Convergent Biochemical Pathways for Xanthine Alkaloid Production in Plants Evolved from Ancestral Enzymes with Different Catalytic Properties

Andrew J. O’Donnell,† Ruiqi Huang,‡ Jessica J. Barboline, and Todd J. Barkman*

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

†Present address: Max Planck Institute for Chemical Ecology, Jena, Germany
‡Present address: Applied Biomedical Science Institute, San Diego, CA, USA

*Corresponding author: E-mail: todd.barkman@wmich.edu.
Associate editor: Belinda Chang

Abstract

Convergent evolution is widespread but the extent to which common ancestral conditions are necessary to facilitate the independent acquisition of similar traits remains unclear. In order to better understand how ancestral biosynthetic catalytic capabilities might lead to convergent evolution of similar modern-day biochemical pathways, we resurrected ancient enzymes of the caffeine synthase (CS) methyltransferases that are responsible for theobromine and caffeine production in flowering plants. Ancestral CS enzymes of Theobroma, Paullinia, and Camellia exhibited similar substrate preferences but these resulted in the formation of different sets of products. From these ancestral enzymes, descendants with similar substrate preference and product formation independently evolved after gene duplication events in Theobroma and Paullinia. Thus, it appears that the convergent modern-day pathways likely originated from ancestral pathways with different inferred flux. Subsequently, the modern-day enzymes originated independently via gene duplication and their convergent catalytic characteristics evolved to partition the multiple ancestral activities by different mutations that occurred in homologous regions of the ancestral proteins. These results show that even when modern-day pathways and recruited genes are similar, the antecedent conditions may be distinctive such that different evolutionary steps are required to generate convergence.

Key words: convergent evolution, ancestral sequence resurrection, enzyme evolution, caffeine.

Introduction

The ubiquity of convergent trait acquisition throughout the tree of life suggests that evolution may be, at least to some extent, repeatable due to some combination of constraint on one hand and strength of natural selection on the other (Storz 2016). One framework by which to conceptualize and study mechanisms of convergence emphasizes dissection of the biological hierarchy through study of the pathways (biochemical or developmental), underlying genes, encoded protein functions, and mutational paths involved (Des Marais and Rausher 2010; Manceau et al. 2010; Losos 2011). At one end of the spectrum, convergent traits may arise from different pathways, genes, and sets of mutations, whereas at the other, in principle, it is possible to have the same pathway, generate a similar phenotype in independent lineages that are realized by orthologous genes that acquired their novel functions by identical mutations to the same ancestral nucleotides (Zhang 2006; Christin et al. 2010; Manceau et al. 2010; Storz 2016). Although convergent evolution has been studied in terms of the underlying modern-day pathways and genes recruited, the ancient underpinnings of the independent origins of traits are less well understood. For instance, in the case of convergently evolved metabolites generated by the same biosynthetic pathways (Pichersky and Lewinsohn 2011), ancestral biochemical flux may or may not have been the same as that exhibited by modern-day descendant species. Computational models indicate that evolutionary changes in pathway flux may be predictable and are influenced by biochemical network structure, gene expression levels, and kinetic properties of the enzymes involved (Wheeler and Smith 2019). One way to investigate the historical sequence of independent pathway evolution involves ancestral sequence resurrection (ASR) (Thornton 2004; Dean and Thornton 2007) which provides an experimental means to directly determine the genetic and biochemical changes leading to convergence (Zhang 2006; Natarajan et al. 2016). This experimental method has provided novel insight into protein functional evolution in several different systems (Chang et al. 2004).
Convergent Biochemical Pathways for Xanthine Alkaloid Production in Plants

Evolution of TB and CF Production in Paullinia, Theobroma, and Camellia Originated from Ancestral CS Enzymes with Similar Substrate Preferences to Form Distinct Sets of Products

Even though whole-genome sequences were queried, CS orthologs appear to be lacking from most asterid lineages like Lamiales and Solanales and seem to be encoded in only a scattered set of angiosperm lineages as shown in figure 2A (see also supplementary fig. S2, Supplementary Material online). The only functionally characterized CS enzymes are known from Sapindales, Malvales, and Ericales. These are also the only lineages from which xanthine alkaloids (CF or precursors) are reliably known of those shown in fig. 2. Nonetheless, primary routes to CF biosynthesis have been hypothesized based on a combination of documented gene coexpression patterns, enzyme substrate preferences, and detection of intermediate metabolites in metabolomic and radioisotope tracer studies (Kato et al. 1996; Mizuno, Okuda, et al. 2003; Uefuji et al. 2003; Kato and Mizuno 2004; Ashihara et al. 2008; Huang et al. 2016; Deng et al. 2020). Angiosperm CS- and XMT-type enzymes have been characterized mainly from CF-producing species; yet, orthologs are known from other non-xanthine alkaloid-accumulating relatives (Huang et al. 2016). In these cases, it is unclear why the genes would have been maintained over long periods of time if they were not involved in CF biosynthesis. In Mangifera, the only non-CF-producing species studied, it was shown that benzoic acid is the preferred substrate for its XMT ortholog with no xanthine alkaloid activity detected. Benzoic and salicylic acid methylation was also shown for the ancestral angiosperm XMT protein which suggests that xanthine alkaloid methylation only recently evolved in descendant enzymes (Huang et al. 2016).

The convergent xanthine alkaloid pathways in Paullinia (Sapindales) and Theobroma (Malvales) catalyzed by recently and independently duplicated orthologous CS enzymes makes this a remarkable case of convergence because they are members of lineages that diverged at least 60 Ma (Huang et al. 2016; Zeng et al. 2017). As they have both converged on the use of orthologous proteins to catalyze the same pathway, we predict that each lineage would have possessed progenitor enzymes with similar catalytic properties that would allow for evolution of the modern-day pathway following the same biochemical steps from X to 3X to TB. Furthermore, it has been demonstrated that single amino acid substitutions within the predicted active site in XMT enzymes were sufficient for specialization on xanthine alkaloid substrates in the Citrus lineage (Huang et al. 2016). Therefore, we predict that mutations leading to changes in substrate preference of CS enzymes are also constrained to the same active site regions in spite of their ancient divergence. In this study, we resurrected ancestral enzymes at key branching points of the CS lineage to test these predictions.

Results and Discussion

Evolution of TB and CF Production in Paullinia, Theobroma, and Camellia Originated from Ancestral CS Enzymes with Similar Substrate Preferences to Form Distinct Sets of Products

Even though whole-genome sequences were queried, CS orthologs appear to be lacking from most asterid lineages like Lamiales and Solanales and seem to be encoded in only a scattered set of angiosperm lineages as shown in figure 2A (see also supplementary fig. S2, Supplementary Material online). The only functionally characterized CS enzymes are known from Sapindales, Malvales, and Ericales. These are also the only lineages from which xanthine alkaloids (CF or precursors) are reliably known of those shown in fig. 2A (Ashihara and Suzuki 2004). Thus, it is unclear what the activity of CS orthologs from Myrtales, Geraniales, and Cornales

2002; Bridgham et al. 2006, 2009; Gaucher et al. 2008; Field and Matz 2010; Smith et al. 2013; Lichman et al. 2020). Utility of ASR was shown to be particularly valuable for revealing ancestral changes to enzymes involved in the corticosteroid pathway that allowed for biosynthetic elaboration in primates (Olson-Manning 2020).

One example of convergent trait evolution in angiosperms is the biosynthesis of xanthine alkaloids, like caffeine (CF), which have several ecological roles (Suzuki and Waller 1987; Baumann et al. 1995; Ashihara and Crozier 2001; Uefuji et al. 2005; Anaya et al. 2006; Wright et al. 2013). CF is produced in various plant tissues via the sequential methylation of xanthine alkaloid precursors by either caffeine synthase (CS) or xanthine methyltransferase (XMT) enzymes that are paralogs in the SABATH family of S-adenosyl-l-methionine (SAM)-dependent methyltransferases (Suzuki and Takahashi 1976; Mazzafra et al. 1994; Ashihara et al. 1996; Mizuno, Kato, et al. 2003; Kato and Mizuno 2004) (fig. 1 and supplementary fig. S1, Supplementary Material online). Coffee spp. (coffee) and Citrus sinensis (citrus) convergently utilize XMT enzymes to produce CF (Uefuji et al. 2003; McCarthy AA and McCarthy JG 2007; Denoeud et al. 2014; Huang et al. 2016; supplementary fig. S1, Supplementary Material online). Yet, these orthologous enzymes catalyze divergent pathways primarily via xanthosine (XR), 7-methylxanthine (7X), and theophylline (TB) in Coffea and xanthine (X), 1- and 3-methylxanthine (1X, 3X), and theophylline (TP) in Citrus (fig. 1). The pathway in Camellia sinensis (tea) is thought to be convergent with Coffea, even though it included CS enzymes to catalyze the same reactions (Kato et al. 2000) (fig. 1 and supplementary fig. S1, Supplementary Material online). Paullinia cupana (guaraná) and Theobroma cacao (cocoa) convergently utilize CS-type enzymes to methylate xanthine alkaloids as in Camellia (Huang et al. 2016) (supplementary fig. S1, Supplementary Material online). However, the CS enzymes of Paullinia and Theobroma primarily convert X to 3X and 3X to TB (fig. 1). Surprisingly, the CS1 enzymes in Paullinia and Theobroma that convert X to 3X appear to have evolved independently after gene duplication, as have their CS2 enzymes that catalyze the conversion of 3X to TB (Huang et al. 2016). A CS-type enzyme has been reported to convert TB to CF in Paullinia but none is yet known for Theobroma (Schimpl et al. 2014; Huang et al. 2016) (fig. 1). The multiple modern-day CS enzymes in Paullinia, Theobroma, and Camellia are part of an ancient lineage but, like the XMT enzymes of Coffea and Citrus, arise by independent duplication in each caffeine-producing group more recently (Huang et al. 2016) (fig. S1, Supplementary Material online).

Individual CS or XMT enzymes may generate considerable biochemical complexity due to their potential to 1) methylate multiple substrates in the pathway and 2) produce multiple products from single substrates due to methylation of different ring N atoms (see N1, N3, or N7 in fig. 1) (Kato et al. 2000; McCarthy AA and McCarthy JG 2007; Huang et al. 2016). This enzymatic complexity, combined with the fact that intermediates in the pathway do not accumulate to appreciable levels, makes it difficult to determine for any one species if flux is linear or highly branched throughout the metabolic network shown in figure 1. Nonetheless, primary routes to CF biosynthesis have been hypothesized based on a combination of documented gene coexpression patterns, enzyme substrate preferences, and detection of intermediate metabolites in metabolomic and radioisotope tracer studies (Kato et al. 1996; Mizuno, Okuda, et al. 2003; Uefuji et al. 2003; Kato and Mizuno 2004; Ashihara et al. 2008; Huang et al. 2016; Deng et al. 2020). Angiosperm CS- and XMT-type enzymes have been characterized mainly from CF-producing species; yet, orthologs are known from other non-xanthine alkaloid-accumulating relatives (Huang et al. 2016). In these cases, it is unclear why the genes would have been maintained over long periods of time if they were not involved in CF biosynthesis. In Mangifera, the only non-CF-producing species studied, it was shown that benzoic acid is the preferred substrate for its XMT ortholog with no xanthine alkaloid activity detected. Benzoic and salicylic acid methylation was also shown for the ancestral angiosperm XMT protein which suggests that xanthine alkaloid methylation only recently evolved in descendant enzymes (Huang et al. 2016).

The convergent xanthine alkaloid pathways in Paullinia (Sapindales) and Theobroma (Malvales) catalyzed by recently and independently duplicated orthologous CS enzymes makes this a remarkable case of convergence because they are members of lineages that diverged at least 60 Ma (Huang et al. 2016; Zeng et al. 2017). As they have both converged on the use of orthologous proteins to catalyze the same pathway, we predict that each lineage would have possessed progenitor enzymes with similar catalytic properties that would allow for evolution of the modern-day pathway following the same biochemical steps from X to 3X to TB. Furthermore, it has been demonstrated that single amino acid substitutions within the predicted active site in XMT enzymes were sufficient for specialization on xanthine alkaloid substrates in the Citrus lineage (Huang et al. 2016). Therefore, we predict that mutations leading to changes in substrate preference of CS enzymes are also constrained to the same active site regions in spite of their ancient divergence. In this study, we resurrected ancestral enzymes at key branching points of the CS lineage to test these predictions.
may be because none of their proteins have been functionally characterized. One potential role for the orthologs from these lineages is trigonelline biosynthesis which also requires ring nitrogen methylation and has been shown for an XMT-type enzyme in *Coffea* (Mizuno et al. 2014). Alternatively, it may be that CS enzymes in all of the lineages shown in figure 2 are involved in xanthine alkaloid biosynthesis to produce CF at low, difficult-to-detect levels. If so, the distribution of this important stimulant could be much more widespread in angiosperms than is currently appreciated.

To investigate the characteristics of ancient CS-type enzymes in angiosperms, we experimentally resurrected and characterized at least two ancestral allelic variants for each of the progenitors predicted for *Paullinia* (PcAncCS1), *Theobroma* (TcAncCS1), and *Camellia* (CsAncCS) (fig. 2C–E and supplementary figs. S1, S3–S5, Supplementary Material online). Surprisingly, all three of these ancestral enzymes had highest relative activity with 7X even though preference for the substrate is not exhibited by the modern-day CS1 and CS2 enzymes derived from each of them (fig. 2C–E and supplementary figs. S3–S5, Supplementary Material online). This similar methylation activity was exhibited in spite of a substantial degree of sequence divergence among these ancestral CS enzymes that ranges from 51% to 63%. All three ancestral enzymes converted 7X to TB by methylation at N3, but in addition, TcAncCS1 and CsAncCS converted 7X to paraxanthine (PX) by methylation at the N1 position (fig. 2C–E and supplementary table S1, Supplementary Material online). Each ancestral enzyme also had lower relative activity with X: In PcAncCS1, X was converted to 3X, whereas TcAncCS1 and CsAncCS methylated X to produce both 1X and 3X (fig. 2C–E and supplementary table S1, Supplementary Material online). The product of the reaction with 3X was below our detection limit for PcAncCS1 and CsAncCS, but in TcAncCS1, 3X was converted to TP (fig. 2D and supplementary table S1, Supplementary Material online). Thus, N1 and N3 methylation were properties of these ancestral enzymes but none appear to methylate the N7 position of any substrate we tested.

Since each of the ancestral enzymes in the *Paullinia*, *Theobroma*, and *Camellia* lineages appears to have been capable of converting xanthine to a monomethylated product (1X and 3X) (fig. 2C–E), they may have provided the beginnings of the modern-day CF biosynthetic pathways. This is predicted under the cumulative hypothesis for pathway
evolution in which earlier steps are predicted to evolve first (Granick 1957; Huang et al. 2016); in addition, TcAncCS1 could also have formed the dimethylated product, TP. If these molecules were to accumulate in ancestral plant tissues, they could have conferred a selective advantage which would likely result in retention of the ancient genes because 1X and 3X have been shown to bind to modern-day rat adenosine receptors (Daly et al. 1983) and TP can modulate Adenylate Cyclase in insects (Nathanson 1984). Their formation in tissues is tenable since PcAncCS1 and TcAncCS1 have apparent $K_M$ estimates for X that are comparable with modern-day CS enzymes (53–417 $\mu$M) (table 1) (Huang et al. 2016). Alternatively, if ancestral X methylation was not physiologically relevant, then their high 7X activity was
Ancestral enzymes and Theobroma and Camellia

Ancestral preference is an enzymatic characteristic also latency that we have not assayed for, it is apparent that ancestral 7X preference is an enzymatic characteristic also maintained by modern-day CS-type enzymes in non-caffeine accumulating Camellia CsAncCS appears to be an ancient characteristic that has been maintained over 100 My and is utilized as an enzymatic step for modern-day CF production (Kato et al. 2000) (fig. 1 and 2). Even if CS genes were maintained for some alternative function outside of xanthine alkaloid metabolism that we have not assayed for, it is apparent that ancestral 7X preference is an enzymatic characteristic also maintained by modern-day CS-type enzymes in non-caffeine accumulating Camellia (Ishida et al. 2009) and Theobroma (Yoneyama et al. 2006), which appears to contribute to TB formation.

Although each of the ancestral enzymes catalyzed a unique set of xanthine alkaloid products, Paullinia and Theobroma ultimately evolved convergent pathways to TB biosynthesis (figs. 1, 2C, and 2D) (Huang et al. 2016). At a minimum, this would have required independent acquisition of N7 methylation of 3X by Paullinia and Theobroma CS enzymes in order for them to produce TB, the second metabolite in the pathway toward CF (fig. 1, 2C, and 2D). To establish how this change to xanthine alkaloid metabolism occurred, we resurrected and experimentally characterized the younger ancestral enzymes that descended from PcAncCS1 and TcAncCS1 and ultimately gave rise to the specialized modern-day CS1 and CS2 paralogs independently in Paullinia and Theobroma.

Convergent Duplication Events of Paullinia and Theobroma Ancestral CS Enzymes Allowed for a Connected Pathway to TB Biosynthesis through Convergent Catalytic Changes

Gene duplication of PcAncCS1 in the Paullinia CS lineage (Node M) gave rise to PcAncCS2 (Node N) and a modern-day descendant, Paullinia “CS-like,” for which no enzymatic activity was detected (fig. 2C). In both allelic variants of PcAncCS2, as in the ancestor PcAncCS1, activity with 7X was still highest to form TB and PX, whereas methylation with X to form 3X remained relatively low; however, in this descendant, higher relative activity with 3X evolved to form TB by N7 methylation (fig. 2C, Node N; supplementary fig. S6 and table S1, Supplementary Material online). As a result, this change would have allowed for extension of the caffeine pathway by sequential methylation from the ancestral step of X to 3X to include 3X to TB in a manner consistent with the cumulative hypothesis which predicts that later evolved biosynthetic steps build upon earlier ones within a pathway (Granick 1957).

In Theobroma, duplication of TcAncCS1 at Node O led to the evolution of TcAncCS2 (Node P) as well as Theobroma BTS (fig. 2D). Whereas BTS retained ancestral preference to methylate 7X (Yoneyama et al. 2006), both allelic variants of TcAncCS2 suggest it lost ancestral preference for 7X and evolved highest relative activity with X, converting it solely to 3X (fig. 2D and supplementary fig. S7 and table S1, Supplementary Material online). This result represents specialization for N3 methylation due to loss of N1 activity with xanthine. TcAncCS2 also converted 3X to TB, which required acquisition of N7 methylation and loss of N1 methylation of the substrate (fig. 2D and supplementary table S1, Supplementary Material online). Thus, both Theobroma and Paullinia ancestral enzymes acquired N7 methylation of 3X following the earliest known duplications that occurred in each lineage (fig. 2C and D). As these independent gene duplication events gave rise to descendant enzymes capable of performing the sequential methylation steps required to convert X to TB, they could be viewed as convergent changes that facilitated the evolution of the same pathway steps in both lineages. In the case of Theobroma, it appears that the first two biochemical steps catalyzed by its modern-day enzymes evolved by a diversion away from ancestral flux to 1X from X and TP from 3X as seen in TcAncCS1 to primarily X to 3X to TB, rather than being gradually assembled one reaction at a time as in Paullinia (fig. 2C and D).

The restructuring of relative substrate preferences and methylation patterns observed in TcAncCS2 (fig. 2D) prompted further investigation into its kinetic properties as well as those of TcAncCS1 from which it descended; it was not clear if the relative preference shift toward X in TcAncCS2 was due to higher affinity for it or merely a relative loss in recognition of other substrates. The Michaelis–Menten kinetic parameter estimates summarized in table 1 indicate that TcAncCS2 had a $k_{cat}/K_M$ of 82.1 s$^{-1}$ M$^{-1}$ with X which is 37-fold higher than 3X (and likely that of 7X for which activity was too low to determine kinetic parameters), whereas in TcAncCS1, the $k_{cat}/K_M$ with X was comparable with that of 3X but four times lower than 7X (table 1). The apparent improvement of TcAncCS2 with xanthine relative to other substrates was concomitant with a switch away from N1 toward N3 methylation and appears to have set ancestral flux into the beginning of the pathway to TB and CF via 3X predicted for modern-day Theobroma (Huang et al. 2016). However, even though TcAncCS2 evolved to perform the

| Enzyme (substrate) | $K_M$ (mM) | $k_{cat}$ (1/s) | $k_{cat}/K_M$ (s$^{-1}$M$^{-1}$) |
|-------------------|-----------|----------------|-------------------------------|
| TcCS1 (X)         | 95.80     | 8.37E$-$05     | 0.87                          |
| TcCS2 (3X)        | 49.10     | 9.81E$-$05     | 2.00                          |
| PcCS1 (X)         | 95.38     | 1.52E$-$03     | 15.94                         |
| PcCS2 (3X)        | 677.00    | 9.33E$-$04     | 1.38                          |
| Ancestral enzymes |           |                |                               |
| TcAncCS1 (X)      | 53.40     | 8.65E$-$05     | 1.62                          |
| TcAncCS1 (3X)     | 138.60    | 1.28E$-$04     | 0.92                          |
| TcAncCS1 (7X)     | 34.60     | 2.47E$-$04     | 7.20                          |
| TcAncCS2 (X)      | 4.14      | 3.39E$-$04     | 82.06                         |
| TcAncCS2 (3X)     | 154.00    | 3.32E$-$04     | 2.16                          |
| PcAncCS1 (X)      | 417.50    | 3.76E$-$04     | 0.90                          |
| CsAncCS (7X)      | 26.70     | 1.05E$-$03     | 39.33                         |
two successive enzymatic steps to TB, kinetic estimates predict that flux may have been low for this single ancestral enzyme. The $k_{cat}/K_m$ for TcAncCS2 with 3X, the product of the reaction with X and second substrate in the pathway to TB, was estimated to be 2.2 s$^{-1}$ M$^{-1}$ (table 1). Because the specificity constant of TcAncCS2 for 3X is nearly 40 times lower than that of xanthine, from which it is biochemically derived, the conversion of 3X to TB is predicted to have been low in the presence of this single ancestral enzyme, unless cellular concentrations of X were to become depleted.

After the origin of PcAncCS2 by duplication from PcAncCS1, it was subsequently duplicated again resulting in the evolution of one descendant, PcCS1, that acquired near-complete preference for X to form 3X and a second descendant, PcCS2, that evolved specificity for 3X to form TB (Huang et al. 2016) (fig. 2C). Likewise, initial gene duplication of TcAncCS1 gave rise to TcAncCS2; this daughter gene was then duplicated later to result in modern-day TcCS1, which prefers to methylate X to form 3X, and TcCS2, which prefers to methylate 3X to form TB (but also has lower relative activity with X to form 7X) (Huang et al. 2016) (fig. 2D). These independent duplications in the two lineages therefore represent additional convergence between Paullinia and Theobroma at the level of genes involved; it was not until the duplications of PcAncCS2 and TcAncCS2 that paralogs with the same substrate preferences emerged in the two lineages. In both cases, this likely resulted in a more catalytically favorable pathway from X to 3X to TB. Although we were not able to estimate kinetic parameters for PcAncCS2 due to low protein yields after purification, kinetic parameters of TcAncCS2 and its descendant enzymes predict that specialists would be favored as 3X would not compete for active site binding; the same is perhaps true for Paullinia enzymes.

Divergent Mutations to Homologous Protein Regions Facilitated Convergent Shifts toward Substrate Specialization in Paullinia and Theobroma

In Paullinia and Theobroma, alignments show that Region I (fig. 3A) was mutated in both CS enzyme lineages and is predicted to interact with substrate molecules within the active site, as defined by the Coffea XMT and Clarkia SAMT crystal structures as well as mutagenesis of Citrus XMT (Zubieta et al. 2003; McCarthy AA and McCarthy JG 2007; Huang et al. 2016). Within Region I of the Paullinia CS lineage, Thr25 of PcAncCS2 (Node N) was replaced by Ser in modern-day PcCS2 (figs. 2C and 3A). Experimental replacement of Thr25 by Ser in PcAncCS2 largely recapitulated the enzymatic shift toward 3X preference to form TB as in its descendant, PcCS2, and caused loss of nearly all ancestral activity with other substrates (figs. 2C and 3B and supplementary table S1, Supplementary Material online). Within the Theobroma lineage leading from TcAncCS2 (Node P) to modern-day TcCS2, Region I was substituted such that SAG21-23 was replaced by AEA (figs. 2D and 3A). Experimental mutation of the three contiguous sites resulted in a convergent shift of relative preference for 3X to form TB, like was shown in the Paullinia lineage, as well as the ability to convert X to 7X making the mutant very similar to modern-day TcCS2 (figs. 2D and 3B, supplementary table S1, Supplementary Material online). Thus, convergent improvement to 3X methylation of the N7 position occurred by divergent mutations to different codons of a broadly homologous region of CS-type enzymes. Support for the hypothesis that Region I is necessary for xanthine alkaloid methylation specificity is further strengthened by the fact that a convergent amino acid replacement to that of PcAncCS2 occurred in the XMT lineage of Citrus CF biosynthetic enzymes (Huang et al. 2016) (fig. 3A).

In this case, instead of T25S as inferred for Paullinia PcAncCS2, Pro25 was replaced by Ser in CisAncXMT2 and this also resulted in improved activity with 3X (Huang et al. 2016). However, although P25S increased activity with 3X in CisAncXMT2, this improved upon N1 methylation unlike the case for PcAncCS2 in which N7 activity increased when Ser replaced Thr in the presumed active site. Thus, this convergent amino acid replacement by Ser in these SABATH paralogs did not result in the evolution of convergent catalytic properties.

Amino acid replacements in Region III in both the Paullinia and Theobroma ancestral CS lineages, also produced similar convergent effects on substrate preference. In Paullinia, Asn307 of PcAncCS2 was mutated to Tyr in modern-day PcCS1 that shows near-complete preference for X to form 3X (figs. 2C and 3A). When we experimentally replaced Asn307 with Tyr in PcAncCS2, all activity with 3X and 7X was lost such that specialization for xanthine methylation at N3 resulted (fig. 3B and supplementary table S1, Supplementary Material online). Therefore, this single replacement largely recapitulated the evolution of enzyme substrate preference of PcCS1 (figs. 2C and 3B). TcAncCS2 descended from TcAncCS1 and switched substrate preference from 7X to X. TcAncCS2 differs from its ancestor by three amino acids in Region III (fig. 3A). When we experimentally replaced Asn307, Leu308, and Ser310 of TcAncCS1 by Gly307, His308, and Cys310 (NLRS307-310GHRC), the mutant showed higher activity with X to produce 3X, much like TcAncCS2, (although it also formed 1X) (fig. 3B and supplementary table S1, Supplementary Material online). Thus, convergence toward N3 methylation of X appears to have evolved in part by different substitutions to the same homologous protein Region III.

Collectively, our results show that the ancient pathways to methylate xanthine alkaloids in Paullinia and Theobroma ancestors likely differed from the similar ones used by species today. Because the pathways later converged it suggests that selection may be important for pathway flux changes over time. After subsequent gene duplication and divergence, single ancient enzymes alone could perform the pathway steps that the multiple modern-day descendants currently catalyze. That convergence subsequently resulted from independent gene duplication events that led to the partitioning of the same biochemical reactions in two enzymes in each of the two lineages studied is remarkable. These results suggest that two modern-day enzymes are better than one ancestral enzyme in terms of pathway flux and product accumulation which is largely consistent with predictions under a model of
escape from adaptive conflict (Des Marais and Rausher 2008) although a rigorous test of this model was beyond the scope of this study (Barkman and Zhang 2009). Similar to the CS enzymes we have studied, ancestral constraints for multistep enzymes in the corticosteroid pathway were shown to be at least partly alleviated after gene duplication in primates (Olson-Manning 2020). Although it appears that the modern-day duplicated CS enzymes are coexpressed in *Paullinia* and *Theobroma* (Huang et al. 2016), knowledge of ancestral tissue-specific expression patterns for these enzymes would further understanding of the mechanisms of convergence; however, such data are difficult to infer with precision.

**Fig. 3.** Ancestral CS enzymes experienced mutations to homologous regions and exhibited convergent changes in substrate preferences. (A) Alignments in Regions I and III of CS enzymes show that both were mutated in ancestral *Theobroma* and *Paullinia* enzymes, whereas only ancestral *Citrus* XMT (Huang et al. 2016) experienced substitution in Region II. Ancestral/derived amino acid states are shown in blue/red, respectively. (B) Correspondence analysis shows that ancient CS enzymes were similar and associate due to 7X methylation preference (node labels and substrate colors in pie charts taken from fig. 2C–E). From these ancestral activities, convergent modern-day enzyme substrate preferences evolved largely by mutations to common protein regions. It appears that mutations to Region I in PcAncCS2 and TcAncCS2 resulted in the convergent evolution of similar enzymes, PcCS2 and TcCS2, which associate due to preference to methylate 3X to form TB. Mutations to Region III of PcAncCS2 and TcAncCS1 resulted in increased relative preference for X to form 3X and ultimately contributed to the convergent evolution of PcCS1 and TcCS1. Arrows between enzyme coordinates represent the enzyme lineages shown in figure 2.
due to our lack of knowledge of ancestral transcriptional regulators and cis-regulatory elements. Finally, because we show that different mutations to homologous sequence regions led to convergence, constraint on the mutational paths available to each lineage is implicated. Future work aimed at testing the relative roles for selection and constraint for this and other cases of convergence may benefit from the use of ancestral sequence resurrection that can reveal evolutionary aspects of the process that may not have been predicted a priori.

Materials and Methods

Phylogenetic Analyses
In order to accurately determine the orthology of Paullinia (Sapindales), Theobroma (Malvales), and Camellia (Ericales) xanthine alkaloid-producing enzymes, amino acid sequences from all previously characterized SABATH gene family members and those from various land plant complete genomes were obtained from GenBank and the PlantTribes database (Wall et al. 2008). We also queried the OneKP database in order to provide more detailed branching relationships of the recently evolved CS enzymes of Paullinia, Theobroma, and Camellia as shown in supplementary figure S1, Supplementary Material online. Alignment of amino acid sequences was achieved using MAFFT version 7 (Katoh and Standley 2013) using the auto search strategy to maximize accuracy and speed. A maximum likelihood phylogenetic estimate for the SABATH family members was obtained using PhyML (Guindon et al. 2010) assuming the Jones, Taylor, and Thornton (JTT) matrix model for amino acid substitution with an invariant and gamma parameter for among-site rate heterogeneity as determined by ProtTest (Abascal et al. 2005). Bootstrap support was evaluated based on 100 pseudoreplicated data sets.

ASR enzyme sequences for nodes M–Q were estimated from the full CS lineage of the SABATH gene family shown in figure 2A using the JTT+ Gamma model of amino acid substitution as implemented in Codeml of PAML 4.0 (Yang 2007). Importantly, where possible, we relied on a combination of genomic sequence as well as supporting transcriptomic data to ensure that high-quality sequence was analyzed in order to avoid introducing potential sequence artifacts into our ancestral state estimates. This was possible for Theobroma and Camellia (Argout et al. 2008, 2011; Taniguchi et al. 2012; Wei et al. 2018); however, no genome exists for Paullinia so in that case we relied on transcriptome data alone (Angelo et al. 2008; Figueiredo et al. 2011) as reported in Huang et al. (2016). Alignments of the ancestral proteins with their modern-day descendants are shown in supplementary figure S8, Supplementary Material online. In order to determine ancestral protein lengths in regions with alignment gaps, we coded each sequence for the number of amino acids possessed and used parsimony to determine ancestral residue numbers as in our previous studies (Huang et al. 2012, 2016). The estimated sequences were synthesized by Genscript Corp. and had codons chosen for optimal protein expression in Escherichia coli. For sites that had relatively low posterior probabilities or that differed when the gamma parameter was not assumed, we generated alternative ancestral alleles by site-directed mutagenesis using the QuickChange® Site-Directed Mutagenesis Kit (Agilent), following the manufacturer’s protocol. This allowed us to assess whether experimentally determined enzyme activities were dependent upon particular amino acid reconstructions. At least two ancestral enzymes were characterized for each node M–Q in figure 2 even though average posterior probabilities were high for most sites (see average site-specific posterior probabilities in supplementary figs. S3–S7, Supplementary Material online). Details of those alternative alleles are provided in supplementary figures S3–S7, Supplementary Material online, including which sites were mutated as well as individual enzyme activity of each allele. Mean relative activity with each substrate is shown in the pie charts of figure. 2C–E.

Cloning, Heterologous Expression, and Purification of Enzymes
Ancestral gene sequences were synthesized by Genscript and were subcloned from the pUC57 cloning vector into the pET-15b (Novagen) expression vector. Plasmid DNA was first digested at 37 °C for 6 h using 1.5 μg of DNA and Ndel and BamHI in 30 μl reactions. Linear fragments corresponding to the expected sizes were gel purified using the QIAEXII gel extraction kit (Qiagen Corp.) according to the manufacturer’s instructions. Purified DNA fragments were ligated into pET15b using T4 DNA ligase from New England Biolabs. Reactions were incubated at 16 °C overnight. Ligation products were transformed into Top10 E. coli cells using 2 μl of ligation reaction. Minipreps of positive transformants were obtained using a Qiaprep spin miniprep kit (Qiagen Corp.). Ten nanogram of each plasmid was used to transform and grow BL21 E. coli cells using standard plating and incubation methods. Induction of His6-protein was achieved in 50 ml BL21 (DE3) cell cultures with the addition of 1 mM IPTG at 23 °C for 6 h as described previously (Huang et al. 2016). Purification of the His6c-tagged protein utilized TALON spin columns (Takara Bio) and followed the manufacturer’s instructions. Bradford assays were used to determine purified protein concentration and recombinant protein purity was evaluated on SDS-PAGE gels. The plasmids used to produce ancestral proteins are freely available upon request.

Enzyme Assays
All enzymes were tested for activity with the eight xanthine alkaloid substrates shown in figure 1. Radiochemical assays were performed in 50 μl reactions with 0.01 μCi (0.5 μl) 14C-labeled SAM, 100 μM methyl acceptor substrate dissolved in 0.5 M NaOH and 10–20 μl purified protein in 50 mM Tris–HCl buffer at 24 °C for 60 min. Negative controls were composed of the same reagents except that the methyl acceptor substrate was omitted and 1 μl of 0.5 M NaOH was added instead. Methylated products were extracted in 200 μl ethyl acetate and quantified using a liquid scintillation counter. The
highest enzyme activity reached with a specific substrate was set to 1.0 and relative activities with remaining substrates were calculated. Each assay was run at least twice so that mean, plus SD, could be calculated as shown in supplementary figures S3–S7, Supplementary Material online. Although the enzymes in this study did not show high activity with some of the substrates (e.g., 1X, XR), we do not believe that this is due to any artifact or limitation of the assays since our previous studies using the same conditions did detect catalysis of those structures by different enzymes (Huang et al. 2016).

Enzyme Kinetics
Kinetic parameters ($k_{\text{cat}}$ and $K_M$) of the methyltransferases with a given substrate were determined using the 50 μl radioassay described above. However, appropriate enzyme concentration and incubation time were determined in time-course assays with low nonsaturating substrate concentrations to ensure that the reaction velocity was linear during the assay period. When varying xanthine alkaloid substrate concentration, the SAM concentration was held constant and saturated at 320 μM. Assays were run in duplicate and initial velocities versus substrate concentration were plotted using GraphPad Prism (GraphPad Software, La Jolla, CA) to fit the hyperbolic Michaelis–Menten equation to calculate $V_{\text{max}}$ and $K_M$. $V_{\text{max}}$ was converted to $k_{\text{cat}}$ based on estimated protein concentrations and expressed in units of s$^{-1}$.

Liquid Chromatography–Tandem Mass Spectrometry
Liquid chromatography MS–tandem mass spectrometry (LC-MS/MS) was used to confirm product identity from ancestral enzymes and substrate preferences was determined with a given substrate were determined using the 50 μl radioassay described above. However, appropriate enzyme concentration and incubation time were determined in time-course assays with low nonsaturating substrate concentrations to ensure that the reaction velocity was linear during the assay period. When varying xanthine alkaloid substrate concentration, the SAM concentration was held constant and saturated at 320 μM. Assays were run in duplicate and initial velocities versus substrate concentration were plotted using GraphPad Prism (GraphPad Software, La Jolla, CA) to fit the hyperbolic Michaelis–Menten equation to calculate $V_{\text{max}}$ and $K_M$. $V_{\text{max}}$ was converted to $k_{\text{cat}}$ based on estimated protein concentrations and expressed in units of s$^{-1}$.

Statistical Analysis
Correspondence analysis (Jackson 1997) was used to ordinate modern-day, ancestral, and mutated ancestral CS enzymes based upon relativized substrate preferences. Symmetric plots were used to visualize the results. Nonindependence of the enzymes and substrate preferences was determined ($P < 0.05$) and total inertia was 1.05 for the analysis described in figure 3. The first two factors of the analysis accounted for a total of 85% of the inertia. Positions of enzymes along the x-axis in the symmetric plot are due to variation in methylation preference for xanthine (coordinates < −0.5) or 3X (coordinates > 1). On the other hand, preference for 7X methylation is explained by position along the y-axis (coordinates < −0.5).

Supplementary Material
Supplementary data are available at Molecular Biology and Evolution online.

Acknowledgments
This work was supported by the National Science Foundation (Grant No. MCB-1120624 to T.J.B.). Greg Cavey is thanked for assistance provided with LC-MS analyses. Ricky Stull and two anonymous reviewers provided helpful feedback on previous versions of the manuscript.

Data Availability
Individuals interested in the data matrices underlying this article may obtain them upon request of the corresponding author. The original data underlying this article are available in https://www.ncbi.nlm.nih.gov/genbank/ as well as https://db.cngb.org/onekp/ (both of which were last accessed on March 01, 2021).

References
Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21(9):2104–2105.
Anaya AL, Cruz-Ortega R, Waller GR. 2006. Metabolism and ecology of purine alkaloids. Front Biosci. 11:2354–2370.
Angelo PCSA, Nunes-Silva CG, Brigido MM, Azevedo JSN, Assuncao EN, Sousa ARB, Patricio FJB, Rego MM, Peixoto JCC, Oliveira WP, et al. 2008. Guarana (Paullinia cupana var. sorbilis), an anciently consumed stimulant from the Amazon rain forest: the seeded-fruit transcriptome. Plant Cell Rep. 27(11):117–124.
Argout X, Fouet O, Wincker P, Gramacho K, Legavre T, Sabau X, Risterucci AM, Da Silva C, Cascarino J, Allegre M, et al. 2008. Towards the understanding of the cocoa transcriptome: production and analysis of an exhaustive dataset of ESTs of Theobroma cacao L. generated from various tissues and under various conditions. BMC Genomics 9:512.
Argout X, Salse J, Aury J-M, Gaultier MN, Drog C, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova SN, et al. 2011. The genome of Theobroma cacao. Nat Genet. 43(2):101–108.
Ashihara H, Crozier A. 2001. Caffeine: a well known but little mentioned stimulant from the Amazon rain forest: the seeded-fruit transcriptome. Plant Cell Rep. 20:912–918.
Ashihara H, Crozier A. 1996. Biochemical synthesis of caffeine in leaves of coffee. Plant Physiol. 111(3):747–753.
Ashihara H, Sano H, Crozier A. 2008. Caffeine and related purine alkaloids: biosynthesis, catabolism, function and genetic engineering. Phytochemistry 69(4):841–856.
Ashihara H, Suzuki T. 2004. Distribution and biosynthesis of caffeine in plants. Front Biosci. 9:1864–1876.
Barkman T, Zhang J. 2009. Evidence for escape from adaptive conflict? Nature 462(7274):E1–E2.
Baumann TW, Schulthess BH, Hanni K. 1995. Guarana (Paullinia cupana) rewards seed dispersers without intoxicating them by caffeine. Phytochemistry 39(5):1063–1070.
Bridgham JT, Carroll SM, Thornton JW. 2006. Evolution of hormone-receptor complexity by molecular exploitation. Science 312(5770):97–101.
Convergent Biochemical Pathways for Xanthine Alkaloid Production in Plants.

Bridgham JT, Ortlund EA, Thornton JW. 2009. An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. Nature 461(7263):515–519.

Chang BSW, Jonsson K, Kazmi MA, Donoghue MJ, Sakmar TP. 2002. Recreating a functional ancient archosaur visual pigment. Mol Biol Evol. 19(9):1483–1489.

Christin P-A, Weinreich DM, Besnard G. 2010. Causes and evolutionary significance of genetic convergence. Trends Genet. 26(9):400–406.

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(1):69–80.

Dean AM, Thornton JW. 2007. Mechanistic approaches to the study of evolution: the functional synthesis. Nat Rev Genet. 8(9):675–688.

Deng C, Xiuping K, Lin-Lin C, Si-pn P, Limao F, Wei-Wei D, Zhao J, Zheng-Zhu Z. 2020. Metabolite and transcriptome profiling on xanthine alkaloids-fed tea plant (Camellia sinensis) shoot tips and roots reveal the complex metabolic network for caffeine biosynthesis and degradation. Front Plant Sci. 11:551288.

Denoëud F, Carretero-Paulet L, Dereeper A, Droc G, Guyot R, Pietrella M, Des Marais DL, Rausher MD. 2010. Parallel evolution at multiple levels in the central nervous-system: interaction with caffeine and related methylxanthines. Cell Mol Neurobiol. 30(3):59–68.

Deuel FJ, Lefort V, Goyrat R, Aprea G, Zheng CF, Alberti A, Anthony F, Aprea G. 2014. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. Science 345(6191):1181–1184.

Des Marais DL, Rausher MD. 2008. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. Nature 454(7205):U762–U785.

Des Marais DL, Rausher MD. 2010. Parallel evolution at multiple levels in the origin of hummingbird pollinated flowers in Siphonochilus. Evolution 64(7):2044–2054.

Field SF, Matz MV. 2010. Retracing evolution of red fluorescence in GFP-like proteins from Favia corals. Mol Biol Evol. 27(2):225–233.

Figueiredo LC, Faria-Campos AC, Astolfi S, Azevedo JL. 2011. Identification and isolation of full-length cDNA sequences by sequencing and analysis of expressed sequence tags from guarana (Paullinia cupana). Genet Mol Res. 10(2):1188–1199.

Gaucher EA, Govindarajan S, Ganesh OK. 2008. Palaeotemperature

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 U S A. 113(38):10613–10618.

Ishida M, Kita N, Mizuno K, Tanikawa N, Kato M. 2009. Occurrence of theobromine synthase genes in purine alkaloid-free species of Camellia plants. Plant Sci. 229(3):559–568.

Jackson DA. 1997. Compositional data in community ecology: the paradigm or peril of proportions. Ecology 78(3):929–940.

Kato M, Kanehara T, Shimizu H, Suzuki T, Gilles FM, Crozier A, Ashihara H. 1996. Caffeine biosynthesis in young leaves of Camellia sinensis: in vitro studies on N-methyltransferase activity involved in the conversion of xanthosine to caffeine. Physiol Plant. 98(3):629–636.

Kato M, Mizuno K. 2004. Caffeine synthase and related methyltransferases in plants. Front Biosci. 9(1–3):1833–1842.

Kato M, Mizuno K, Crozier A, Fujimura T, Ashihara H. 2000. Plant biotechnology—caffeine synthase gene from tea leaves. Nature 406(6800):956–957.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.

Lichman BR, Godden GT, Hamilton JP, Palmer L, Kamileen MO, Zhao D, Vaillancourt B, Wood JC, Sun M, Kinser TJ, et al. 2020. The evolutionary origins of the cat attractant nepetalactone in catnip. Sci Adv. 6(20):eaab0721.

Losos JB. 2011. Convergence, adaptation, and constraint. Evolution 65(7):1827–1840.

Manceau M, Domingues VS, Linnen CR, Rosenblum EB, Hoekstra HE. 2010. Convergence in pigmentation at multiple levels: mutations, genes and function. Philos Trans R Soc Lond B Biol Sci. 365(1552):2439–2450.

Mazzafera P, Wingle G, Olsson O, Sandberg G. 1994. S-Adenosyl-L-methionine-theobromine 1-N-methyltransferase, an enzyme catalyzing the synthesis of caffeine in coffee. Phytochemistry 37(6):1577–1584.

McCarthy AA, McCarthy JG. 2007. The structure of two N-methyltransferases from the caffeine biosynthetic pathway. Plant Physiol. 144(2):879–889.

Mizuno K, Kato M, Irino F, Yoneyama N, Fujimura T, Ashihara H. 2003. The first committed step reaction of caffeine biosynthesis: 7-methylxanthosine synthase is closely homologous to caffeine synthases in coffee (Coffea arabica L.). Fests Lett. 547(1–3):56–60.

Mizuno K, Matsuzaki M, Kanazawa S, Tokiwano T, Yoshizawa Y, Kato M. 2014. Conversion of nicotinic acid to trigonelline is catalyzed by N-methyltransferase belonging to motif B‘ methyltransferase family in Coffea arabica. Biochem Biophys Res Commun. 452(4):1060–1066.

Mizuno K, Okuda A, Kato M, Yoneyama N, Tanaka H, Ashihara H, Fujimura T. 2003. Isolation of a new dual-functional caffeine synthase gene encoding an enzyme for the conversion of 7-methylxanthine to caffeine from coffee (Coffea arabica L.). Fests Lett. 531(1–3):75–81.

Nakajima C, Hoffman FG, Weber RE, Fago A, Witt CC, Storz JF. 2016. Predictive convergence in hemoglobin function has unpredictable molecular underpinnings. Science 354(6310):336–339.

Nathanson JA. 1984. Caffeine and related methylxanthines: possible naturally-occurring pesticides. Science 226(4671):184–187.

Olson-Manning CF. 2020. Elaboration of the corticosterone synthesis pathway in primates through a multistep enzyme. Mol Biol Evol. 37(8):2257–2267.

Pichersky E, Lewinsohn E. 2011. Convergent evolution in plant specialized metabolism. Annu Rev Plant Biol. 62:549–566.

Schimpl FC, Kyota E, Mayer JLS, Goncalves JFD, da Silva JF, Mazzafera P. 2014. Molecular and biochemical characterization of caffeine synthase and purine alkaloid concentration in guanara fruit. Phytochemistry 105:25–36.

Smith SD, Wang S, Rausher MD. 2013. Functional evolution of an anthocyananin pathway enzyme during a flower color transition. Mol Biol Evol. 30(3):602–612.

Storz JF. 2016. Causes of molecular convergence and parallelism in protein evolution. Nat Rev Genet. 17(4):239–250.

Suzuki T, Takahashi E. 1976. Caffeine biosynthesis in Camellia sinensis. Phytochemistry 15(8):1235–1239.

Suzuki T, Waller GR. 1987. Allelopathy due to purine alkaloids in tea seeds during germination. Plant Soil. 98(1):131–136.

Taniguchi F, Fukuoka H, Tanaka J. 2012. Expressed sequence tags from organ-specific cDNA libraries of tea (Camellia sinensis) and polymorphisms and transferability of EST-SSRs across Camellia species. Breed Sci. 62(2):186–195.

Thornton JW. 2004. Resurrecting ancient genes: experimental analysis of extinct molecules. Nat Rev Genet. 5(5):366–375.

Uefuji H, Ogita S, Yamaguchi Y, Koizumi N, Sano H. 2003. Molecular cloning and functional characterization of three distinct N-methyltransferases involved in the caffeine biosynthetic pathway in coffee plants. Plant Physiol. 132(1):372–380.

Uefuji H, Tatsuyama Y, Morimoto M, Kaithien-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(2):221–227.

Wall PK, Leebens-Mack J, Muller KF, Field D, Altman NS, dePamphilis CW. 2008. Plant Tribes: a gene and gene family resource for comparative genomics in plants. Nucleic Acids Res. 36(Database issue):D970–D976.

Wei C, Yang H, Wang S, Zhao J, Liu C, Gao L, Xia E, Lu Y, Tai Y, She G, et al. 2018. Draft genome sequence of Camellia sinensis var. sinensis...
provides insights into the evolution of the tea genome and tea quality. Proc Natl Acad Sci U S A. 115(18):E4151–E4158.
Wheeler LC, Smith SD. 2019. Computational modeling of anthocyanin pathway evolution: biases, hotspots, and trade-offs. Integr Comp Biol. 59(3):585–598.
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(6124):1202–1204.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24(8):1586–1591.
Yoneyama N, Morimoto H, Chuang-Xing Y, Ashihara H, Mizuno K, Kato M. 2006. Substrate specificity of N-methyltransferase involved in purine alkaloids synthesis is dependent upon one amino acid residue of the enzyme. Mol Genet Genomics. 275(2):1338–1354.
Zeng L, Zhang Q, Endress PK, Huang J, Ma H. 2017. Resolution of deep eudicot phylogeny and their temporal diversification using nuclear genes from transcriptomic and genomic datasets. New Phytol. 214(3):1338–1354.
Zhang JZ. 2006. Parallel adaptive origins of digestive RNases in Asian and African leaf monkeys. Nat Genet. 38(7):819–823.
Zubieta C, Ross JR, Koscheski P, Yang Y, Pichersky E, Noel JP. 2003. Structural basis for substrate recognition in the salicylic acid carboxyl methyltransferase family. Plant Cell. 15(8):1704–1716.