sequence in which the most likely amino acid is exchanged with the second-best amino acid at *every*

site where the alternative is above a particular probability threshold (e.g. 0.2), followed by functional characterization to compare with the results obtained from the optimal estimate [91]. Another approach is to use multiple different datasets and compare the functions of the optimal sequence obtained from each. This might include using different alignments or models of protein substitution or comparing the results of a nucleotide and corresponding amino acid alignment; one advantage of using nucleotides is that ancestral codons may be estimated that encode amino acids not seen in any modern-day protein sequences. Ancestral sequences estimated by different statistical methods can also be compared [50,92]. Bayesian approaches are intuitively appealing since they explicitly incorporate variations of tree topology, branch lengths and model parameters into estimates [23,92,93] although their accuracy does not appear to exceed that of ML [25]. Methods with different assumptions, such as ProtASR [94], which strives to account for protein folding stability and how it might evolve over time, could be used to compare with traditional approaches. Other studies approach the question of uncertainty in the optimal protein sequence by generating a second set of ancestral estimates for the same nodes using different datasets. This could be achieved using differently curated alignments that only include genomic sequences or combine genome and transcriptome data. Also, owing to the rapidly increasing amount of data available every year, sequence estimates could change over time. Catania *et al.* [45] used a 40-fold larger dataset to investigate the uncertainty of ancestral amino acid sequence estimates obtained in a previous study [43] that had a different tree topology, branch lengths and amino acid substitution model [45]. In this particular case, the sequence identity for the alternative alleles for the same nodes ranged from only 75 to 83% but showed similar enzyme activity. In conclusion, there is no single best way to assess uncertainty in an ASR study and any approach to experimentally characterize the variation associated with a sequence estimate is better than ignoring its existence. The crucial point is that if experimentally determined enzyme assay results of multiple ancestral variants are comparable, then we will have higher confidence that the results are not artefactual or biased.

8. Future studies

ASR appears to be a valuable approach to inform studies of plant specialized metabolism. In particular, ASR has provided insight that is impossible to glean from the study of modern-day enzymes only. For instance, the required mutational order leading to enzyme functional divergence, bygone ancestral enzymatic activities and the role of promiscuity in evolution can only be guessed without ASR. Therefore, several long-standing questions regarding the evolution of plant specialized metabolism should continue to benefit from its application. One underexplored area of research is to resurrect entire ancestral metabolic networks to understand pathway flux evolution more broadly. This might require the resurrection and co-expression of numerous ancestral enzymes within a single cellular system or could be conducted *in vitro*. These studies will be important to enhance our understanding of how ancestral components could be integrated into a network in a patchwork fashion and to what extent underground metabolism is pervasive [12,84]. While studies of ancestral metabolism *in vitro* can provide some insight into the molecular evolutionary aspects of specialized pathways, systems that would allow selection using pathogens or herbivores on transgenic plants (or other model systems) are sorely needed. These would allow a better understanding of whether promiscuous enzyme activities or differing ancestral pathway fluxes have fitness consequences [95]. Finally, since there is evidence for the importance of metabolons for specialized metabolite formation [96], investigations using ASR could reveal how and why these are evolutionarily assembled.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. This article has no additional data.

Declaration of AI use. I have not used AI-assisted technologies in creating this article.

Authors' contributions. T.J.B.: conceptualization, writing—original draft, writing—review and editing.

Conflict of interest declaration. I declare I have no competing interests.

Funding. This work was supported by National Science Foundation grant MCB-2325341 to T.J.B.

Acknowledgements. Andrew O'Donnell and two anonymous reviewers are thanked for providing valuable comments.

References

1. Pichersky E, Gang DR. 2000 Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends Plant Sci.* **5**, 439–445. (doi:10.1016/s1360-1385(00)01741-6)
2. Dudareva N, Pichersky E, Gershenzon J. 2004 Biochemistry of plant volatiles. *Plant Physiol.* **135**, 1893–1902. (doi:10.1104/pp.104.049981)
3. Maeda HA, Fernie AR. 2021 Evolutionary history of plant metabolism. *Annu. Rev. Plant Biol.* **72**, 185–216. (doi:10.1146/annurev-aplant-080620-031054)
4. 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 J.* **66**, 212–229. (doi:10.1111/j.1365-313X.2011.04520.x)
5. Kruse LH *et al.* 2022 Orthology-based analysis helps map evolutionary diversification and predict substrate class use of BAHD acyltransferases. *Plant J.* **111**, 1453–1468. (doi:10.1111/tpj.15902)
6. Köllner TG, Lenk C, Zhao N, Seidl-Adams I, Gershenzon J, Chen F, Degenhardt J. 2010 Herbivore-induced SABATH methyltransferases of maize that methylate anthranilic acid using S-adenosyl- \texttilde -methionine. *Plant Physiol.* **153**, 1795–1807. (doi:10.1104/pp.110.158360)
7. Wang J, De Luca V. 2005 The biosynthesis and regulation of biosynthesis of Concord grape fruit esters, including "foxy" methylanthranilate. *Plant J.* **44**, 606–619. (doi:10.1111/j.1365-313X.2005.02552.x)

8. DeMars MD, O'Connor SE. 2024 Evolution and diversification of carboxylesterase-like [4+2] cyclases in aspidosperma and iboga alkaloid biosynthesis. *Proc. Natl Acad. Sci. USA* **121**, e2318586121. (doi:10.1073/pnas.2318586121)
9. McCarthy AA, McCarthy JG. 2007 The structure of two *N*-methyltransferases from the caffeine biosynthetic pathway. *Plant Physiol.* **144**, 879–889. (doi:10.1104/pp.106.094854)
10. Pichersky E, Lewinsohn E. 2011 Convergent evolution in plant specialized metabolism. *Annu. Rev. Plant Biol.* **62**, 549–566. (doi:10.1146/annurev-aplant-042110-103814)
11. Lou YR, Pichersky E, Last RL. 2022 Deep roots and many branches: origins of plant-specialized metabolic enzymes in general metabolism. *Curr. Opin. Plant Biol.* **66**, 102192. (doi:10.1016/j.jpb.2022.102192)
12. Weng JK, Philippe RN, Noel JP. 2012 The rise of chemodiversity in plants. *Science* **336**, 1667–1670. (doi:10.1126/science.1217411)
13. Thornton JW. 2004 Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat. Rev. Genet.* **5**, 366–375. (doi:10.1038/nrg1324)
14. Selberg AGA, Gaucher EA, Liberles DA. 2021 Ancestral sequence reconstruction: from chemical paleogenetics to maximum likelihood algorithms and beyond. *J. Mol. Evol.* **89**, 157–164. (doi:10.1007/s00239-021-09993-1)
15. Harms MJ, Thornton JW. 2013 Evolutionary biochemistry: revealing the historical and physical causes of protein properties. *Nat. Rev. Genet.* **14**, 559–571. (doi:10.1038/nrg3540)
16. Harms MJ, Thornton JW. 2010 Analyzing protein structure and function using ancestral gene reconstruction. *Curr. Opin. Struct. Biol.* **20**, 360–366. (doi:10.1016/j.sbi.2010.03.005)
17. Lichman BR. 2022 Ancestral sequence reconstruction for exploring alkaloid evolution. *Methods Mol. Biol.* **2505**, 165–179. (doi:10.1007/978-1-0716-2349-7_12)
18. Scossa F, Fernie AR. 2021 Ancestral sequence reconstruction – an underused approach to understand the evolution of gene function in plants? *Comput. Struct. Biotechnol. J.* **19**, 1579–1594. (doi:10.1016/j.csbj.2021.03.008)
19. Hochberg GKA, Thornton JW. 2017 Reconstructing ancient proteins to understand the causes of structure and function. *Annu. Rev. Biophys.* **46**, 247–269. (doi:10.1146/annurev-biophys-070816-033631)
20. Cai W, Pei J, Grishin NV. 2004 Reconstruction of ancestral protein sequences and its applications. *BMC Evol. Biol.* **4**, 33. (doi:10.1186/1471-2148-4-33)
21. Yang ZH. 2007 PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591. (doi:10.1093/molbev/msm088)
22. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015 IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274. (doi:10.1093/molbev/msu300)
23. Randall RN, Radford CE, Roof KA, Natarajan DK, Gaucher EA. 2016 An experimental phylogeny to benchmark ancestral sequence reconstruction. *Nat. Commun.* **7**, 12847. (doi:10.1038/ncomms12847)
24. Zhang J, Nei M. 1997 Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. *J. Mol. Evol.* **44**, S139–S146. (doi:10.1007/pl00000067)
25. Hanson-Smith V, Kolaczkowski B, Thornton JW. 2010 Robustness of ancestral sequence reconstruction to phylogenetic uncertainty. *Mol. Biol. Evol.* **27**, 1988–1999. (doi:10.1093/molbev/msq081)
26. Vialle RA, Tamuri AU, Goldman N. 2018 Alignment modulates ancestral sequence reconstruction accuracy. *Mol. Biol. Evol.* **35**, 1783–1797. (doi:10.1093/molbev/msy055)
27. Li G, Steel M, Zhang L. 2008 More taxa are not necessarily better for the reconstruction of ancestral character states. *Syst. Biol.* **57**, 647–653. (doi:10.1080/10635150802203898)
28. Del Amparo R, Arenas M. 2022 Consequences of substitution model selection on protein ancestral sequence reconstruction. *Mol. Biol. Evol.* **39**, msac144. (doi:10.1093/molbev/msac144)
29. Salichos L, Rokas A. 2013 Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* **497**, 327–331. (doi:10.1038/nature12130)
30. Arenas M, Posada D. 2010 The effect of recombination on the reconstruction of ancestral sequences. *Genetics* **184**, 1133–1139. (doi:10.1534/genetics.109.113423)
31. Groussin M, Hobbs JK, Szöllösi GJ, Gribaldo S, Arcus VL, Gouy M. 2015 Toward more accurate ancestral protein genotype–phenotype reconstructions with the use of species tree-aware gene trees. *Mol. Biol. Evol.* **32**, 13–22. (doi:10.1093/molbev/msu305)
32. Jowkar G, Pečerska J, Maiolo M, Gil M, Anisimova M. 2023 ARPIP: ancestral sequence reconstruction with insertions and deletions under the Poisson indel process. *Syst. Biol.* **72**, 307–318. (doi:10.1093/sysbio/syac050)
33. Edwards RJ, Shields DC. 2004 GASP: gapped ancestral sequence prediction for proteins. *BMC Bioinform.* **5**, 123. (doi:10.1186/1471-2105-5-123)
34. Aadland K, Kolaczkowski B. 2020 Alignment-integrated reconstruction of ancestral sequences improves accuracy. *Genome Biol. Evol.* **12**, 1549–1565. (doi:10.1093/gbe/eva164)
35. Shih PM, Occhialini A, Cameron JC, Andralojc PJ, Parry MAJ, Kerfeld CA. 2016 Biochemical characterization of predicted Precambrian RuBisCO. *Nat. Commun.* **7**, 10382. (doi:10.1038/ncomms10382)
36. Lin MT, Salihovic H, Clark FK, Hanson MR. 2022 Improving the efficiency of Rubisco by resurrecting its ancestors in the family Solanaceae. *Sci. Adv.* **8**, eabm6871. (doi:10.1126/sciadv.abm6871)
37. Schulz L *et al.* 2022 Evolution of increased complexity and specificity at the dawn of form I Rubiscos. *Science* **378**, 155–160. (doi:10.1126/science.abq1416)
38. Priyadarshini N, Steube N, Wiens D, Narikawa R, Wilde A, Hochberg GKA, Enomoto G. 2023 Evidence for an early green/red photocycle that precedes the diversification of GAF domain photoreceptor cyanobacteriochromes. *Photochem. Photobiol. Sci.* **22**, 1415–1427. (doi:10.1007/s43630-023-00387-4)
39. Zhang Z, Coenen H, Ruelens P, Hazarika RR, Al Hindi T, Ogus GK, Vandeperre A, van Noort V, Geuten K. 2018 Resurrected protein interaction networks reveal the innovation potential of ancient whole-genome duplication. *Plant Cell* **30**, 2741–2760. (doi:10.1105/tpc.18.00409)
40. Steube N *et al.* 2023 Fortuitously compatible protein surfaces primed allosteric control in cyanobacterial photoprotection. *Nat. Ecol. Evol.* **7**, 756–767. (doi:10.1038/s41559-023-02018-8)
41. Chantreau M, Poux C, Lensink MF, Brysbaert G, Vekemans X, Castric V. 2019 Asymmetrical diversification of the receptor-ligand interaction controlling self-incompatibility in *Arabidopsis*. *eLife* **8**, e50253. (doi:10.7554/eLife.50253)
42. Jia H, Aadland K, Kolaczkowski O, Kolaczkowski B. 2021 Direct molecular evidence for an ancient, conserved developmental toolkit controlling posttranscriptional gene regulation in land plants. *Mol. Biol. Evol.* **38**, 4765–4777. (doi:10.1093/molbev/msab201)
43. Huang R, Hippauf F, Rohrbeck D, Haustein M, Wenke K, Feike J, Sorrelle N, Piechulla B, Barkman TJ. 2012 Enzyme functional evolution through improved catalysis of ancestrally nonpreferred substrates. *Proc. Natl Acad. Sci. USA* **109**, 2966–2971. (doi:10.1073/pnas.1019605109)
44. Huang R, O'Donnell AJ, Barboline JJ, Barkman TJ. 2016 Convergent evolution of caffeine in plants by co-option of exapted ancestral enzymes. *Proc. Natl Acad. Sci. USA* **113**, 10613–10618. (doi:10.1073/pnas.1602575113)
45. Catania EM, Dubs NM, Soumen S, Barkman TJ. 2024 The mutational road not taken: using ancestral sequence resurrection to evaluate the evolution of plant enzyme substrate preferences. *Genome Biol. Evol.* **16**, evae016. (doi:10.1093/gbe/evae016)
46. O'Donnell AJ, Huang R, Barboline JJ, Barkman TJ. 2021 Convergent biochemical pathways for xanthine alkaloid production in plants evolved from ancestral enzymes with different catalytic properties. *Mol. Biol. Evol.* **38**, 2704–2714. (doi:10.1093/molbev/msab059)

47. Smith SD, Wang S, Rausher MD. 2013 Functional evolution of an anthocyanin pathway enzyme during a flower color transition. *Mol. Biol. Evol.* **30**, 602–612. (doi:10.1093/molbev/mss255)

48. Lichman BR *et al.* 2020 The evolutionary origins of the cat attractant nepetalactone in catnip. *Sci. Adv.* **6**, eaba0721. (doi:10.1126/sciadv.aba0721)

49. Kaltenbach M, Burke JR, Dindo M, Pabis A, Munsberg FS, Rabin A, Kamerlin SCL, Noel JP, Tawfik DS. 2018 Evolution of chalcone isomerase from a noncatalytic ancestor. *Nat. Chem. Biol.* **14**, 548–555. (doi:10.1038/s41589-018-0042-3)

50. Devamani T *et al.* 2016 Catalytic promiscuity of ancestral esterases and hydroxynitrile lyases. *J. Am. Chem. Soc.* **138**, 1046–1056. (doi:10.1021/jacs.5b12209)

51. Rauwerdink A, Lunzer M, Devamani T, Jones B, Mooney J, Zhang ZJ, Xu JH, Kazlauskas RJ, Dean AM. 2016 Evolution of a catalytic mechanism. *Mol. Biol. Evol.* **33**, 971–979. (doi:10.1093/molbev/msv338)

52. Vinde MH *et al.* 2022 Ancestral sequence reconstruction of the CYP711 family reveals functional divergence in strigolactone biosynthetic enzymes associated with gene duplication events in monocot grasses. *New Phytol.* **235**, 1900–1912. (doi:10.1111/nph.18285)

53. Luck K, Chen X, Norris AM, Chen F, Gershenzon J, Köllemer TG. 2020 The reconstruction and biochemical characterization of ancestral genes furnish insights into the evolution of terpene synthase function in the Poaceae. *Plant Mol. Biol.* **104**, 203–215. (doi:10.1007/s11103-020-01037-4)

54. Zhang Z, Huang L, Zhang ZJ, Xu JH, Yu HL. 2023 Rational design of taxadiene hydroxylase by ancestral enzyme construction and the elucidation of key amino acids. *Biochemistry* **62**, 3214–3221. (doi:10.1021/acs.biochem.3c00411)

55. Sun Y, Calderini E, Kourist R. 2021 A reconstructed common ancestor of the fatty acid photo-decarboxylase clade shows photo-decarboxylation activity and increased thermostability. *ChemBioChem* **22**, 1833–1840. (doi:10.1002/cbic.202000851)

56. Broun P, Shanklin J, Whittle E, Somerville C. 1998 Catalytic plasticity of fatty acid modification enzymes underlying chemical diversity of plant lipids. *Science* **282**, 1315–1317. (doi:10.1126/science.282.5392.1315)

57. Leong BJ, Last RL. 2017 Promiscuity, impersonation and accommodation: evolution of plant specialized metabolism. *Curr. Opin. Struct. Biol.* **47**, 105–112. (doi:10.1016/j.sbi.2017.07.005)

58. Khersonsky O, Roodveldt C, Tawfik DS. 2006 Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr. Opin. Chem. Biol.* **10**, 498–508. (doi:10.1016/j.cbpa.2006.08.011)

59. Conant GC, Wolfe KH. 2008 Turning a hobby into a job: how duplicated genes find new functions. *Nat. Rev. Genet.* **9**, 938–950. (doi:10.1038/nrg2482)

60. Jensen RA. 1976 Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* **30**, 409–425. (doi:10.1146/annurev.mi.30.100176.002205)

61. Bergthorsson U, Andersson DI, Roth JR. 2007 Ohno's dilemma: evolution of new genes under continuous selection. *Proc. Natl. Acad. Sci. USA* **104**, 17004–17009. (doi:10.1073/pnas.0707158104)

62. Piatigorsky J, Wistow G. 1991 The recruitment of crystallins: new functions precede gene duplication. *Science* **252**, 1078–1079. (doi:10.1126/science.252.5009.1078)

63. Barve A, Wagner A. 2013 A latent capacity for evolutionary innovation through exaptation in metabolic systems. *Nature* **500**, 203–206. (doi:10.1038/nature12301)

64. Schulenburg C, Miller BG. 2014 Enzyme recruitment and its role in metabolic expansion. *Biochemistry* **53**, 836–845. (doi:10.1021/bi401667f)

65. Copley SD, Newton MS, Widney KA. 2023 How to recruit a promiscuous enzyme to serve a new function. *Biochemistry* **62**, 300–308. (doi:10.1021/acs.biochem.2c00249)

66. Voordeckers K, Brown CA, Vanneste K, van der Zande E, Voet A, Maere S, Verstrepen KJ. 2012 Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication. *PLoS Biol.* **10**, e1001446. (doi:10.1371/journal.pbio.1001446)

67. Jones BJ *et al.* 2020 Larger active site in an ancestral hydroxynitrile lyase increases catalytically promiscuous esterase activity. *PLoS One* **15**, e0235341. (doi:10.1371/journal.pone.0235341)

68. O'Brien PJ, Herschlag D. 1999 Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.* **6**, R91–R105. (doi:10.1016/S1074-5521(99)80033-7)

69. Aharoni A, Gaidukov L, Khersonsky O, McQ Gould S, Roodveldt C, Tawfik DS. 2005 The 'evolvability' of promiscuous protein functions. *Nat. Genet.* **37**, 73–76. (doi:10.1038/ng1482)

70. Ortlund EA, Bridgman JT, Redinbo MR, Thornton JW. 2007 Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* **317**, 1544–1548. (doi:10.1126/science.1142819)

71. Starr TN, Thornton JW. 2016 Epistasis in protein evolution. *Protein Sci.* **25**, 1204–1218. (doi:10.1002/pro.2897)

72. Kaltenbach M, Tokuriki N. 2014 Dynamics and constraints of enzyme evolution. *J. Exp. Zool. B Mol. Dev. Evol.* **322**, 468–487. (doi:10.1002/jez.b.22562)

73. Weinreich DM, Delaney NF, Depristo MA, Hartl DL. 2006 Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114. (doi:10.1126/science.1123539)

74. Miller SP, Lunzer M, Dean AM. 2006 Direct demonstration of an adaptive constraint. *Science* **314**, 458–461. (doi:10.1126/science.1133479)

75. Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans SJ. 2007 Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* **445**, 383–386. (doi:10.1038/nature05451)

76. Tufts DM, Natarajan C, Revsbech IG, Projecto-Garcia J, Hoffmann FG, Weber RE, Fago A, Moriyama H, Storz JF. 2015 Epistasis constrains mutational pathways of hemoglobin adaptation in high-altitude pikas. *Mol. Biol. Evol.* **32**, 287–298. (doi:10.1093/molbev/msu311)

77. Smith SD, Rausher MD. 2011 Gene loss and parallel evolution contribute to species difference in flower color. *Mol. Biol. Evol.* **28**, 2799–2810. (doi:10.1093/molbev/msr109)

78. Yang G *et al.* 2019 Higher-order epistasis shapes the fitness landscape of a xenobiotic-degrading enzyme. *Nat. Chem. Biol.* **15**, 1120–1128. (doi:10.1038/s41589-019-0386-3)

79. Natarajan C, Inoguchi N, Weber RE, Fago A, Moriyama H, Storz JF. 2013 Epistasis among adaptive mutations in deer mouse hemoglobin. *Science* **340**, 1324–1327. (doi:10.1126/science.1236862)

80. Xu M, Wilderman PR, Peters RJ. 2007 Following evolution's lead to a single residue switch for diterpene synthase product outcome. *Proc. Natl. Acad. Sci. USA* **104**, 7397–7401. (doi:10.1073/pnas.0611454104)

81. O'Maille PE *et al.* 2008 Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases. *Nat. Chem. Biol.* **4**, 617–623. (doi:10.1038/nchembio.113)

82. Miton CM, Chen JZ, Ost K, Anderson DW, Tokuriki N. 2020 Statistical analysis of mutational epistasis to reveal intramolecular interaction networks in proteins. *Methods Enzymol.* **643**, 243–280. (doi:10.1016/bs.mie.2020.07.012)

83. Miton CM, Tokuriki N. 2016 How mutational epistasis impairs predictability in protein evolution and design. *Protein Sci.* **25**, 1260–1272. (doi:10.1002/pro.2876)

84. Noda-Garcia L, Liebermeister W, Tawfik DS. 2018 Metabolite–enzyme coevolution: from single enzymes to metabolic pathways and networks. *Annu. Rev. Biochem.* **87**, 187–216. (doi:10.1146/annurev-biochem-062917-012023)

85. Wright GA, Baker DD, Palmer MJ, Stabler D, Mustard JA, Power EF, Borland AM, Stevenson PC. 2013 Caffeine in floral nectar enhances a pollinator's memory of reward. *Science* **339**, 1202–1204. (doi:10.1126/science.1228806)

86. Uefuji H, Tatsumi Y, Morimoto M, Kaothien-Nakayama P, Ogita S, Sano H. 2005 Caffeine production in tobacco plants by simultaneous expression of three coffee *N*-methyltransferases and its potential as a pest repellent. *Plant Mol. Biol.* **59**, 221–227. (doi:10.1007/s11103-005-8520-x)

87. Olson-Manning CF. 2020 Elaboration of the corticosteroid synthesis pathway in primates through a multistep enzyme. *Mol. Biol. Evol.* **37**, 2257–2267. (doi:10.1093/molbev/msaa080)
88. Daly JW, Butts-Lamb P, Padgett W. 1983 Subclasses of adenosine receptors in the central nervous system: interaction with caffeine and related methylxanthines. *Cell. Mol. Neurobiol.* **3**, 69–80. (doi:10.1007/BF00734999)
89. Weng JK, Li Y, Mo H, Chapple C. 2012 Assembly of an evolutionarily new pathway for α -pyrone biosynthesis in *Arabidopsis*. *Science* **337**, 960–964. (doi:10.1126/science.1221614)
90. Oliva A, Pulicani S, Lefort V, Bréhélin L, Gascuel O, Guindon S. 2019 Accounting for ambiguity in ancestral sequence reconstruction. *Bioinformatics* **35**, 4290–4297. (doi:10.1093/bioinformatics/btz249)
91. Eick GN, Bridgham JT, Anderson DP, Harms MJ, Thornton JW. 2017 Robustness of reconstructed ancestral protein functions to statistical uncertainty. *Mol. Biol. Evol.* **34**, 247–261. (doi:10.1093/molbev/msw223)
92. Hobbs JK, Shepherd C, Saul DJ, Demetras NJ, Haanen S, Monk CR, Daniel RM, Arcus VL. 2012 On the origin and evolution of thermophily: reconstruction of functional Precambrian enzymes from ancestors of *Bacillus*. *Mol. Biol. Evol.* **29**, 825–835. (doi:10.1093/molbev/msr253)
93. Williams PD, Pollock DD, Blackburne BP, Goldstein RA. 2006 Assessing the accuracy of ancestral protein reconstruction methods. *PLoS Comput. Biol.* **2**, e69. (doi:10.1371/journal.pcbi.0020069)
94. Arenas M, Weber CC, Liberles DA, Bastolla U. 2017 ProtASR: an evolutionary framework for ancestral protein reconstruction with selection on folding stability. *Syst. Biol.* **66**, 1054–1064. (doi:10.1093/sysbio/syw121)
95. Hobbs JK, Prentice EJ, Groussin M, Arcus VL. 2015 Reconstructed ancestral enzymes impose a fitness cost upon modern bacteria despite exhibiting favourable biochemical properties. *J. Mol. Evol.* **81**, 110–120. (doi:10.1007/s00239-015-9697-5)
96. Laursen T *et al.* 2016 Characterization of a dynamic metabolon producing the defense compound dhurrin in *Sorghum*. *Science* **354**, 890–893. (doi:10.1126/science.aag2347)