# Convergent evolution of caffeine in plants by co-option of exapted ancestral enzymes

Ruiqi Huang[a], Andrew J. O'Donnell[a,1], Jessica J. Barboline[a], and Todd J. Barkman[a,2]

[a]Department of Biological Sciences, Western Michigan University, Kalamazoo, MI 49008

Edited by Ian T. Baldwin, Max Planck Institute for Chemical Ecology, Jena, Germany, and approved July 18, 2016 (received for review March 25, 2016)

**Convergent evolution is a process that has occurred throughout the tree of life, but the historical genetic and biochemical context promoting the repeated independent origins of a trait is rarely understood. The well-known stimulant caffeine, and its xanthine alkaloid precursors, has evolved multiple times in flowering plant history for various roles in plant defense and pollination. We have shown that convergent caffeine production, surprisingly, has evolved by two previously unknown biochemical pathways in chocolate, citrus, and guaraná plants using either caffeine synthase- or xanthine methyltransferase-like enzymes. However, the pathway and enzyme lineage used by any given plant species is not predictable from phylogenetic relatedness alone. Ancestral sequence resurrection reveals that this convergence was facilitated by co-option of genes maintained over 100 million y for alternative biochemical roles. The ancient enzymes of the *Citrus* lineage were exapted for reactions currently used for various steps of caffeine biosynthesis and required very few mutations to acquire modern-day enzymatic characteristics, allowing for the evolution of a complete pathway. Future studies aimed at manipulating caffeine content of plants will require the use of different approaches given the metabolic and genetic diversity revealed by this study.**

convergent evolution | caffeine biosynthesis | enzyme evolution | paleomolecular biology

Convergent evolution has resulted in the independent origins of many traits dispersed throughout the tree of life. Whereas some convergent traits are known to be generated via similar developmental or biochemical pathways, others arise from different paths (1–5). Likewise, similar (orthologous) or different (paralogous or even unrelated) genes may encode for the regulatory or structural proteins composing the components of pathways that build convergent traits (6–10). One of the most prominent examples of convergence in plants is that of caffeine biosynthesis, which appears to have evolved at least five times during flowering plant history (11). The phylogenetic distribution of caffeine, or xanthine alkaloids more generally, is highly sporadic and usually restricted to only a few species within a given genus (12, 13). Caffeine accumulates in various tissues, where it may deter herbivory (14, 15) or enhance pollinator memory (16). Numerous studies over the past 30 y have indicated that although several possible routes exist, the same canonical pathway to caffeine biosynthesis has evolved independently in *Coffea* (coffee) and *Camellia* (tea) involving three methylation reactions to sequentially convert xanthosine to 7-methylxanthine to theobromine to caffeine (Fig. 1) (17, 18). In *Coffea*, three xanthine methyltransferase (XMT)-type enzymes from the SABATH (salicylic acid, benzoic acid, theobromine methyltransferase) family (19) are used to catalyze the methylation steps of the pathway, whereas *Camellia* uses a paralogous, convergently evolved caffeine synthase (CS)-type enzyme (20–22) (Fig. 1). Because most SABATH enzymes catalyze the methylation of oxygen atoms of a wide diversity of carboxylic acids such as anthranilic, benzoic, gibberellic, jasmonic, loganic, salicylic, and indole-3-acetic acid for floral scent, defense, and hormone modulation (23–25), methylation of xanthine alkaloid nitrogen atoms by XMT and CS is likely a recently evolved activity.

Although convergence has been documented at multiple hierarchical levels, fundamental questions remain unanswered about the evolutionary gain of traits such as caffeine that are formed via a multistep pathway. First, although convergently co-opted genes, such as *XMT* or *CS*, may evolve to encode enzymes for the same biosynthetic pathway, it is unknown what ancestral functions they historically provided that allowed for their maintenance over millions of years of divergence. Second, it is unknown how multiple protein components are evolutionarily assembled into an ordered, functional pathway like that for caffeine biosynthesis. Under the cumulative hypothesis (26), it is predicted that enzymes catalyzing earlier reactions of a pathway must evolve first; otherwise, enzymes that perform later reactions would have no substrates with which to react. Subsequently, duplication of the gene encoding the first enzyme would give rise to enzymes catalyzing later steps. This hypothesis assumes that, initially, the intermediates in a pathway are advantageous, because it is unlikely that multiple enzymatic steps in a pathway could evolve simultaneously. Alternatively, the retrograde hypothesis (27) states that enzymes catalyzing reactions that occur at the end of a pathway evolved first. Gene duplication of the sequence encoding the first-evolved enzyme would eventually result in new enzymes that perform the preceding pathway steps. This hypothesis assumes that the intermediates of a given pathway would be produced nonenzymatically and be available for catalysis; as such, it may have less general explanatory application. Finally, the patchwork hypothesis (28, 29) explains the origins of novel pathways by the recruitment of enzymes from alternative preexisting pathways. This hypothesis assumes that the older, recruited enzymes were ancestrally promiscuous with respect to the substrates catalyzed such that they were exapted for the activities that they later become specialized for in the novel pathway. Unlike the cumulative and retrograde hypotheses, there is no prediction for the relative ages of enzymes performing each step of the novel pathway under the patchwork hypothesis. The patchwork hypothesis is compatible with the innovation, amplification, and duplication

## Significance

Convergent evolution is responsible for generating similar traits in unrelated organisms, such as wings that allow flight in birds and bats. In plants, one of the most prominent examples of convergence is that of caffeine production, which has independently evolved in numerous species. In this study, we reveal that even though the caffeine molecule is identical in the cacao, citrus, guaraná, coffee, and tea lineages, it is produced by different, previously unknown, biosynthetic pathways. Furthermore, by resurrecting extinct enzymes that ancient plants once possessed, we show that the novel pathways would have evolved rapidly because the ancestral enzymes were co-opted from previous biochemical roles to those of caffeine biosynthesis for which they were already primed.

Author contributions: R.H., A.J.O., J.J.B., and T.J.B. designed research; R.H., A.J.O., and J.J.B. performed research; R.H., A.J.O., J.J.B., and T.J.B. analyzed data; and T.J.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]Present address: Max Planck Institute for Chemical Ecology, 07745 Jena, Germany.

[2]To whom correspondence should be addressed. Email: todd.barkman@wmich.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1602575113/-/DCSupplemental.

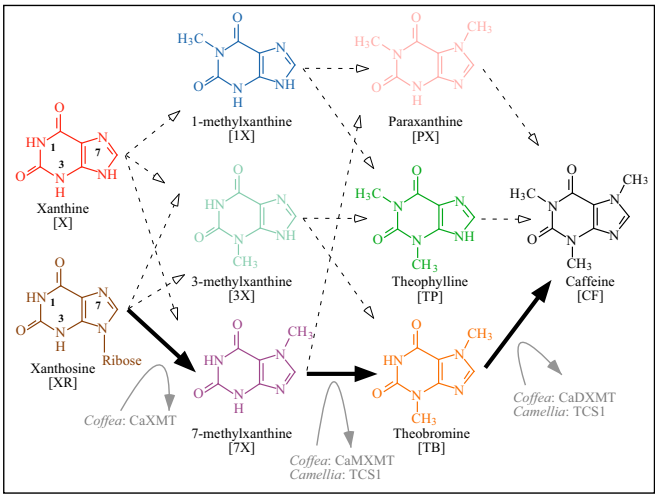


**Fig. 1.** Caffeine biosynthetic network has 12 potential paths. The only path characterized from plants is shown by solid black arrows and involves sequential methylation of xanthosine at N-7, 7-methylxanthine at N-3, and theobromine at N-1 of the heterocyclic ring. Each methylation step is performed by a separate xanthine alkaloid methyltransferase in *Coffea*. In contrast, *Camellia* employs the distantly related caffeine synthase enzyme, TCS1, for both the second and third methylation steps, whereas the enzyme that catalyzes the first reaction remains uncharacterized. Other potential biochemical pathways to caffeine are shown by dashed arrows, but enzymes specialized for those conversions are unknown. Cleavage of ribose from 7-methylxanthosine is not shown, but may occur concomitantly with N-7 methylation of xanthosine. CF, caffeine; PX, paraxanthine; TB, theobromine; TP, theophylline; X, xanthine; 1X, 1-methylxanthine; 3X, 3-methylxanthine; 7X, 7-methylxanthine; XR, xanthosine.

model of protein functional change (30) and those that have emerged from protein engineering studies (31) in that promiscuous enzyme activities are nearly universal properties of modern-day enzymes and have been shown to serve as the basis for evolution of specialized, novel enzyme activities.

Here we report on a comparative molecular and biochemical approach that dissects how caffeine convergence occurred at the level of the genes involved and biochemical pathways catalyzed in the five economically important plants *Theobroma* (chocolate), *Paullinia* (guaraná), *Citrus* (orange), *Camellia*, and *Coffea*. We further use the *Citrus* lineage to test hypotheses related to the mechanisms allowing for the convergent evolution of caffeine by using paleomolecular biology coupled with experimental mutagenesis, thereby demonstrating how multistep pathways may independently evolve.

## Results and Discussion

**Novel Biosynthetic Pathways for Caffeine in Modern-Day Plants.** To uncover the genes and pathways used by modern-day plants to synthesize caffeine, bioinformatic and phylogenetic analyses were used to reveal that both *Theobroma cacao* (Tc) (Malvales) and *Paullinia cupana* (Pc) (Sapindales) express multiple *CS*-type sequences in their caffeinated leaves and/or fruits that are orthologous to those used by *Camellia sinensis* (Ericales) in leaves and shoots (Fig. 2*A* and Figs. S1 and S2) (32, 33). However, contrary to expectations from *Camellia* (Fig. 1), heterologous expression and assays of the *Theobroma* and *Paullinia* CS enzymes indicate that they catalyze a different pathway to synthesize xanthine alkaloids. Specifically, both species possess one enzyme (CS1) that preferentially methylates xanthine to produce 3-methylxanthine, as well as a second enzyme (CS2) that preferentially methylates 3-methylxanthine to produce theobromine (Fig. 2 *B* and *C*). Surprisingly, even though the four enzymes are part of the CS lineage, TcCS1 is more closely related to TcCS2 rather than the enzymatically similar PcCS1, which, in turn, is more closely related to PcCS2 (Fig. 2 *B* and *C* and Fig. S1). This pattern of relationships indicates that the relative methylation preferences of these enzymes, although similar, have convergently evolved in *Theobroma* and *Paullinia*, likely after gene duplication independently occurred in each lineage (Fig. 2 *B* and *C* and Fig. S1). No enzymes have been previously reported to specialize in the methylation of xanthine or 3-methylxanthine, and the biochemical route to caffeine implied by these enzyme activities (Fig. 2 *B* and *C*) has not been implicated as the primary pathway in these or any other plants. However, there is evidence for this pathway in *Theobroma* fruits and leaves from metabolomic analyses and radiolabeled tracer studies that showed 3-methylxanthine as an intermediate formed during theobromine accumulation when xanthine or various purine bases and nucleosides are provided as substrates (34, 35). Analyses of fruit and leaf extracts and isotope tracer studies also report the accumulation of 7-methylxanthine to a lesser extent (34, 35). Our liquid chromatography-mass spectrometry (LC-MS) results indicate that this metabolite can be formed by methylation of xanthine by TcCS2 as a secondary activity (Fig. 2*B* and Fig. S3). Thus, in *Theobroma*, it is possible that theobromine is produced via methylation of this intermediate by BTS (36) and/or TcCS1 (Fig. 2*B*), in addition to 3-methylxanthine by TcCS2 (Fig. 2*B*). It is not yet clear which enzyme contributes to the low levels of caffeine accumulation in *Theobroma*, but this is not surprising, because its biosynthesis is reported to be very slow (35). In *Paullinia*, a third enzyme, PcCS, is reported to convert theobromine to caffeine (37) (Fig. 2*C*). Theobromine is reported from *Paullinia* tissues (37, 38), which is consistent with the pathway shown in Fig. 2*C*, but no analyses have surveyed for the presence of intermediates such as 3-methylxanthine or others. For *Camellia*, bioinformatic and phylogenetic analyses show that *TCS1* and *TCS2* are expressed in leaves and recently duplicated (Fig. 2*E* and Figs. S1 and S2*C*) (36). Although the biochemical role of TCS1 is clear (39), an associated activity for TCS2 has remained elusive (36); however, we were able to demonstrate maximal methyl transfer activity at N-7 of xanthosine (Fig. 2*E*), consistent with the reported pathway (Fig. 1).

In contrast, bioinformatic and phylogenetic analyses revealed that *Citrus sinensis* (Cis) (Sapindales) expresses two recently duplicated *XMT*-type enzymes in caffeinated flowers orthologous to those found in *Coffea arabica* (Gentianales) tissues (Fig. 2*A* and Figs. S1 and S4*A*). Surprisingly, assays of the *Citrus* XMT enzymes imply yet another pathway, different from that catalyzed by *Coffea* XMT enzymes (Fig. 1), which has led to the convergent evolution of caffeine. Specifically, CisXMT1 not only methylates xanthine to produce 1- and 3-methylxanthine; it also methylates both 1- and 3-methylxanthine to produce theophylline (Fig. 2*D* and Fig. S5). A second enzyme, CisXMT2, preferentially methylates theophylline to produce caffeine (Fig. 2*D*). LC-MS analyses of flower buds (Fig. S6) and assays of crude enzyme extracts from *Citrus x limon* stamens (40) are consistent with the pathway shown in Fig. 2*D*. Furthermore, in *Citrus*, theophylline is a conspicuous metabolite in developing flower buds that accumulates early and decreases in concentration as caffeine levels increase (41), suggesting that it is involved in the accumulation of caffeine. On the other hand, theobromine, the long-assumed universal precursor to caffeine in plants (11), is undetectable or present only at low levels in developing buds (41). These findings are particularly intriguing, given that no enzymes have been previously reported to be specialized for these methylation reactions to form theophylline or caffeine and because theophylline is usually considered a degradation product of caffeine (42, 43). We expected *Citrus* to use the same gene family members and pathway as *Paullinia* because both are members of Sapindales (Fig. 2*A*). However, we could neither detect in vitro activity with xanthine alkaloid substrates by the single *Citrus* CS-type enzyme (Fig. S1) nor is it represented by ESTs in flowers (Fig. S4*A*), the principal site of caffeine accumulation (41).

For more than 30 y, published studies have indicated that caffeine is produced via a single canonical pathway in plants (44, 45). Our results show that flowering plants have a much broader

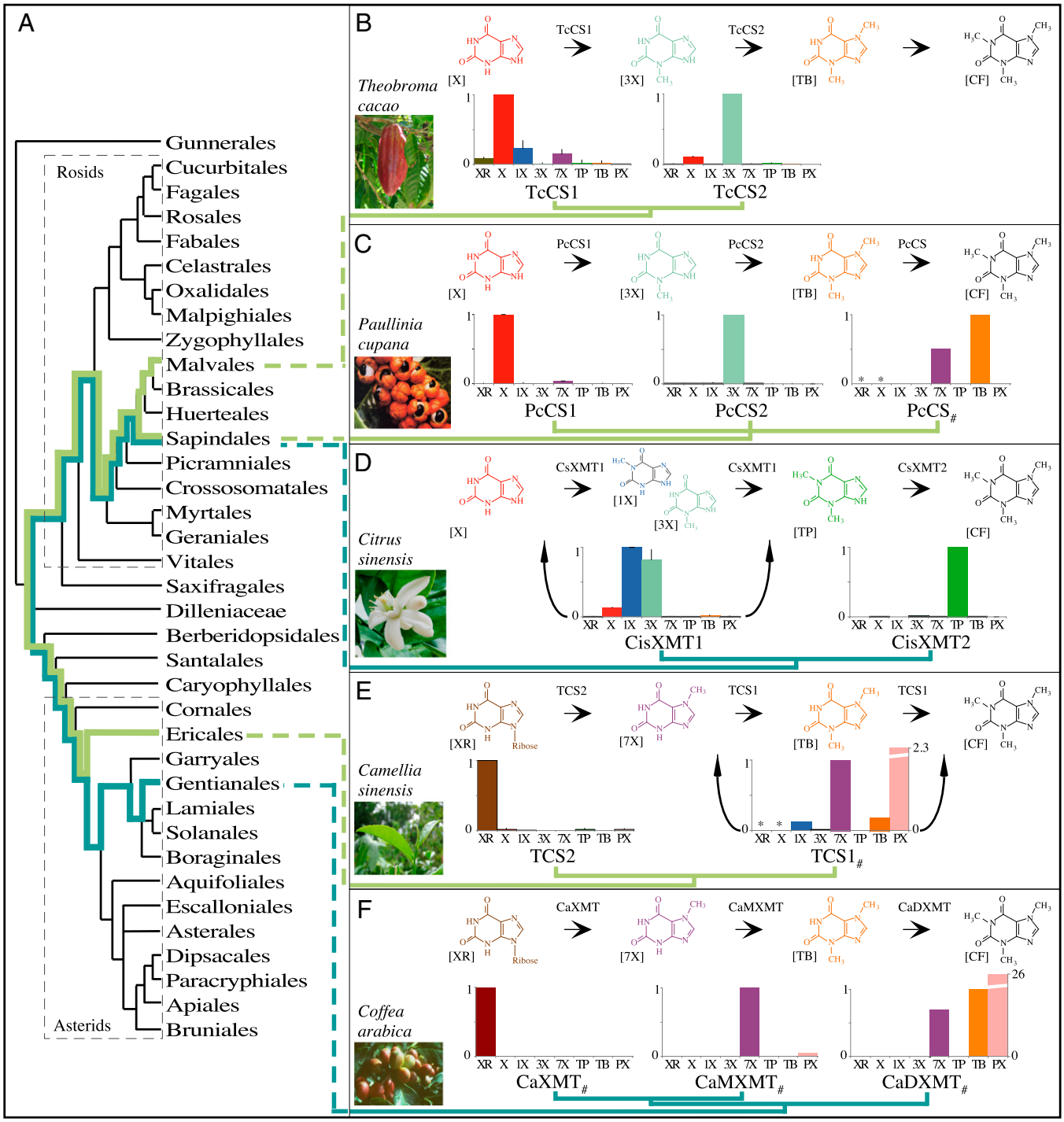


**Fig. 2.** Caffeine has convergently evolved in five flowering plant species using different combinations of genes and pathways. (*A*) Phylogenetic relationships among orders of Rosids and Asterids show multiple origins of caffeine biosynthesis. Lime-green lineages trace the ancient CS lineage of enzymes that has been independently recruited for use in caffeine-accumulating tissues in *Theobroma*, *Paullinia*, and *Camellia*. Turquoise lineages trace the ancient XMT lineage that was independently recruited in *Citrus* and *Coffea*. (*B* and *C*) *Theobroma* and *Paullinia* have converged upon similar biosynthetic pathways catalyzed by CS-type enzymes. (*D*) *Citrus* has evolved a different pathway catalyzed by XMT-type enzymes, despite its close relationship to *Paullinia*. (*E* and *F*) *Camellia* and *Coffea* catalyze the same pathway using different enzymes. Proposed biochemical pathways are based on relative enzyme activities shown by corresponding bar charts that indicate mean relative activities (from 0 to 1) with eight xanthine alkaloid substrates. CisXMT1 and TCS1 catalyze more than one reaction in the proposed pathways. XMT and CS have recently and independently duplicated in each of the five lineages (see Fig. S1 for a detailed gene tree). #Data taken from the literature; *substrate not assayed.

biochemical repertoire whereby at least three pathways lead to caffeine biosynthesis catalyzed by enzymes that derive from one of two methyltransferase lineages. These enzymes have substrate affinities (as measured by $K_M$) comparable to XMT and CS in *Coffea* and *Camellia*, respectively, as well as those of other SABATH family members, which are in the 10–1,000 μM range

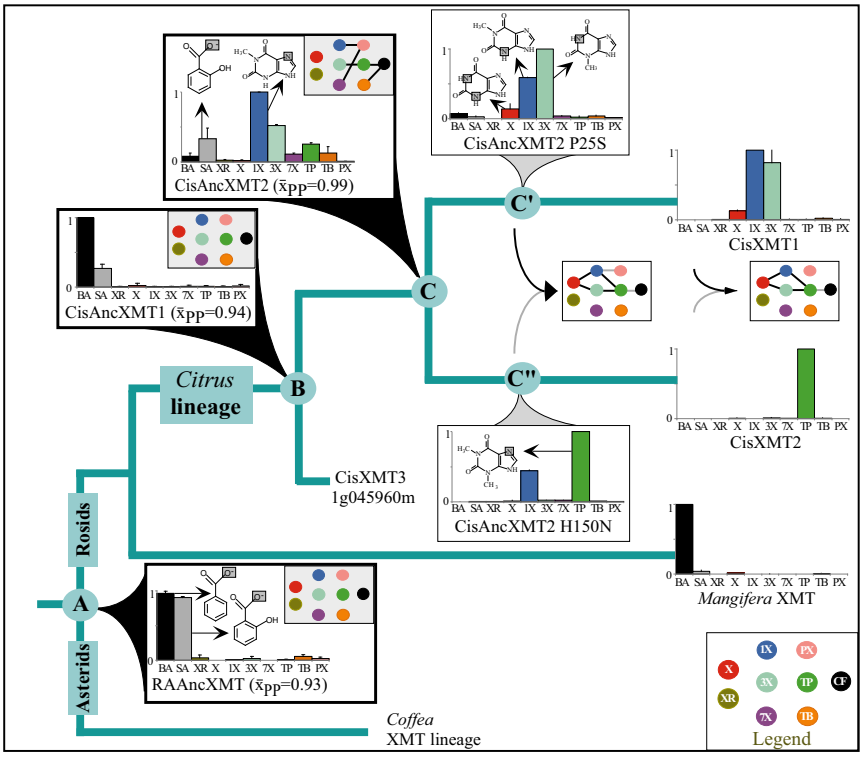


**Fig. 3.** Resurrected ancestral XMT proteins reveal the historical context for convergent evolution of caffeine biosynthesis. Bar charts show mean relative enzyme activities (from 0 to 1) for 10 substrates. BA, benzoic acid; SA, salicylic acid; all others are as in Fig. 1. Node A shows the resurrected enzyme of the >100-My-old ancestor of Rosids and Asterids that exhibits high relative activity with benzoic and salicylic acid. Although those ancestral activities were maintained in CisAncXMT1 at node B and modern-day *Mangifera*, they were eventually replaced by increased relative preference for xanthine alkaloid methylation as seen at node C and its descendants. CisAncXMT2 mutants (P25S and H150N indicated on lineages C′ and C″, respectively) show that very few amino acid replacements are necessary to re-evolve modern-day enzyme activity patterns and form a complete caffeine biosynthetic pathway. Product formation from assays and implied pathway connections are shown by color-coded dots (*Insets*). For example, a connection between TP (green) and CF (black) implies that the enzyme in question converts theophylline to caffeine. Unshaded rectangles exhibit complete metabolic connections to caffeine, whereas shaded rectangles do not. Average site-specific posterior probabilities are shown for each resurrected ancestral enzyme. Select substrate structures are shown to specify the atom to which a methyl group is transferred.

(21, 23, 46, 47) (Table S1). Additionally, although xanthine alkaloids may not be homogeneously distributed at the subcellular and tissue level (48), reported concentrations of the relevant intermediates in *Theobroma*, *Paullinia*, and *Citrus* may be conservatively estimated to be in the 10–1,000 μM range (34, 38, 41), which are comparable to the $K_M$ values we obtained (Table S1). Although it is apparent that three analogous pathways for caffeine biosynthesis have evolved in flowering plants, it is unclear what historical genetic and biochemical conditions facilitated this convergence.

**Historical Maintenance of Ancestral XMT Enzymes Allowed for Convergence.** In the case of convergent caffeine production in *Citrus* and *Coffea*, XMT needed to be maintained for more than 100 My from their common ancestor (49) to then independently give rise to xanthine alkaloid-methylating enzymes, because it is unlikely that their progenitor was producing caffeine given that they currently use completely different biosynthetic pathways (Fig. 2 *D* and *F*). To understand how this long-term maintenance occurred, we resurrected ancestral enzymes (50) for the XMT lineage at nodes A–C (Fig. 3 and Fig. S1). Surprisingly, both the putative ca. 100-My-old Rosid-Asterid ancestral enzyme, RAAncXMT (node A), as well as its descendant, CisAncXMT1 (node B), exhibit high relative activity with benzoic acid and salicylic acid (to form methyl benzoate and methyl salicylate, respectively) but very little with xanthine alkaloids (Fig. 3). Ancestral O-methylation of benzoic acid is maintained in a modern-day XMT from *Mangifera*, which is a relative of *Citrus* in Sapindales but is not known to synthesize xanthine alkaloids (Fig. 3 and Fig. S1). On the other hand, ancestral activities with benzoic and salicylic acid were completely lost in the modern-day descendant enzymes of *Citrus*, CisXMT1 and CisXMT2, and they now appear specialized for only N-methylation of xanthine alkaloids (Fig. 3). These specialized modern-day enzymes were most recently derived from CisAncXMT2 (node C), which exhibits both O- and N-methylation activities and seems to be a transitional enzyme associated with the gain of xanthine alkaloid production (Fig. 3). Although CisAncXMT2 appears to have low levels of activity with benzoic and salicylic acid compared with that of 1-methylxanthine, the specific activity with these substrates (0.6 and 2.3 pkat/mg, respectively) is comparable to heterologously expressed, modern-day SAMT- and BSMT-type enzymes (51). Today, modern-day *Citrus* possesses an SAMT that is capable of methylating both benzoic and salicylic acid, thereby compensating for the eventual loss of those activities from ancestral XMT enzymes (Fig. S4*B*). Ancestral O-methylation of the carboxyl moiety of benzoic and salicylic acid might have promoted the evolution of N-methylation of xanthine alkaloids because of common attributes of the active sites that would need to accommodate the largely planar rings of both classes of substrates. Indeed, a paralogous SABATH methyltransferase specialized for methylation of the carboxyl group of nicotinic acid (which is also an N-heterocyclic substrate) also recently arose from ancestral enzymes that exhibited activities with benzoic and salicylic acid (46). Although these data indicate that ancestral activity with benzoic and salicylic acid in the XMT lineage allowed for subsequent co-option of the descendant enzymes to form caffeine, how recruitment of the enzymes into a functional pathway occurred remains unknown.

**Exaptation Facilitates Multistep Pathway Evolution in a Cumulative Manner.** To understand how convergent caffeine production evolved via an entirely novel pathway in the *Citrus* lineage, we mapped ancestral XMT pathway connections of the caffeine biosynthetic network (Fig. 3, *Insets*, dot boxes). At nodes A and B, ancestral enzymes exhibited very low activity with xanthine alkaloids, such that quantities were too low to allow for product identification by HPLC, making it unlikely that a complete pathway existed at those times. Subsequently, the derived ancestral *Citrus* enzyme, CisAncXMT2 (node C), had activity with numerous xanthine alkaloids; in particular, highest relative activity with 1-methylxanthine resulted in paraxanthine formation, and 3-methylxanthine was methylated to form theophylline (Fig. S7 *A* and *B*). CisAncXMT2 could also convert theophylline to caffeine, such that it would have performed two of the three steps necessary to form caffeine from 3-methylxanthine (Fig. S7*C*). However, this ancestral enzyme exhibits only a low level of activity

and specificity with xanthine to form 1-methylxanthine (Fig. 3, Fig. S8, and Table S1), which could have subsequently been converted to paraxanthine by CisAncXMT2, but not caffeine (Fig. 3, node C). Thus, if CisAncXMT2 was used for methylation of benzoic and salicylic acid, then it appears to have been exapted for several later reactions of the xanthine alkaloid biosynthetic network used by modern-day *Citrus*, because a complete pathway to caffeine was not likely catalyzed by this enzyme alone. Alternatively, it remains formally possible that the ancestor of *Citrus* possessed a different, now extinct, enzyme that could have converted xanthine to 3-methylxanthine, so that caffeine may have been produced by CisAncXMT2, yet only modern-day XMT- and CS-type enzymes (Fig. 2) are capable of that conversion, making it unclear why an enzyme specialized for that reaction would subsequently be lost given its importance today.

Next, to recapitulate the evolutionary steps required to generate complete caffeine biosynthetic pathway linkages, we performed experimental mutagenesis of CisAncXMT2 (node C), which was duplicated to give rise to the two modern-day enzymes of *Citrus*, CisXMT1 and CisXMT2. In the lineage leading to CisXMT1, 17 amino acids were replaced and resulted in the evolution of increased activity with xanthine as well as specialization with 1- and 3-methylxanthine (Fig. 3). We experimentally replaced Pro25 by Ser in CisAncXMT2 (Fig. 3; lineage C′), because this site is predicted to be part of the active site of *Coffea* DXMT (52) and differs in CisXMT1 and CisXMT2 (Fig. S9). This single mutation resulted in the evolution of three important biochemical changes. First, near-complete loss of ancestral activity with theophylline as well as with benzoic and salicylic acid occurred such that CisAncXMT2 P25S acquired a relative activity profile very similar to modern-day CisXMT1 (Fig. 3). Second, CisAncXMT2 P25S exhibited a 2.5-fold increased catalytic efficiency with xanthine compared with CisAncXMT2 (Table S1) to produce both 1- and 3-methylxanthine (Fig. 3 and Fig. S8). Third, activity of CisAncXMT2 P25S changed such that 1-methylxanthine is methylated to form theophylline instead of paraxanthine (Fig. S10). The importance of this single amino acid replacement is that a connected biosynthetic network from xanthine to theophylline via both 1- and 3-methylxanthine would have rapidly evolved, in part, due to exaptation of CisAncXMT2 (Fig. 3). The existence of exapted ancestral enzymes such as CisAncXMT2 resolves one of the fundamental problems of the cumulative hypothesis because multiple steps of a pathway could evolve simultaneously, thereby avoiding the need to assume the existence of selectively advantageous intermediates. These results also point to a crucial role for ancestral promiscuous activities postulated as part of the patchwork hypothesis and protein engineering studies (31) and reported previously for other SABATH enzymes (46).

Finally, although the immediate, postduplication daughter enzyme of lineage C′′ would have initially retained the ancestral activities of it progenitor, CisAncXMT2, it eventually gave rise to the modern-day descendant CisXMT2, which exhibits near-complete specialization with theophylline (Fig. 3). A total of 16 amino acid replacements occurred along this lineage, one of which was His150, which was replaced by Asn (Fig. S9). This residue is likely part of the active site and known to control substrate preference in other SABATH enzymes (46). Experimental mutagenesis of His150 to Asn in CisAncXMT2 (Fig. 3, node C′′) resulted in an enzyme that is similar to modern-day CisXMT2. Specifically, H150N nearly completely abolished methylation activity with every substrate except theophylline and, to a lesser extent, 1-methylxanthine (Fig. 3). Because methylation of theophylline results in the formation of caffeine, the combined activities of CisAncXMT2 H150N and CisAncXMT2 P25S would have allowed for a complete caffeine biosynthetic pathway given xanthine as a starting substrate, much like the two modern-day XMT enzymes in *Citrus*. Although mutations other than H150N and P25S might have shifted ancestral enzymes toward the modern-day specialized activities of CisXMT1 and CisXMT2, we show that only these two replacements need to be implicated in specialization, because other mutations do not recapitulate the inferred relative activity changes (Fig. S9*C*).

## Conclusion

The results for the XMT lineage indicate that convergent evolution of caffeine biosynthesis was possible partly because ancient lineages of enzymes were maintained over 100 My for alternative biochemical functions. Furthermore, like the fortuitous roles of feathers for flight in birds (53) or ligand binding in ancestral hormone receptors (54), it appears that exapted activities of the ancestral XMT enzymes ultimately promoted their co-option for caffeine biosynthesis. These exaptations became biochemically relevant when, as predicted under the cumulative hypothesis (26), the initial reactions of the caffeine pathway evolved. The fact that very few substitutions to CisAncXMT2 were required to promote substrate preference switches suggests relatively facile mutational basis for the evolution of caffeine biosynthetic pathways. Therefore, it is likely that caffeine biosynthesis would evolve in flowering plants again if the evolutionary tape of life were to be replayed (55). What is more difficult to predict is which of the 12 potential biochemical pathways any particular lineage will use, which methyltransferase enzyme will be co-opted, or which amino acids will be substituted to provide for particular substrate preferences due to the role of historical contingency associated with any given evolutionary transition.

## Materials and Methods

**Heterologous Expression and Purification of Enzymes.** Gene sequences for *Theobroma* and *Paullinia* were synthesized with codon use optimized for gene expression in *Escherichia coli* (GenScript). Gene sequences for *Citrus* and *Camellia* were cloned from fresh flowers or leaves, respectively, using primers designed from the EST and genomic sequences. cDNA was generated using the SuperScript II/Platinum Taq One-Step RT-PCR Kit (Invitrogen). Protein overexpression used either pET-15b (Novagen) or Expresso T7 SUMO (Lucigen) expression vectors, and induction of $His_6$–protein was achieved in 50-mL BL-21 (DE3) cell cultures with the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside at 23 °C for 6 h. Purification of the $His_6$-tagged protein was achieved by TALON spin columns (Clontech) according to the manufacturer's instructions. To determine protein concentration, a standard Bradford assay was used. Recombinant protein purity was evaluated by SDS/PAGE.

**Enzyme Assays.** All enzymes were tested for activity with the eight xanthine alkaloid substrates shown in Fig. 1. In addition, all enzymes were tested with benzoic and salicylic acid, but we only report results for XMT enzymes, as shown in Fig. 3, because CS enzymes do not show activity with those substrates. Xanthine alkaloid substrates were dissolved in 0.5 M NaOH, whereas benzoic and salicylic acid were in ethanol. Radiochemical assays were performed in 50-μL reactions with 0.01 μCi (0.5 μL) $^{14}$C-labeled SAM, 100 μM methyl acceptor substrate, and 10–20 μL purified protein in 50 mM Tris·HCl buffer at 24 °C for 20 min. Negative controls were composed of the same reagents except that the methyl acceptor substrate was omitted and the corresponding solvent was added instead. Methylated products were extracted in ethyl acetate and quantified using a liquid scintillation counter. Raw disintegrations per min obtained from the scintillation counter were corrected using empirically determined extraction efficiencies of products in ethyl acetate. The highest enzyme activity reached with a specific substrate was set to 1.0, and relative activities with the remaining substrates were calculated. Each assay was run at least twice so that mean plus SD, could be calculated. All assays shown were performed on purified protein unless activity was abolished after purification. In such cases (only CisXMT2 and TCS2), we present total protein data. The specific activity for TCS2 with xanthosine was 0.025 pkat/mg, and the specific activity for CisXMT2 with theophylline was 0.12 pkat/mg.

**Ancestral Sequence Resurrection and Mutagenesis.** CODEML (56) was used to estimate ancestral sequences for the XMT lineage of enzymes of the SABATH family assuming the Jones, Taylor, Thornton (JTT) + gamma model of amino acid substitution. Regions with alignment gaps were analyzed with parsimony to determine ancestral residue numbers. The estimated sequences were subsequently synthesized by GenScript with codons chosen for optimal protein expression in *E. coli*. Alternative ancestral alleles were generated by site-directed mutagenesis using the QuikChange Site-Directed Mutagenesis Kit (Stratagene) to change amino acids that differed among analyses using different subsets of

sequences, trees, and models of substitution. Although posterior probabilities were high for most sites of most alleles (see average site-specific posterior probabilities in Fig. 3), different analyses did result in different estimated ancestral alleles in some cases. Therefore, at least two ancestral enzymes were characterized for each node A–C in Fig. 3, and an alignment showing each resurrected allele is provided in Fig. S9*A*. Amino acid sites differing between alternate alleles are shown in Fig. S9*B*, with posterior probabilities listed for each amino acid that was mutated. The positions of mutations are shown in the alignment of Fig. S9*A*. Because assays at each node were represented by at least two alleles, mean and SE were calculated for relative activity with each substrate assayed.

Detailed procedures for bioinformatic, phylogenetic, enzyme kinetics, HPLC, and LC-MS/MS analyses are provided in *SI Materials and Methods*. See Table S2 for MS/MS parameters and LC retention time for target xanthine alkaloids.

**ACKNOWLEDGMENTS.** We thank Greg Cavey, Talline Martins, Andre Venter, James Kiddle, Matt Gibson, Logan Rowe, Rachel Steuf, and Kevin Blair for helpful discussions and assistance; Gane Ka-Shu Wong for providing access to the oneKP database to confirm placements of gene sequences shown in Fig. S1; the Chemistry Department at Western Michigan University for facilitating our HPLC analyses; and Thomas Baumann for allowing us to use his photos of *Camellia* and *Paullinia* in Fig. 2. This study was supported by National Science Foundation Grant MCB-1120624 (to T.J.B.).

1. Rajakumar R, et al. (2012) Ancestral developmental potential facilitates parallel evolution in ants. *Science* 335(6064):79–82.
2. Sanger TJ, Revell LJ, Gibson-Brown JJ, Losos JB (2012) Repeated modification of early limb morphogenesis programmes underlies the convergence of relative limb length in *Anolis* lizards. *Proc Biol Sci* 279(1729):739–748.
3. Pichersky E, Lewinsohn E (2011) Convergent evolution in plant specialized metabolism. *Annu Rev Plant Biol* 62:549–566.
4. Des Marais DL, Rausher MD (2010) Parallel evolution at multiple levels in the origin of hummingbird pollinated flowers in *Ipomoea*. *Evolution* 64(7):2044–2054.
5. Yoon HS, Baum DA (2004) Transgenic study of parallelism in plant morphological evolution. *Proc Natl Acad Sci USA* 101(17):6524–6529.
6. Reed RD, et al. (2011) *optix* drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science* 333(6046):1137–1141.
7. Berens AJ, Hunt JH, Toth AL (2015) Comparative transcriptomics of convergent evolution: Different genes but conserved pathways underlie caste phenotypes across lineages of eusocial insects. *Mol Biol Evol* 32(3):690–703.
8. Schwarze K, et al. (2014) The globin gene repertoire of lampreys: Convergent evolution of hemoglobin and myoglobin in jawed and jawless vertebrates. *Mol Biol Evol* 31(10):2708–2721.
9. Rosenblum EB, Römpler H, Schöneberg T, Hoekstra HE (2010) Molecular and functional basis of phenotypic convergence in white lizards at White Sands. *Proc Natl Acad Sci USA* 107(5):2113–2117.
10. Losos JB (2011) Convergence, adaptation, and constraint. *Evolution* 65(7):1827–1840.
11. Ashihara H, Suzuki T (2004) Distribution and biosynthesis of caffeine in plants. *Front Biosci* 9:1864–1876.
12. Hammerstone JF, Romanczyk LJ, Aitken WM (1994) Purine alkaloid distribution within *Herrania* and *Theobroma*. *Phytochemistry* 35(5):1237–1240.
13. Weckerle CS, Stutz MA, Baumann TW (2003) Purine alkaloids in *Paullinia*. *Phytochemistry* 64(3):735–742.
14. Nathanson JA (1984) Caffeine and related methylxanthines: Possible naturally occurring pesticides. *Science* 226(4671):184–187.
15. Kim YS, et al. (2011) Resistance against beet armyworms and cotton aphids in caffeine-producing transgenic chrysanthemum. *Plant Biotechnol* 28(4):393–395.
16. Wright GA, et al. (2013) Caffeine in floral nectar enhances a pollinator's memory of reward. *Science* 339(6124):1202–1204.
17. Suzuki T, Takahashi E (1976) Caffeine biosynthesis in *Camellia sinensis*. *Phytochemistry* 15(8):1235–1239.
18. Ashihara H, Monteiro AM, Gillies FM, Crozier A (1996) Biosynthesis of caffeine in leaves of coffee. *Plant Physiol* 111(3):747–753.
19. D'Auria JC, Chen F, Pichersky E (2003) The SABATH family of MTS in *Arabidopsis thaliana* and other plant species. *Recent Advances in Phytochemistry*, ed Romeo JT (Elsevier Science, Oxford), pp 253–283.
20. Kato M, et al. (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.
21. 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.
22. Denoeud F, et al. (2014) The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* 345(6201):1181–1184.
23. Effmert U, et al. (2005) Floral benzenoid carboxyl methyltransferases: From in vitro to in planta function. *Phytochemistry* 66(11):1211–1230.
24. Seo HS, et al. (2001) Jasmonic acid carboxyl methyltransferase: A key enzyme for jasmonate-regulated plant responses. *Proc Natl Acad Sci USA* 98(8):4788–4793.
25. Varbanova M, et al. (2007) Methylation of gibberellins by *Arabidopsis* GAMT1 and GAMT2. *Plant Cell* 19(1):32–45.
26. Granick S (1957) Speculations on the origins and evolution of photosynthesis. *Ann N Y Acad Sci* 69(2):292–308.
27. Horowitz NH (1945) On the evolution of biochemical syntheses. *Proc Natl Acad Sci USA* 31(6):153–157.
28. Ycas M (1974) On earlier states of the biochemical system. *J Theor Biol* 44(1):145–160.
29. Jensen RA (1976) Enzyme recruitment in evolution of new function. *Annu Rev Microbiol* 30:409–425.
30. Bergthorsson U, Andersson DI, Roth JR (2007) Ohno's dilemma: Evolution of new genes under continuous selection. *Proc Natl Acad Sci USA* 104(43):17004–17009.
31. Aharoni A, et al. (2005) The 'evolvability' of promiscuous protein functions. *Nat Genet* 37(1):73–76.
32. Angelo PCSA, et al.; Brazilian Amazon Consortium for Genomic Research (REALGENE) (2008) Guarana (*Paullinia cupana* var. *sorbilis*), an anciently consumed stimulant from the Amazon rain forest: The seeded-fruit transcriptome. *Plant Cell Rep* 27(1):117–124.
33. Argout X, 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.
34. Koyama Y, Tomoda Y, Kato M, Ashihara H (2003) Metabolism of purine bases, nucleosides and alkaloids in theobromine-forming *Theobroma cacao* leaves. *Plant Physiol Biochem* 41(11–12):977–984.
35. Zheng XQ, Koyama Y, Nagai C, Ashihara H (2004) Biosynthesis, accumulation and degradation of theobromine in developing *Theobroma cacao* fruits. *J Plant Physiol* 161(4):363–369.
36. Yoneyama N, et al. (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):125–135.
37. Schimpl FC, et al. (2014) Molecular and biochemical characterization of caffeine synthase and purine alkaloid concentration in guarana fruit. *Phytochemistry* 105:25–36.
38. Baumann TW, Schulthess BH, Hanni K (1995) Guaraná (*Paullinia cupana*) rewards seed dispersers without intoxicating them by caffeine. *Phytochemistry* 39(5):1063–1070.
39. Kato M, Mizuno K, Crozier A, Fujimura T, Ashihara H (2000) Caffeine synthase gene from tea leaves. *Nature* 406(6799):956–957.
40. Stutz MA (2001) Purinalkaloide in Blüten. Zeitliche und räumliche Allokation. Aspekte der Biosynthese. Masters thesis (University of Zurich, Zurich).
41. Kretschmar JA, Baumann TW (1999) Caffeine in *Citrus* flowers. *Phytochemistry* 52(1): 19–23.
42. Mazzafera P (2004) Catabolism of caffeine in plants and microorganisms. *Front Biosci* 9:1348–1359.
43. Ashihara H, Crozier A (1999) Biosynthesis and metabolism of caffeine and related purine alkaloids in plants. *Adv Bot Res* 30:117–205.
44. Ashihara H, Crozier A (2001) Caffeine: A well known but little mentioned compound in plant science. *Trends Plant Sci* 6(9):407–413.
45. Suzuki T, Takahashi E (1975) Biosynthesis of caffeine by tea-leaf extracts. Enzymic formation of theobromine from 7-methylxanthine and of caffeine from theobromine. *Biochem J* 146(1):87–96.
46. Huang R, et al. (2012) Enzyme functional evolution through improved catalysis of ancestrally nonpreferred substrates. *Proc Natl Acad Sci USA* 109(8):2966–2971.
47. Kato M, et al. (1999) Purification and characterization of caffeine synthase from tea leaves. *Plant Physiol* 120(2):579–586.
48. van Breda SV, van der Merwe CF, Robbertse H, Apostolides Z (2013) Immunohistochemical localization of caffeine in young *Camellia sinensis* (L.) O. Kuntze (tea) leaves. *Planta* 237(3):849–858.
49. Wikstrom N, Savolainen V, Chase MW (2001) Evolution of the angiosperms: Calibrating the family tree. *Proc Biol Sci* 268(1482):2211–2220.
50. Thornton JW (2004) Resurrecting ancient genes: Experimental analysis of extinct molecules. *Nat Rev Genet* 5(5):366–375.
51. Hippauf F, et al. (2010) Enzymatic, expression and structural divergences among carboxyl O-methyltransferases after gene duplication and speciation in *Nicotiana*. *Plant Mol Biol* 72(3):311–330.
52. McCarthy AA, McCarthy JG (2007) The structure of two N-methyltransferases from the caffeine biosynthetic pathway. *Plant Physiol* 144(2):879–889.
53. Prum RO, Brush AH (2002) The evolutionary origin and diversification of feathers. *Q Rev Biol* 77(3):261–295.
54. Bridgham JT, Carroll SM, Thornton JW (2006) Evolution of hormone-receptor complexity by molecular exploitation. *Science* 312(5770):97–101.
55. Gould SJ (1989) *Wonderful Life* (WW Norton, New York).
56. Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8):1586–1591.
57. Forment J, et al. (2005) Development of a citrus genome-wide EST collection and cDNA microarray as resources for genomic studies. *Plant Mol Biol* 57(3):375–391.
58. Azam M, et al. (2013) Comparative analysis of flower volatiles from nine citrus at three blooming stages. *Int J Mol Sci* 14(11):22346–22367.
59. Figueirêdo LC, Faria-Campos AC, Astolfi-Filho 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.
60. Wall PK, et al. (2008) PlantTribes: A gene and gene family resource for comparative genomics in plants. *Nucleic Acids Res* 36(Database issue):D970–D976.
61. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 30(4):772–780.
62. Guindon S, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* 59(3):307–321.