Phytochemistry xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Phytochemistry

journal homepage: www.elsevier.com/locate/phytochem

Molecular and biochemical characterization of caffeine synthase and purine alkaloid concentration in guarana fruit

Flávia Camila Schimpl [a], Eduardo Kiyota [a], Juliana Lischka Sampaio Mayer [b], José Francisco de Carvalho Gonçalves [c], José Ferreira da Silva [d], Paulo Mazzafera [a,*]

[a] Departamento de Biologia Vegetal-IB, CP 6109, Universidade Estadual de Campinas, 13083-970 Campinas, SP, Brazil
[b] Departamento de Biologia, Universidade Estadual do Centro-Oeste, Guarapuava, PR, Brazil
[c] Instituto Nacional de Pesquisas da Amazônia, Manaus, AM, Brazil
[d] Departamento de Produção Animal e Vegetal, Faculdade de Ciências Agrárias, Universidade Federal do Amazonas, Manaus, AM, Brazil

ARTICLE INFO

Article history:
Received 13 November 2013
Received in revised form 17 February 2014
Available online xxxx

Keywords:
Amazon
Venezuelan guarana
Paullinia cupana var. *sorbilis*
Sapindaceae
Caffeine
Theobromine
Recombinant protein
Seed

ABSTRACT

Guarana seeds have the highest caffeine concentration among plants accumulating purine alkaloids, but in contrast with coffee and tea, practically nothing is known about caffeine metabolism in this Amazonian plant. In this study, the levels of purine alkaloids in tissues of five guarana cultivars were determined. Theobromine was the main alkaloid that accumulated in leaves, stems, inflorescences and pericarps of fruit, while caffeine accumulated in the seeds and reached levels from 3.3% to 5.8%. In all tissues analysed, the alkaloid concentration, whether theobromine or caffeine, was higher in young/immature tissues, then decreasing with plant development/maturation. Caffeine synthase activity was highest in seeds of immature fruit. A nucleotide sequence (*PcCS*) was assembled with sequences retrieved from the EST database REALGENE using sequences of caffeine synthase from coffee and tea, whose expression was also highest in seeds from immature fruit. The *PcCS* has 1083 bp and the protein sequence has greater similarity and identity with the caffeine synthase from cocoa (BTS1) and tea (TCS1). A recombinant PcCS allowed functional characterization of the enzyme as a bifunctional CS, able to catalyse the methylation of 7-methylxanthine to theobromine (3,7-dimethylxanthine), and theobromine to caffeine (1,3,7-trimethylxanthine), respectively. Among several substrates tested, PcCS showed higher affinity for theobromine, differing from all other caffeine synthases described so far, which have higher affinity for paraxanthine. When compared to previous knowledge on the protein structure of coffee caffeine synthase, the unique substrate affinity of PcCS is probably explained by the amino acid residues found in the active site of the predicted protein.

1. Introduction

Guarana is the name given to *Paullinia cupana* var. *sorbilis* Kunth, a native species to the Maués region in the Brazilian Amazon. The variety *sorbilis* is known as the Brazilian guarana, while *Paullinia cupana* var. *typica* is the Venezuelan guarana (Schimpl et al., 2013). Except for *Paullinia pinnata* L., which is also present in tropical Africa, all other species of the genus are restricted to tropical and subtropical America (Missouri Botanical Garden, 2012). In this study, guarana will designate *P. cupana* var. *sorbilis*, which is the only commercially exploited species and has been studied most intensively.

Guarana has the highest caffeine (1,3,7-trimethylxanthine) (**1**) concentration reported in plants and its seeds may contain from 2.5% to 6.5% caffeine at dry weight (Oliveira, 2010; Spoladore et al., 1987). Theobromine (3,7-dimethylxanthine) (**2**) and theophylline (1,3-dimethylxanthine) (**3**) are present in smaller quantities (see structures 1, 2 and 3 in Fig. 1). Indians and natives of the Amazon have consumed guarana for centuries as an aqueous beverage made from the powder obtained from roasted seeds, whose best known physiological effect is improved alertness because of caffeine (Henman, 1986; Schimpl et al., 2013; Schmidt, 1941).

Caffeine (**1**) has been reported in 13 plant orders, mostly eu-dicotyledonous plants. With minor variations, the main route of its synthesis is highly conserved in the plants studied so far, mainly coffee (*Coffea arabica*) and tea (*Camellia sinensis*) (Ashihara et al., 2008; Ashihara and Suzuki, 2004; Mazzafera, 2004).

Caffeine (**1**) biosynthesis has three methylation steps and xanthosine is the first compound to be methylated. It has been suggested that there are four possible routes in the origin of

* Corresponding author. Tel.: +55 19 3521 6213.
E-mail address: pmazza@unicamp.br (P. Mazzafera).

http://dx.doi.org/10.1016/j.phytochem.2014.04.018
0031-9422/

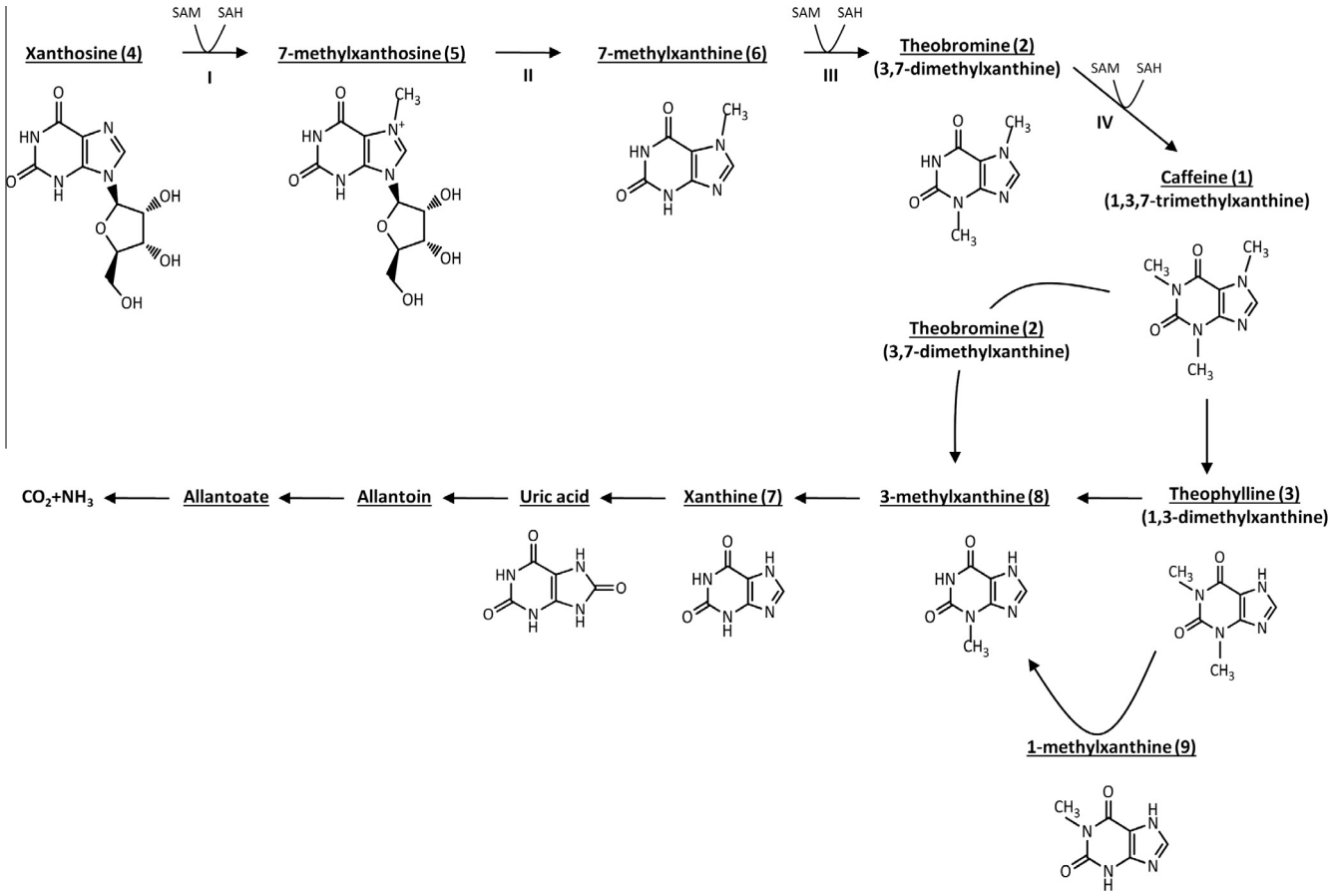

Fig. 1. Main metabolic pathway for the biosynthesis and biodegradation of caffeine. SAM = S-adenosyl-L-methionine, SAH = S-adenosyl-L-homo-cysteine. I = 7-methylxanthosine synthase (xanthosine methyltransferase), II = nucleosidase, III = theobromine synthase or caffeine synthase, IV = caffeine synthase.

xanthosine (**4**) (Ashihara, 2006; Ashihara et al., 2008). In coffee and tea, the caffeine biosynthesis pathway is: xanthosine (**4**) → 7-methylxanthosine (**5**) → 7-methylxanthine (**6**) → theobromine (3,7-dimethylxanthine) (**2**) → caffeine (1,3,7-trimethlxanthine) (**1**) (Fig. 1) (Ashihara et al., 2008; Mizuno et al., 2003a; Uefuji et al., 2003). S-adenosyl-L-methionine (SAM) is the methyl donor in caffeine (**1**) biosynthesis (Suzuki, 1972). The genes coding for the *N*-methyltransferase enzymes catalysing the three methylations have been studied mainly in coffee, and to a lesser extent in tea (Ashihara et al., 2008, 2011; Kato and Mizuno, 2004).

So far studies on guarana have mostly been restricted to characterization of caffeine (**1**) and other purine alkaloid concentration in the seeds. Recently, with the aim of increasing genetic knowledge of this plant, an EST database was made from guarana fruit at three different stages of maturation [Rede Amazônia Legal de Pesquisas Genômicas – Realgene (Ângelo et al., 2008)]. Recently, Figueiredo et al. (2011) analysed the 15,490 ESTs of the Realgene database and, among 4697 full-length clones identified, 84 clones were identified as *N*-methyltransferases and 18 were sequenced for confirmation of their identity. Phylogenetic analyses were carried out for three clones (with only 0.017% dissimilarity), which were close to a postulated caffeine synthase of *Theobroma cacao* (BTS1/BCS1). However, the recombinant protein from the cocoa sequence and other sequences resembling caffeine synthase only had 3-*N*-methyltransferase activity, i.e., theobromine synthase activity (Yoneyama et al., 2006).

In this study, the purine alkaloid concentration in different tissues of five guarana cultivars grown in Brazil was analyzed and, in one of them, caffeine synthase activity was assayed. Using the Realgene database, a nucleotide sequence displaying was assembled high similarity to caffeine synthase from cocoa, and the recombinant protein proved to be a bifunctional caffeine synthase, being able to methylate 7-methylxanthine (**8**) and theobromine (**2**). In contrast with all caffeine synthases studied so far, which have paraxanthine (**8**) as the best substrate, *P. cupana* caffeine synthase (PcCS) has theobromine as the best substrate and no activity against paraxanthine. *PcCS* had higher expression in immature seeds compared to other guarana tissues.

2. Results

2.1. Methylxanthines concentration

Theobromine (**2**) was the main alkaloid found in leaves, followed by theophylline (**3**), and higher concentrations of both alkaloids were found in young leaves. Caffeine (**1**) was detected only in intermediate and mature leaves. Theobromine (**2**) concentration ranged from 3.47% to 4.18% (Fig. 2).

Similar to leaves, apical and basal portions of stems had theobromine (**2**) as the predominant alkaloid, followed by theophylline (**3**). Caffeine (**1**) was detected only in the basal portion. Among the tissues analysed, stems were those containing the lowest methylxanthine (**8**) concentrations (Fig. 3).

Theobromine (**2**) was also the most abundant alkaloid in the inflorescences of all cultivars and, for most of them, the theophylline (**3**) concentration was higher than caffeine (**1**) (Fig. 4).

In the pericarp, theobromine (**2**) was also the predominant alkaloid but, in this tissue, caffeine (**1**) was more abundant than theophylline (**3**) (Fig. 5). In general, the alkaloid concentration

Please cite this article in press as: Schimpl, F.C., et al. Molecular and biochemical characterization of caffeine synthase and purine alkaloid concentration in guarana fruit. Phytochemistry (2014), http://dx.doi.org/10.1016/j.phytochem.2014.04.018

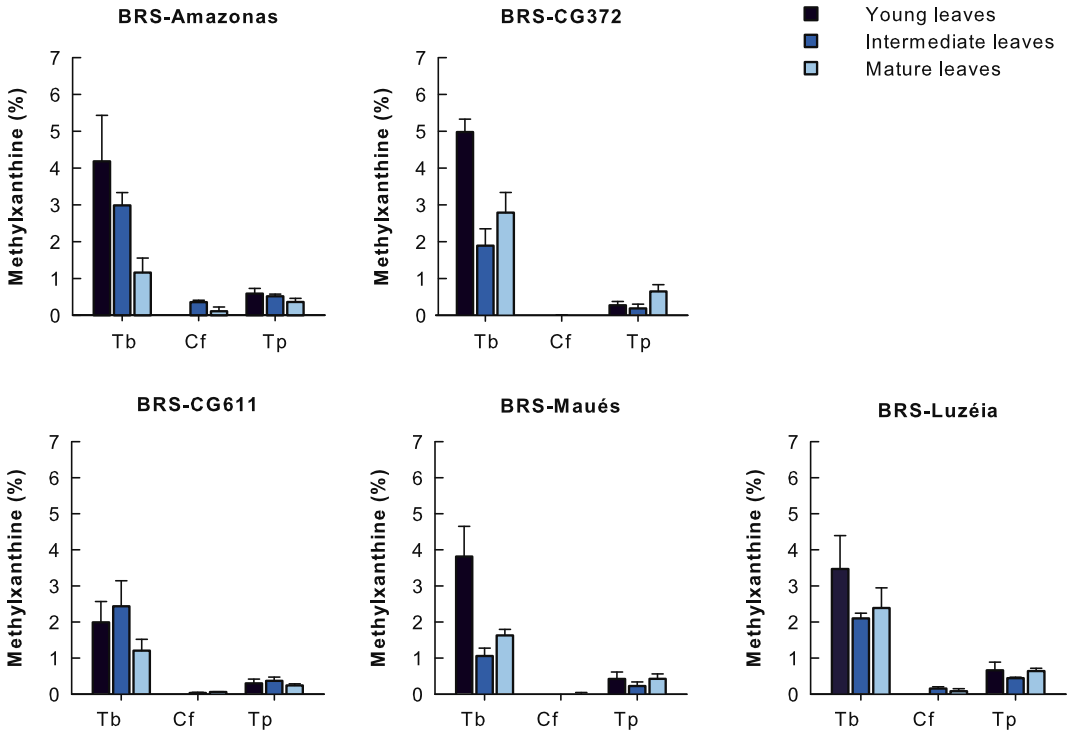

Fig. 2. Levels of theobromine (**2**) (Tb), caffeine (**1**) (Cf) and theophylline (**3**) (Tp) in leaves at three developmental stages of five guarana cultivars. Bars indicate standard deviation of three replicates.

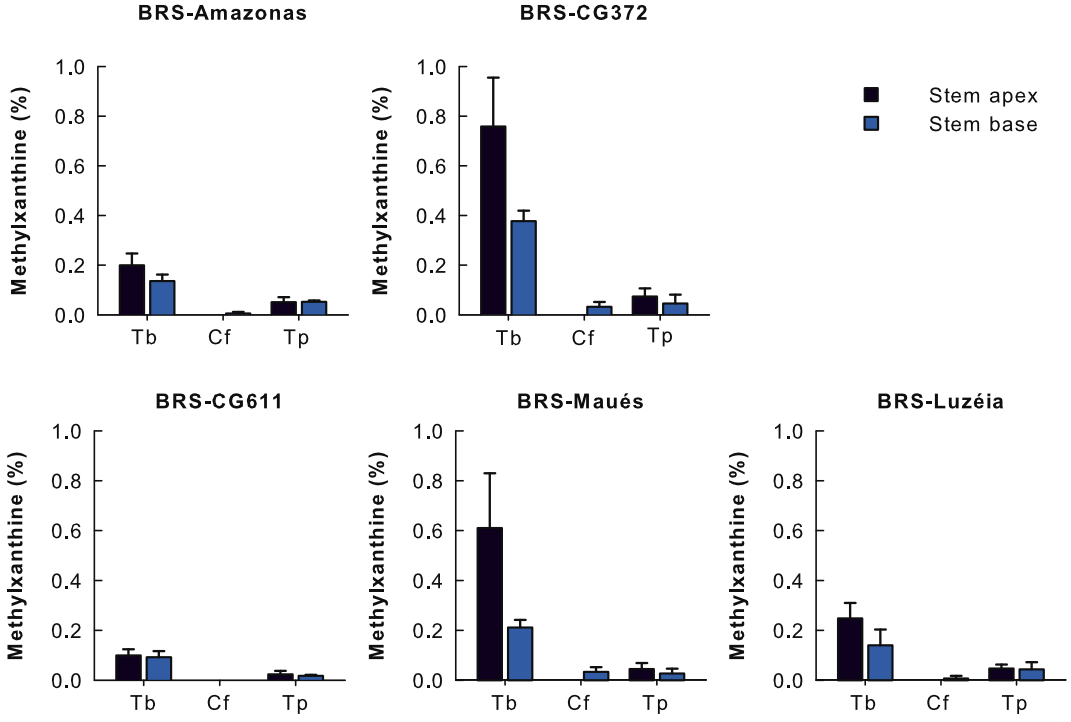

Fig. 3. Levels of theobromine (**2**) (Tb), caffeine (**1**) (Cf) and theophylline (**3**) (Tp) in apical and basal stems of five cultivars of guarana. Bars indicate standard deviation of three replicates.

decreased in the pericarp with fruit maturation (Fig. 5). The caffeine (**1**) concentration was higher than theobromine (**2**) only in seeds, and the values ranged from 3.3% to 5.8% (Fig. 5). As in leaves and stems, caffeine concentration was higher in young tissues (immature seeds). Theophylline (**2**) concentrations in the pericarp and seeds were less than 0.04% (Fig. 5).

2.2. Multiple alignment and phylogenetic analysis

The translated amino acid sequence obtained from the *PcCS* nucleotide sequence showed high similarity with three sequences previously assembled from the Realgene EST database (Figueiredo et al., 2011), whose expression or functionality was not tested

Please cite this article in press as: Schimpl, F.C., et al. Molecular and biochemical characterization of caffeine synthase and purine alkaloid concentration in guarana fruit. Phytochemistry (2014), http://dx.doi.org/10.1016/j.phytochem.2014.04.018

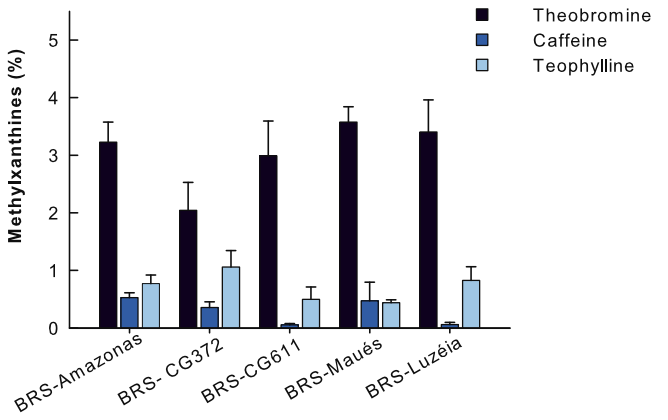

Fig. 4. Levels of theobromine (**2**) (Tb), caffeine (**1**) (Cf) and theophylline (**3**) (Tp) in inflorescences of cultivars of guarana. Bars indicate standard deviation of three replicates.

(Supplementary Fig. S1). PcCS was also very similar to other *N*-methyltransferases of caffeine-accumulating plants and presents the highly conserved domains A, B′, C (specific for binding of S-adenosylmethionine) and YFFF (Fig. 6). Phylogenetic analysis showed two very distinct clades (Fig. 7). The guarana caffeine synthase amino acid sequence was phylogenetically grouped in a clade with sequences of caffeine synthase of theobromine-accumulating species: *C. sinensis* (TCS1: AB031280), *C. irrawadiensis* (ICS1: AB056108), *C. ptilophylla* (PCS1: AB207817) and *T. cacao* (BTS1, AB096699) (Fig. 7). The other clade grouped sequences of caffeine synthase, as well as theobromine synthase and 7-methylxanthine synthase from coffee species (*C. arabica* and *C. canephora*).

2.3. Gene expression of PcCS in guarana fruit

PcCS is expressed in all tissues of the BRS-Amazonas cultivar (Fig. 8). Roots and old leaves showed the lowest transcript levels among the vegetative tissues. The highest expression in fruit tissues was observed in seeds, but in both seeds and pericarps, the expression level decrease with maturation. *PcCS* expression in seeds from green fruit was 7–20 times higher than in the other tissue and stages. The results obtained from RT-qPCR were confirmed by *in situ* hybridization analysis, as *PcCS* was expressed in the cotyledon of seeds of immature and intermediate stages of fruit maturation (Fig. 9A and B, see arrows) but not in mature fruits (Fig. 9D). The probe signal intensity in seed tissues was low probably because of the presence of many starch granules and other interfering substances. No positive probe signal was observed in other parts of the fruit (seed coat, aril and pericarp – not shown).

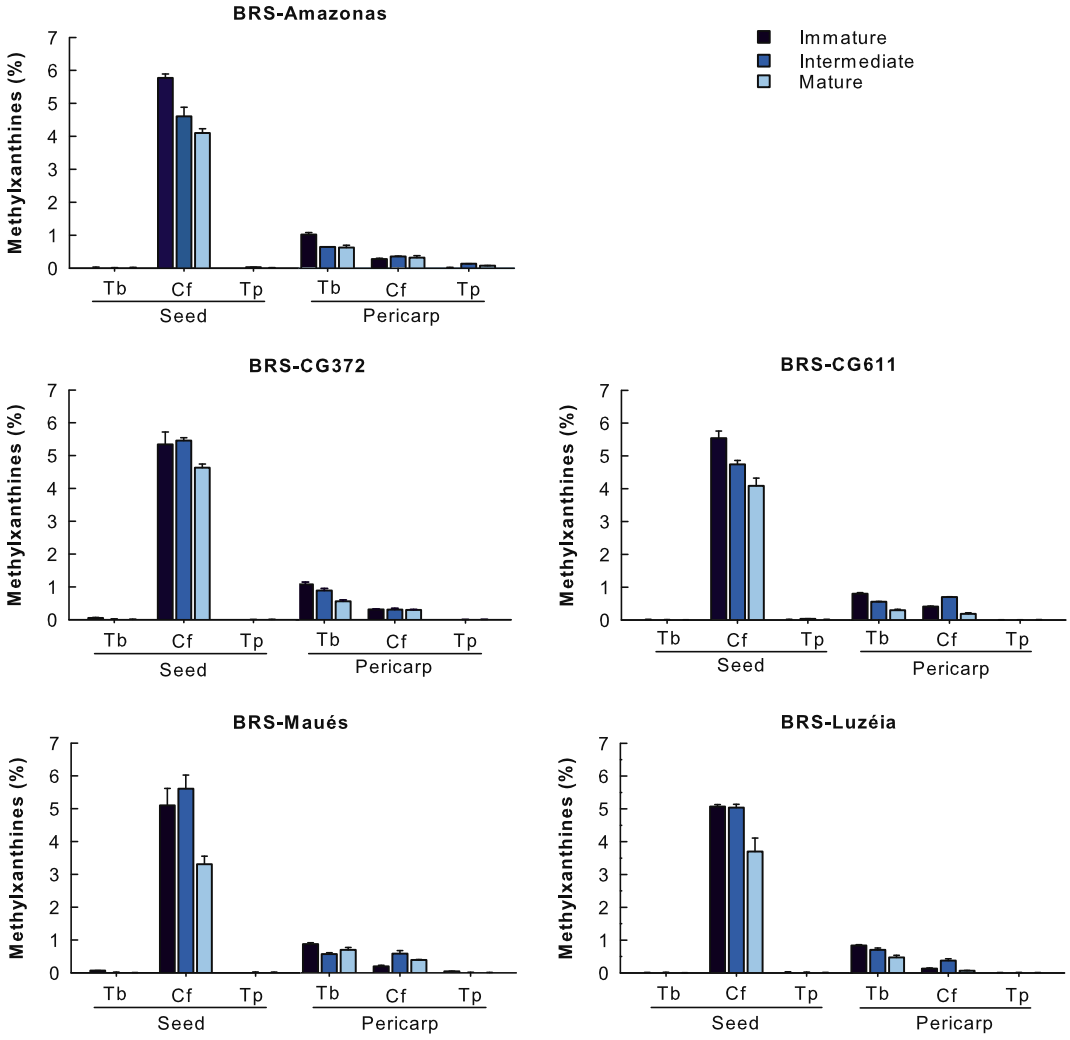

Fig. 5. Levels of theobromine (**2**) (Tb), caffeine (**1**) (Cf) and theophylline (**3**) (Tp) in seeds and pericarp of fruit at different stages of maturation of five guarana cultivars. Bars indicate standard deviation of three replicates.

Please cite this article in press as: Schimpl, F.C., et al. Molecular and biochemical characterization of caffeine synthase and purine alkaloid concentration in guarana fruit. Phytochemistry (2014), http://dx.doi.org/10.1016/j.phytochem.2014.04.018

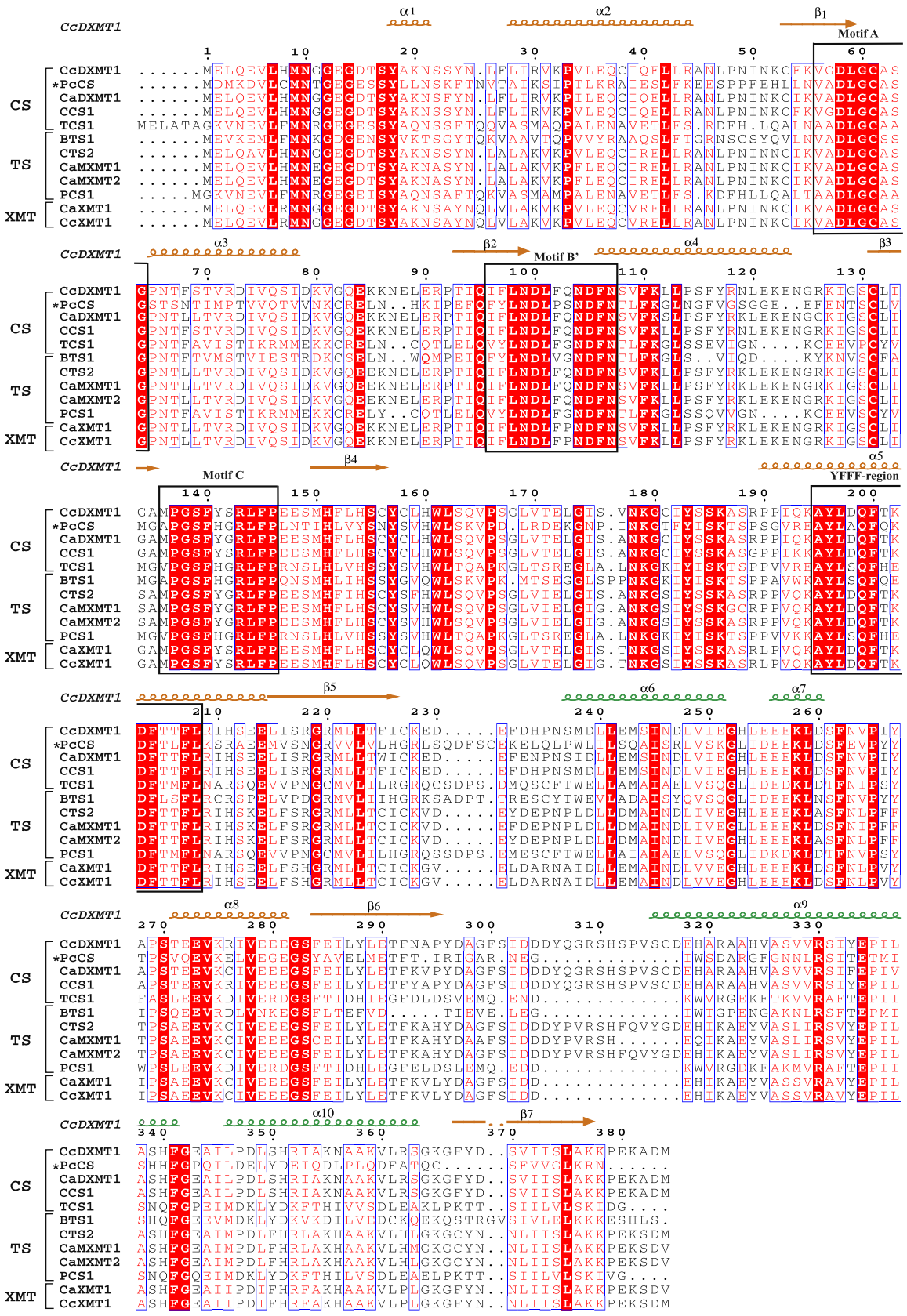

Fig. 6. Alignment of PcCS translated amino acid sequence with *N*-methyltransferase sequences from other species and secondary structures moulded from *C. canephora* CcDXMT1 (caffeine synthase). The core SAM binding domain is indicated in green and the folding helical domain is indicated in orange. The SAM binding motifs (A, B' and C) and YFFF conserved region are shown by black boxes. Access Numbers: CcDXMT: DQ422955; CaDXMT1: AB084125; CCS1: AB086414; TCS1: AB031280; BTS1: AB096699; CTS2: AB054841; CaMXMT1: AB048794; CaMXMT2: AB084126; PCS1, AB207817; CaXMT1: AB048793; CcXMT1: DQ422954. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Schimpl, F.C., et al. Molecular and biochemical characterization of caffeine synthase and purine alkaloid concentration in guarana fruit. Phytochemistry (2014), http://dx.doi.org/10.1016/j.phytochem.2014.04.018

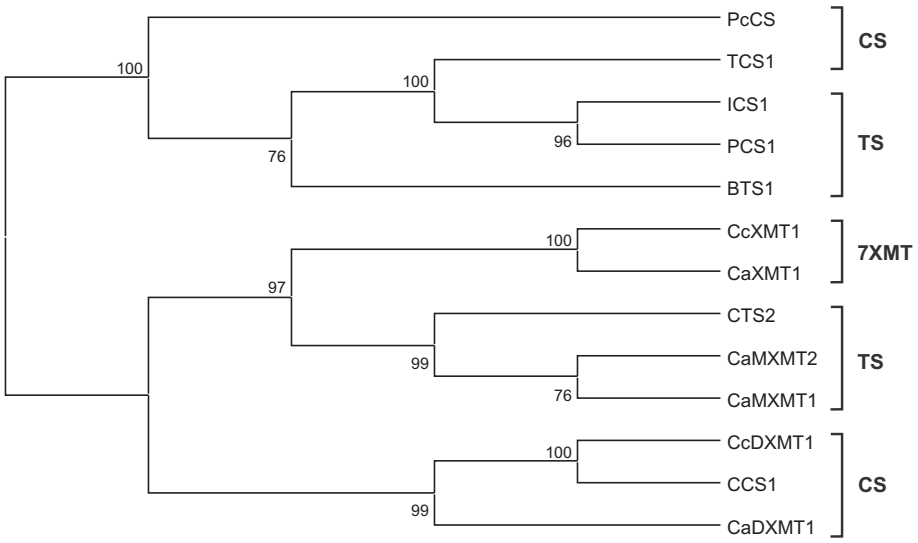

Fig. 7. Phylogenetic tree of *N*-methyltransferase amino acid sequences. *Coffea arabica* (CaXMT1: AB048793, CTS2: AB054841, CaMXMT1: AB084794, CaMXMT2: AB084126, CaDXMT1: AB084125, CCS1: AB086414), *Coffea canephora* (CcXMT1: DQ422954, CcDXMT: DQ422955), *Camellia sinensis* (TCS1: AB031280), *Camellia irrawadiensis* (ICS1: AB056108), *Camellia ptilophylla* (PCS1: AB207819) and *Theobroma cacao* (BTS1: AB096699). Functional activities are indicated as 7MXS: 7-methylxanthine synthase, TS: theobromine synthase, CS: caffeine synthase.

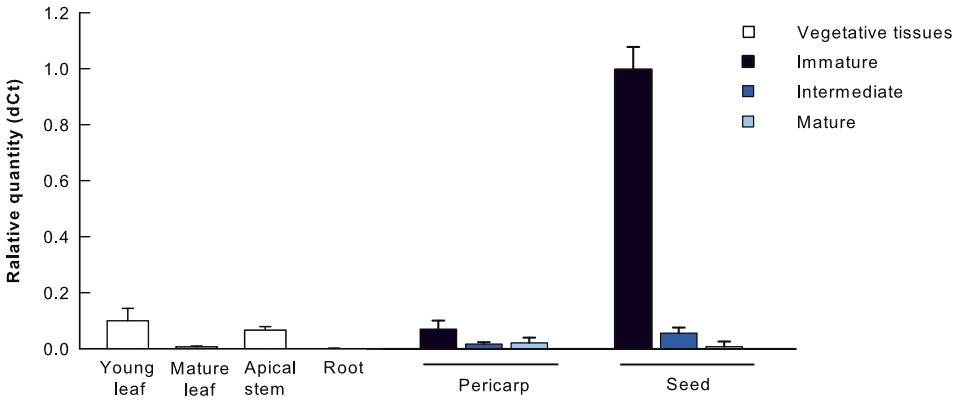

Fig. 8. *PcCS* expression in vegetative and fruit tissues of the cultivar BRS-Amazonas. Bars indicate standard deviation of three replicates.

2.4. Recombinant caffeine synthase

PcCS was expressed in *Escherichia coli* BL21-pRil (Supplementary Fig. S2) and although most of the recombinant protein was retained in the pellet, the soluble fraction purified by affinity chromatography in Ni–NTA agarose yielded a reasonably pure protein fraction (lane 5, Supplementary Fig. S2). Studies on this purified fraction showed a K_m = 328 μM and V_{max} = 20.7 nmol min^{-1} for 7-methylxanthine (**6**) and a K_m = 277 μM and V_{max} = 10.0 nmol min^{-1} for theobromine (**2**). The recombinant PcCS was able to catalyse the methylations of 7-methylxanthine (**6**) to theobromine (**2**) and of theobromine (**2**) to caffeine (**1**) (Table 1). The protein has a theoretical molecular mass of 40.1 kDa and theoretical pI of 5.14 (http://ca.expasy.org/).

2.5. Caffeine synthase activity

Caffeine synthase activity was determined in protein extracts from guarana seeds at three stages of maturation. Activity was higher in seeds of fruit at the green stage using theobromine as a substrate: green stage = 20.9 ± 2.4 fkat; intermediate stage = 6.6 ± 1.0 fkat; mature stage = 3.8 ± 0.9 fkat. Caffeine synthase of seeds has K_m = 258 μM and V_{max} = 0.5 nmol min^{-1}.

3. Discussion

3.1. Caffeine distribution in guarana tissues

In general, caffeine (**1**) is the main alkaloid in plants containing purine alkaloids and its biosynthesis is highest in young tissues (Ashihara et al., 2011). Other plants accumulating purine alkaloids, such as the tea species *C. irrawadiensis* (Nagata and Sakai, 1985) and *C. ptilophylla* (Ye et al., 1997), have theobromine (**2**) as the main alkaloid in young leaves. Theobromine (**2**) is also the predominant alkaloid in seeds of *T. cacao* (Senanayake and Wijesekera, 1971). In coffee seeds and leaves, caffeine (**1**) is the main alkaloid (Ashihara et al., 2011; Ashihara et al., 2008; Mazzafera, 2004; Mazzafera and Carvalho, 1991; Mazzafera et al., 1991).

In this study, analysis of five guarana cultivars showed that the levels of caffeine (**1**), theobromine (**2**) and theophylline (**3**) in guarana varied significantly among tissues and at different developmental stages. In all cultivars, theobromine (**2**) was the most abundant alkaloid in vegetative tissues (leaves and stems), flowers and pericarp, followed by theophylline (**3**), while caffeine (**1**) was predominant only in seeds. In general, the concentrations of all alkaloids were higher in young tissues, decreasing with tissue maturity/aging.

Please cite this article in press as: Schimpl, F.C., et al. Molecular and biochemical characterization of caffeine synthase and purine alkaloid concentration in guarana fruit. Phytochemistry (2014), http://dx.doi.org/10.1016/j.phytochem.2014.04.018

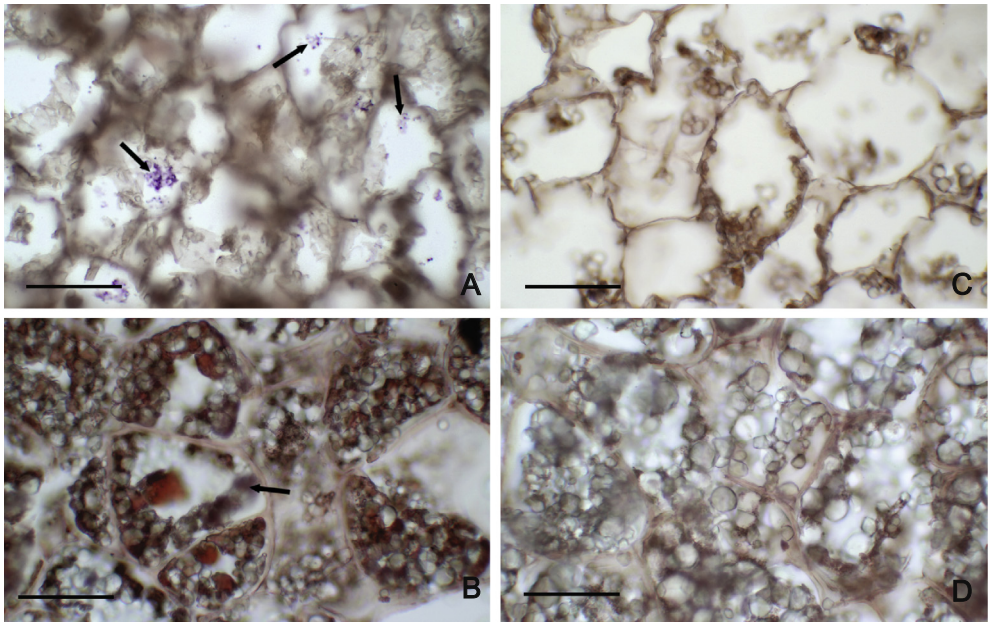

Fig. 9. *In situ* hybridization of *PcCS* in longitudinal sections of guarana fruit using an antisense probe. Arrows indicate the purple coloration resulting from detection of gene transcripts. A: Cotyledon of immature seed; B: Cotyledon of seed at intermediate maturation stage; C: Control with sense probe (cotyledon of immature seed); D: Cotyledon of mature seed. Bars = 50 μm.

Table 1
Specificity of recombinant PcCS and other *N*-methyltransferases to different substrates.

Recombinant enzyme	Substrate/methylation position						References
	1-mX/3*N*	3-mX/1*N*	7-mX/3*N*	pX/3*N*	Tp/7*N*	Tb/1*N*	
Theobromine synthase	Relative activity (%)						
CaMXMT1	ND	ND	**100**	5.0	ND	ND	Uefuji et al. (2003)
CaMXMT2	ND	ND	**100**	5.3	ND	ND	Uefuji et al. (2003)
CTS1	ND	ND	**100**	1.4	ND	ND	Mizuno et al. (2003b)
BTS1	ND	ND	**100**	ND	ND	ND	Yoneyama et al. (2006)
ICS1	ND	ND	**100**	ND	ND	ND	Yoneyama et al. (2006)
PCS1	ND	ND	**100**	11	ND	ND	Yoneyama et al. (2006)
Caffeine synthase							
CCS1	0.53	0.82	24	**100**	ND	25	Mizuno et al. (2003b)
CaDXMT1	ND	ND	1.0	**100**	ND	3.8	Uefuji et al. (2003)
TCS1	5.2	0.4	43.5	**100**	0.0	7.8	Kato et al. (2000)
PcCS	ND	ND	53	ND	ND	**100**	
Product	Tp	Tp	Tb	Cf	Cf	Cf	

Relative activity presented as percentage of the highest activity (100%), indicated in bold print. ND = not detected; 1-mX = 1-methylxanthine (**9**); 3-mX = 3-methylxanthine (**8**); 7-mX = 7-methylxanthine (**6**); pX = paraxanthine; Tp = theophylline (**3**); Tb = theobromine (**2**); Cf = caffeine (**1**). The activity of the recombinant PcCS using theobromine (**2**) as a substrate was 9 fkat mg^{-1} protein. The values represent mean values of three replicates. Enzymes of *Coffea arabica* = CTS1, CCS1, CaDXMT1; *Camellia sinensis* = TCS1; *Camellia irrawadiensis* = ICS1; *Camellia ptilophylla* = PCS1; *Theobroma cacao* = BTS1.

Theobromine (**2**) was also detected as the main alkaloid in *P. cupana* leaves by Weckerle et al. (2003), being higher in young leaves (1.3%) than mature ones (0.03%). On the other hand, in *P. Yoco*, the caffeine (**1**) concentration in the basal portion of stems (0.45%) is about five times higher than in the apical portion, while theobromine (**2**) ranged from 0.048% (apical portion) to 0.006/0.009% (basal portion) (Weckerle et al., 2003). Theobromine (**2**) was the only alkaloid detected in *P. Pachycarpa*, but at very low levels (adult leaves – 0.025%; apical stem – 0.006%, basal stem – 0.001/0.018%) (Weckerle et al., 2003). We have no knowledge of any report on N-methylxanthine concentration in guarana flowers. Coffee (Raju and Gopal, 1979) and tea (Fujimori and Ashihara, 1993) flowers have high levels of caffeine (**1**). Curiously, *Citrus* flowers have caffeine (**1**), but only in parts of the male organs, and theophylline (**3**) also accumulates in minor proportions (Kretschmar and Baumann, 1999).

Similar to flowers and vegetative tissues of guarana, the fruit pericarp also had theobromine (**2**) as the main alkaloid, but depending on the cultivar and the fruit developmental stage, caffeine (**1**) was present in comparable quantities. Caffeine (**1**) was also the main alkaloid in seeds, and at levels much higher than those found for theobromine (**2**) in any other guarana tissue. Several reports have described caffeine concentration in guarana seeds ranging from 2.5% to 6% in the seed and from zero to 1% in the pericarp (Baumann et al., 1995; Maravalhas, 1965; Oliveira, 2010; Pires, 1949; Spoladore et al., 1987). It has been suggested that the prevalence of caffeine (**1**) in guarana seeds is a defense strategy against predators (Mithöfer and Boland, 2012; Weckerle et al., 2003). Baumann et al. (1995) suggested that caffeine (**1**) helps guarana seed dispersion by birds. Some birds swallow the fruit without breaking the seed, which passes through the digestive tract intact without releasing caffeine (**1**), thus avoiding intoxication.

Please cite this article in press as: Schimpl, F.C., et al. Molecular and biochemical characterization of caffeine synthase and purine alkaloid concentration in guarana fruit. Phytochemistry (2014), http://dx.doi.org/10.1016/j.phytochem.2014.04.018

However, the aril, which is free of caffeine (**1**) and rich in sugars, is digested.

Caffeine (**1**) can be degraded via theophylline (**3**) and theobromine (**2**) (Fig. 1). Usually theophylline (**1**) is the main route occurring with high efficiency, which explains its low concentration, for example, in coffee tissues (Mazzafera, 2004; Suzuki and Waller, 1984a,b). Theobromine (**2**) is a precursor and a product of degradation of caffeine (**1**) in most of the species where the metabolism of this alkaloid was studied (Ashihara et al., 2008; Ashihara and Suzuki, 2004; Suzuki and Waller, 1984a,b). A transient accumulation of theobromine (**2**) in young tissues of plants accumulating caffeine (**1**) has been related to a high biosynthesis rate of the latter alkaloid (Ashihara et al., 2008; Ashihara and Suzuki, 2004). However, transient accumulation of theobromine (**2**) in *C. dewevrei* seems to be related to a more significant participation of the theobromine (**2**) degradation route (Mazzafera et al., 1994a).

The significant accumulation of theobromine (**2**) in the vegetative tissues and flowers of guarana might be a consequence of low conversion to caffeine (**1**), as observed in coffee mutants almost free of caffeine (**1**) (Silvarolla et al., 2004). Support for this is the very low expression of *PcCS* in vegetative tissues of guarana (Fig. 8). However, based on previous knowledge of other plants, this is not consistent with the amount of theophylline found in these tissues, which was higher than caffeine (**1**). Unless guarana presents a unique metabolism, in the plants studied so far, most of the theophylline (**3**) was found as a product of degradation of caffeine (Mazzafera, 2004), although the pathway xanthine (**7**) → 3-methylxanthine (**8**) → theophylline (**3**) was found as an alternative in tea seedlings (Deng and Ashihara, 2010). It may be speculated that, coupled with low biosynthesis from theobromine (**2**), caffeine (**1**) is low in these guarana tissues due to a higher degradation rate, but theophylline (**3**) degradation might be slower than in other plants investigated up to now for caffeine metabolism. Only studies using labelled substrates may explain the methylxanthine concentrations and their metabolism in guarana tissues, which was outside the scope of this study.

3.2. Caffeine synthase activity and expression in guarana

Biosynthesis of caffeine (**1**) in plants is mediated by the activity of three *N*-methyltransferases responsible for the methylation of xanthosine (**4**), 7-methylxanthine (**6**) and theobromine (**2**) (3,7-dimethylxanthine). Genes coding for these enzymes have been characterized in some caffeine (**1**)-containing plants, but they were particularly well characterized in coffee (for a review see Ashihara et al. (2011) and Kato and Mizuno (2004)). Three genes are known to code for 3,7-methylxanthine methyltransferase (caffeine synthase – EC 2.1.1.160; CaDXMT1 – AB084125; CCS1 – AB086414; CtCS7 – AB086415). Regarding CtCS7, it is a tentative caffeine synthase because the heterologous protein has very low activity and only against paraxanthine (1,7-dimethylxanthine) but not theobromine (**2**). CaDXMT1 and CCS1 caffeine synthases seem to have unique characteristics as they differ in regard to their affinity for theobromine (**2**) (1200 μM and 157 μM, respectively), and their expression profile, CaDXMT1, is expressed exclusively in immature fruits while CCS1 is expressed in all tissues containing caffeine (Ashihara et al., 2011; Kato and Mizuno, 2004). Curiously, both enzymes have higher affinity for paraxanthine, a methylxanthine found in minute amounts in a coffee cell suspension (Baumann and Frischknecht, 1988; Sartor and Mazzafera, 2000) or in tea leaves (Kato et al., 1996). However, a key difference between these two enzymes is the fact that CaDXMT1 has low activity against 7-methylxanthine (**6**) and theobromine (**2**) (1 and 4%, respectively, considering 100% activity for paraxanthine) while CCS1 has much higher and comparable activities (16–24% for 7-methylxanthine (**6**) and 23–25% for theobromine (**2**)). CCS1 has been suggested to be an enzyme with dual functionality, similar to TCS1 of tea (Kato and Mizuno, 2004).

N-Methyltransferases of methylxanthine-containing plants have greater similarity to each other within the same genus than among different species (Ashihara et al., 2008; Figueiredo et al., 2011; Yoneyama et al., 2006). Thus, theobromine (**2**) and caffeine (**1**) synthases from *Coffea arabica* have more than 80% identity, while tea and coffee caffeine synthases have only 34% identity. The identity of the translated guarana PcCS was greater with theobromine synthases (from 44% to 48%) than with caffeine synthases (less than 35%) identified in other plants (Fig. 7). The contigs assembled here using the EST sequences of the Realgene EST database had 1083 bp and were similar to other assemblings carried out with sequences retrieved from the same database [Supplementary Fig. S1, Figueiredo et al., 2011]. However, as these sequences were not assessed for expression or their heterologous proteins were not produced and assayed for activity, it is not possible to affirm they were caffeine synthases.

The translated amino acid sequence was aligned with theobromine and caffeine (**1**) synthases from other species and typical A, B′, and C motifs and the YFFF region of methyltransferases related to caffeine metabolism were found (Kato and Mizuno, 2004). A, B and C motifs are characteristics of SAM-dependent *O*-methyltransferases (Joshi and Chiang, 1998). The B′ motif and the YFFF region are specific features of the methyltransferase family, including caffeine synthase (Kato and Mizuno, 2004). Salicylic acid carboxyl methyltransferase, benzoic acid carboxyl methyltransferase and jasmonic acid carboxyl methyltransferase are other members of this family (Ishida et al., 2009; Mizuno et al., 2003a,b; Yoneyama et al., 2006). In addition to the presence of these motifs in the PcCS, phylogenetic analysis showed that PcCS has higher identity with methyltransferase sequences from theobromine-accumulating species such as tea and cacao (Fig. 7), although these sequences may code for theobromine (BTS1, ICS1, PCS1 – Yoneyama et al. 2006) or caffeine synthases (TCS1 – Kato et al., 2000 and Yoneyama et al., 2006).

Consistent with the data of methylxanthines concentration in seeds (Fig. 5), the expression of *PcCS* was significantly higher in seeds of immature fruit (Figs. 8 and 9). Complementing this information, attempts to determine caffeine synthase activity were successful only using seed protein extracts from immature fruit (data not shown). In order to prove that the PcCS was a caffeine synthase, a heterologous protein was produced and activity was tested against several substrates, including paraxanthine. In contrast with all other caffeine synthases characterized until now, recombinant PcCS showed theobromine (**2**) as the best substrate (caffeine synthase activity), followed by 7-methylxanthine (**6**) (theobromine synthase activity) (Table 1). No activity was found against any other substrate tested, including paraxanthine, the best substrate identified for other caffeine synthases (Ashihara et al., 2008; Kato and Mizuno, 2004). Theobromine synthase activity of PcCS was approximately 50% of the caffeine synthase activity.

Small changes in the amino acid sequence may partially explain the affinity of PcCS for theobromine instead of paraxanthine. The translated PcCS was aligned with other methyltransferase sequences and the secondary structure of the caffeine synthase de *C. canephora* CcDXMT1 can be used as reference for comparison (Fig. 6). McCarthy and McCarthy (2007) studied the crystallographic structure of caffeine synthase (DXMT) from *Coffea canephora* and, based on the characteristics of the 15 amino acid residues within 5 Å of the theobromine binding site, they suggested that only three have crucial significance for discriminating the substrate specificity of each enzyme, residues 27, 237 and 266. The residues identified in theobromine synthase are Ala27, Pro237 and Phe266 and, in caffeine synthase, they are Phe27, Ser237 and Ile266. It is not clearly understood in which way positions 27

Please cite this article in press as: Schimpl, F.C., et al. Molecular and biochemical characterization of caffeine synthase and purine alkaloid concentration in guarana fruit. Phytochemistry (2014), http://dx.doi.org/10.1016/j.phytochem.2014.04.018

and 237 influence substrate discrimination, while a Phe residue at position 266 would prevent theobromine binding and, therefore, theobromine synthase activity. However, while the 266 residue substitution may be important in coffee, it probably is not in tea since a Ser is the amino acid present in both theobromine (Yoneyama et al., 2006) and caffeine synthase (Kato et al., 2000). In PcCS, the amino acids in these positions are Thr28, Pro238 and Tyr267.

Yoneyama et al. (2006) made site-directed mutagenesis in PCS1 changing the His221 to Arg221 and observed that the recombinant protein, while still keeping 7-methylxanthine (**6**) (theobromine synthase activity) as the best substrate, showed a fivefold increase of activity against paraxanthine (caffeine synthase activity). Although the result is not conclusive, they suggested that the His221 position plays an important role in defining substrate specificity but other amino acid residues might be involved in substrate recognition.

PcCS is particularly different from the other caffeine synthases at the active site, especially in the positions Leu236, Pro237 and Trp238, although it is not possible to conclude how important these changes are with the little knowledge there is on the protein structure of caffeine synthases. In these positions, the caffeine synthase (DXMT) from coffee has Asp236, Ser237 and Met238 (*C. arabica* – CaDXMT) or Ile238 (*C. canephora* – CcDXMT). Thus, if the residue at position 238 is not the same for these coffee species, the residues 236 and 237 might be determinant to define the affinity of PcCS for theobromine instead of paraxanthine. McCarthy and McCarthy (2007) observed that Ser237 in caffeine synthase (DMXT) is mutated to a Pro in theobromine synthase (MXMT), similar to PcCS. This reinforces the indication that it is not possible to draw solid conclusions based on the protein sequences from different species, as well as that more than one amino acid residue defines substrate specificity. Only crystallography structural studies of PcCS and site-directed mutagenesis may explain the unique specificity of PcCS for theobromine but not for paraxanthine.

4. Concluding remarks

Guarana is distinguished from other caffeine-accumulating plants because of high caffeine (**1**) concentration in its seeds whereas, in other parts of the plant, theobromine (**2**) is the major purine alkaloid. This indicates that guarana presents peculiar characteristics regarding the control of caffeine (**1**) biosynthesis in different organs. Like coffee, it seems that the genes coding for methyltransferases of caffeine (**1**) biosynthesis in guarana form a large family. In a search in the guarana EST database, several incomplete methyltransferases sequences with small differences regarding their similarity were found, thereby supporting this conclusion. The affinity of the PcCS isolated in this study, and comparison of its amino acid sequence with other caffeine synthases already reported in the literature, show that an understanding what residues in the active site of the enzyme define the best substrate is still not obtained.

5. Experimental

5.1. Plant material

Guarana plant tissue was collected on the farm Agropecuária Jayoro Ltda, Presidente Figueiredo, Amazonas, Brazil. Fruit (immature, intermediate and mature immediately separated into seeds and pericarps), leaves (young, intermediate and mature), inflorescences (containing floral buds and open flowers, but with no fruit), apical stems (a 20-cm portion from the apex) and basal stems (a 20-cm portion from the base) were collected from adult plants of the cultivars BRS-Amazonas, BRS-Maués, BRS-Luzéia, BRS-CG372, BRS-CG611. Samples were taken from three different plants and immediately after collection they were frozen in dry ice and then, in the laboratory, in a freezer kept at −80 °C. Part of the material was dried at 50 °C for methylxanthines determination.

5.2. Methylxanthine quantification

The material dried at 50 °C was ground to a powder in a mortar with a pestle, and 100 mg was extracted with MeOH–H_2O (80:20 v/v) in a water-bath at 50 °C for 2 h, with occasional stirring. After cooling on the bench, the samples were centrifuged at 15,000*g* for 20 min, and diluted in distilled H_2O at a ratio of 1:10 (v/v; seeds) or 1:2 (v/v; other tissues). Caffeine (**1**), theobromine (**2**) and theophylline (**3**) were quantified by high performance liquid chromatography (HPLC Shimadzu System) using a photodiode array detector operating at 272 nm. A C18 reversed-phase column (25 cm × 4.6 mm × 5 μm; Supelco™) was used to separate the alkaloids using aqueous 0.5% AcOH in H_2O (A) and MeOH (B) as mobile phases, at a flow rate of 1.2 mL min^{-1}. The gradient used was: 0 min 10% B, 18 min 45% B, 19 min 100% B and 20–30 min 10% B. Standard curves for methylxanthine calculation were obtained from known amounts of pure caffeine (**1**), theobromine (**2**) and theophylline (**3**) (Sigma).

5.3. Analysis of EST database, primer design and PcCS sequencing

The EST database Realgene (Ângelo et al., 2008) was used for searching caffeine synthase sequences using the tool BLASTx (Altschul et al., 1990). Caffeine synthase sequences of *Coffea arabica* (CaDXMT1, AB084125; CCS1, AB086414), *Coffea canephora* (CcDXMT, DQ422955) and *Camellia sinensis* (TCS1, AB031280) were used as bait. Among the retrieved sequences, the longest ones with e-values lower than e^{-50} were selected and used to assemble a contig of 1369 bp (CDS = 1083 bp) using the ClustalW tool in the software BioEdit (Thompson et al., 1994). Total RNA was extracted (Rezaian and Krake, 1987) from immature seeds, treated with DNase (Turbo DNA-free, Ambion, Inc.) and the first cDNA strand was synthesized (SuperScript III First-Strand, Invitrogen) and used in RT-PCR reactions. Primers at the 5′UTR and 3′UTR regions and also ones internal to the CDS were designed for amplification (see Fig. 10 for primer sets). The two fragments amplified were separated in 1% agarose gel, purified from the gel (GeneJET™ Gel Extraction Kit, Fermentas) and inserted in the cloning vector pGEM-T Easy Vector System (Promega), which was used to transform the thermo-competent *E. coli* DH5α strain (Novagen). Positive colonies were miniprepped (PureLink Quick Plasmid Miniprep Kit, Invitrogen), and the plasmid insert was sequenced using the flanking vector primers (T7 and SP6) and the internal primers already mentioned. The sequencing reactions were performed using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and an ABI PRISM 377 sequencer (Applied Biosystems). The sequence obtained will hereinafter be referred to as *Paullinia cupana* caffeine synthase (*PcCS* – GenBank accession number BK008796).

5.4. PcCS expression analysis

Total RNA was extracted from different guarana tissues (Rezaian and Krake, 1987) and first strand cDNA was obtained as described above. Expression of *PcCS* was performed by quantitative RT-PCR (qRT-PCR). The reactions (processed in technical triplicates) were prepared with the QuantiFast SYBR Green PCR Kit (Qiagen) and run in an iCycler iQ5 Multicolour Real-Time PCR

Please cite this article in press as: Schimpl, F.C., et al. Molecular and biochemical characterization of caffeine synthase and purine alkaloid concentration in guarana fruit. Phytochemistry (2014), http://dx.doi.org/10.1016/j.phytochem.2014.04.018

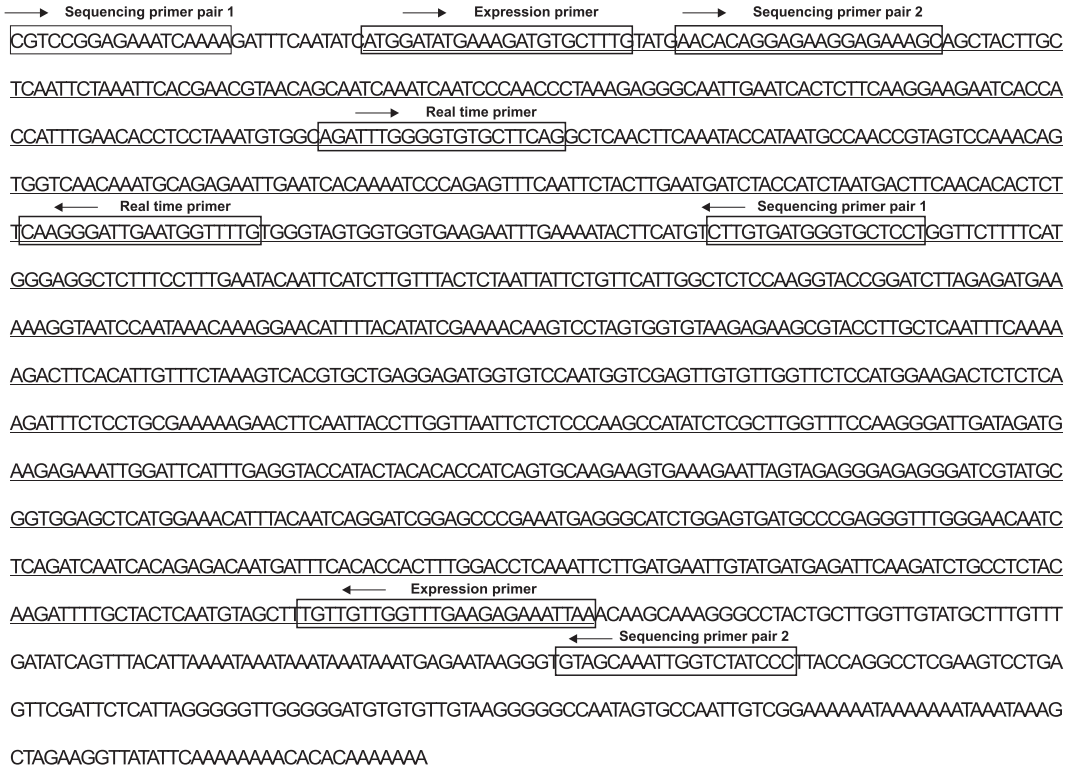

Fig. 10. Contig assembled from guarana EST sequences. Coding region (1083 bp) is underlined from methionine codon (ATG) to the stop codon (TAA). The primers used for cloning and sequencing which generated two fragments by amplification are identified as "Sequencing primers pair" 1 and 2; the primers used in RT-qPCR analyses are identified as "Real time primer"; and the primers used for cloning PcCS for heterologous expression in pET28a are identified as "Expression primer", to which was added the restriction sites for the BamHI (forward) and SalI (reverse) restriction site sequences.

Detection System (Bio-Rad) under the following conditions: 95 °C for 3 min, followed by 40 cycles at 95 °C for 10 s and 60 °C for 30 s. The primers used are shown in Fig. 10 and the $2^{-\Delta Ct}$ method (Livak and Schmittgen, 2001) was used for the quantification of relative expression. Four reference genes were used (Schimpl, 2013): actin (ACT, forward: 5′-ACCATCACCAGAATCCAACACAATACC-3′, reverse: 5′-ACCATCACCAGAATCCAACACAATACC-3′), ubiquitin 11 (UQT, forward: 5′-TAAGGAAGGCATTCCACCTG-3′, reverse: 5′-TAATCC GCTCAATGGT GTCA-3′), Alcohol dehydrogenase (ADH, forward: 5′-GTTGCAGCAT CAGGTCAAGA-3′, reverse: 5′-ACGTCCATTTTCAGC ACACA-3′), and 40S ribosomal protein S24-1 (S24, forward: 5′-TCATCCTGGAAGGCCTAATG-3′, reverse: 5′-CAACCTTGGTGTCAA GTCCA-3′). *PcCS* expression was also analysed by *in situ* hybridization. A probe was produced using the 173 bp fragment amplified with the qRT-PCR primers. All subsequent steps for probe preparation followed the protocol of Araújo et al. (2012). Immature, intermediate and mature fruits were analysed, and the images were captured with an Olympus DP71 video camera connected to an Olympus BX 51 microscope.

5.5. Sequence alignment, phylogenetic analysis and 2D protein structure prediction

Translated amino acid sequences of coffee *N*-methyltransferases (*C. arabica* and *C. canephora*), tea (*Camellia sinensis*, *C. ptilophylla* and *C. irrawadiensis*) and cocoa (*Theobroma cacao*) were aligned with PcCS using the software ClustalW (Thompson et al., 1994). The software MEGA 5 (Tamura et al., 2011) was used to construct a phylogenetic tree using the neighbour-joining algorithm (Saitou and Nei, 1987). Gap regions were excluded by manual adjustments and bootstrap values were obtained after 10,000 repetitions. The prediction of the PcCS 2D structure was obtained with the software ESPript (Gouet et al., 1999) using the 2D structure of the caffeine synthase from *C. canephora* (CcDMXT1, DQ422955; PDB ID: 2EFJ) as reference (McCarthy and McCarthy, 2007).

5.6. Recombinant PcCS

cDNA from immature seeds was used in RT-PCR to amplify the whole *PcCS* sequence (primer sets are shown in Fig. 10), which was initially cloned into the pGEM-T Easy Vector System (Promega) and then into the pET28a (Invitrogen) between the BamHI and SalI sites, containing the recombinant 6X HiS-tag at the N terminus. The quality of the insertion was verified by sequencing. The pET28a-*PcCS* was expressed in the *E. coli* BL21 (pRil) electrocompetent strain. The transformed cells were grown for 4 h in 50 mL of LB medium until the OD_{600} = 0.6 and then induced with 1 mM isopropyl-β-D-thiogalactoside at 37 °C, 200 rpm. The cells were collected by centrifugation (5000*g* at 4 °C) and the bacterial pellet was solubilized in affinity buffer (50 mM sodium phosphate, pH 7.3, 100 mM NaCl, 5% glycerol) and then lysed by sonication (20 cycles of 3 s, with intervals of 5 s, at 20 kHz – Vibra-Cell, SONICS). The presence of the recombinant PcCS in the total extract was verified by 12% SDS–PAGE (Laemmli, 1970). To obtain the purified protein, the starting medium was scaled up to 2 L culture and the bacterial pellet obtained by centrifugation was solubilized in 40 mL of affinity buffer and treated with lysozyme (80 μg mL^{-1}) for 30 min and then sonicated as indicated previously. The sonicated extract was centrifuged (12,000*g* for 15 m at 4 °C) and 1 mL of Ni–NTA agarose resin (Qiagen) previously equilibrated with affinity buffer was added to the supernatant and this was stirred for 30 min. The extract-resin mixture was packed into a plastic column (1.0 cm diameter and 5.0 cm long) and the proteins were eluted by addition of solutions containing increasing concentrations of imidazole

(20, 50, 100, 200 and 500 mM, 2 mL of each solution). All fractions were collected and analysed by 12% SDS–PAGE.

5.7. Plant PcCS activity and recombinant PcCS kinetic study

Plant PcCS was partially purified from immature guarana seeds of the cultivar BRS-Amazonas. The seeds were ground in a mortar with a pestle and extracted (1 g) with 5 mL of phosphate buffer (200 mM, pH 7.3, 4 °C, containing 5 mM EDTA, 10 mM 2-β-mercaptoethanol, 0.5% ascorbic acid, 0.5% Triton X-100 and 4% polyvinylpolypyrrolidone). The supernatant obtained after centrifugation (15 min at 30,000*g*, 4 °C) was saturated to 80% with $(NH_4)_2SO_4$ and pelleted by centrifugation (15 min at 30,000*g*, 4 °C). The pellet was solubilized in phosphate buffer and immediately desalted and concentrated by ultrafiltration (Amicon Ultra-4 Centrifugal Filter Devices, Millipore). Protein concentration was determined using the ready-to-use reagent BioRad (Bradford, 1976).

The determination of PcCS activity was based on transfer of the methyl group from methyl [^{3}H]S-adenosylmethionine to the substrate (Mazzafera et al., 1994b). Reactions were assembled in 1.5 mL Eppendorf tubes containing 4.07 kBq methyl [^{3}H]SAM, 100 μg protein, 10 μL theobromine (**2**) (at six different final concentrations), completed to a final volume of 200 μL with 100 mM Tris–HCl buffer, pH 8.0. Reactions were incubated at 28 °C for 30 min and stopped by adding $CHCl_3$ (1 mL), followed by vortexing for 30 s. After quick centrifugation in a benchtop centrifuge for phase separation, the organic fraction was collected and transferred to scintillation vials, and then evaporated under the air stream of a fume hood. Scintillation fluid was added (5 mL) and the radioactivity incorporated in caffeine was determined in a scintillation counter (Mazzafera et al., 1994b). A reaction without theobromine (**2**) was used as a control. The kinetic parameters K_m and V_{max} were calculated by Lineweaver–Burk curve fitting ($1/V \times 1/[S]$).

To determine PcCS affinity for different substrates, recombinant protein was obtained as described and dialysed against the phosphate buffer (25 mM, pH 7.3, 1 mM EDTA and 1 mM DTT) for 24 h at 4 °C. The enzyme reaction was carried out as previously described but unlabelled SAM was used instead of [methyl-^{3}H]-SAM. The products formed were identified and quantified by UPLC (Ultra Performance Liquid Chromatography) coupled to a mass spectrometry triple quadrupole (Waters-Micromass, Manchester, England) based on their retention times, *m/z* and MS/MS fragmentation. Chromatographic separation was performed using a Waters Acquity BEH C18 (2.1 mm × 50 mm 1.7 mM) column and an 8 min gradient elution of 5% to 100% CH_3CN (solvent B) in aqueous 0.1% HCO_2H (solvent A). The flow rate was 0.200 mL min^{-1} and column temperature was maintained at 30 °C. Electrospray ionization was used in the positive mode under the following conditions: capillary 3.0 kV, cone 50 V, ion source temperature of 150 °C, desolvation temperature of 300 °C. The MS/MS spectra were obtained by induced dissociation collision energy of 20 V. The reactions were performed using 1-methylxanthine (**9**), 3-methylxanthine (**8**), 7-methylxanthine (**6**), paraxanthine, theophylline (**3**) and theobromine (**2**) at 600 μM final concentration. The kinetic parameters K_m and V_{max} were obtained using six different concentrations of 7-methylxanthine (**6**) and theobromine (**2**), and calculated by Lineweaver–Burk curve fitting ($1/V \times 1/[S]$).

Acknowledgments

F.C.S. thanks the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) for a MSc fellowship (2011/03266-6), and JFCG and PM thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for research fellowships.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.phytochem.2014.04.018.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Ângelo, P.C.S., Nunes-Silva, C.G., Brígido, M.M., Azevedo, J.S.N., Assunção, E.N., Sousa, A.R.B., Patrício, F.J.B., Rego, M.M., Peixoto, J.C.C., Oliveira Jr., W.P., Freitas, D.V., Almeida, E.R.P., Viana, A.M.H.A., Souza, A.F.P.N., Andrade, E.V., Acosta, P.O.A., Batista, J.S., Walter, M.E.M.T., Leomil, L., Anjos, D.A.S., Coimbra, R.C.M., Barbosa, M.H.N., Honda, E., Pereira, S.S., Silva, A., Pereira, J.O., Silva, M.L., Marins, M., Holanda, F.J., Abreu, R.M.M., Pando, S.C., Gonçalves, J.F.C., Carvalho, M.L., Leal-Mesquita, E.R.R.B.P., Da Silveira, M.A., Batista, W.C., Atroch, A.L., França, S.C., Porto, J.I.R., Schneider, M.P.C., Astolfi-Filho, S., 2008. Guarana (*Paullinia cupana* var. *sorbilis*), an anciently consumed stimulant from the Amazon rain forest: the seeded-fruit transcriptome. Plant Cell Rep. 27, 117–124.

Araújo, P., Cesarino, I., Carmello-Guerreiro, S.M., Dornelas, M.C., 2012. *Citrus sinensis* L. Osbeck orthologs of FRUITFULL and SHATTERPROOF are differentially expressed during fruit development. Plant Growth Regul. 70, 1–13.

Ashihara, H., 2006. Metabolism of alkaloids in coffee plants. Braz. J. Plant Physiol. 18, 1–8.

Ashihara, H., Suzuki, T., 2004. Distribution and biosynthesis of caffeine in plants. Front. Biosci. 9, 1864–1876.

Ashihara, H., Sano, H., Crozier, A., 2008. Caffeine and related purine alkaloids: biosynthesis, catabolism, function and genetic engineering. Phytochemistry 69, 841–856.

Ashihara, H., Ogita, S., Crozier, A., 2011. Purine alkaloid metabolism. In: Ashihara, H., Crozier, A., Komamine, A. (Eds.), Plant Metabolism and Biotechnology. John Wiley & Sons Ltd, Chichester, pp. 163–189.

Baumann, T., Frischknecht, P., 1988. Caffeine: production by plant (*Coffea* spp.) cell cultures. In: Bajaj, Y. (Ed.), Medicinal and Aromatic Plants I, vol. 4. Springer, Verlag, Berlin, Heidelberg, pp. 264–281.

Baumann, T.W., Schulthess, B.H., Hanni, K., 1995. Guarana (*Paullinia cupana*) rewards seed dispersers without intoxicating them by caffeine. Phytochemistry 39, 1063–1070.

Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein dye binding. Anal. Biochem. 72, 248–254.

Deng, W.-W., Ashihara, H., 2010. Profiles of purine metabolism in leaves and roots of *Camellia sinensis* seedlings. Plant Cell Physiol. 51, 2105–2118.

Figueiredo, L.C., Faria-Campos, A.C., Astolfi-Filho, S., Azevedo, J.L., 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, 1188–1199.

Fujimori, N., Ashihara, H., 1993. Biosynthesis of caffeine in flower buds of *Camellia sinensis*. Ann. Bot. 71, 279–284.

Gouet, P., Courcelle, E., Stuart, D.I., Métoz, F., 1999. ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics 15, 305–308.

Henman, A.R., 1986. Vida Natural, O Guaraná: Sua cultura, propriedades, formas de preparação e uso. Global/Ground, São Paulo.

Ishida, M., Kitao, N., Mizuno, K., Tanikawa, N., Kato, M., 2009. Occurrence of theobromine synthase genes in purine alkaloid-free species of *Camellia* plants. Planta 229, 559–568.

Joshi, C.P., Chiang, V.L., 1998. Conserved sequence motifs in plant S-adenosyl-L-methionine-dependent methyltransferases. Plant Mol. Biol. 37, 663–674.

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

Kato, M., Kaneharu, T., Shimizu, H., Suzuki, T., Gillies, F.M., 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, 629–636.

Kato, M., Mizuno, K., Crozier, A., Fujimura, T., Ashihara, H., 2000. Caffeine synthase gene from tea leaves. Nature 406, 956–957.

Kretschmar, J.A., Baumann, T.W., 1999. Caffeine in *Citrus* flowers. Phytochemistry 52, 19–23.

Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680–685.

Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta}$CT method. Methods 25, 402–408.

Maravalhas, N., 1965. Estudos sobre o guaraná e outras plantas produtoras de cafeína. Instituto Nacional de Pesquisas da Amazônia (INPA), Manaus, p. 25.

Mazzafera, P., 2004. Catabolism of caffeine in plants and microorganisms. Front. Biosci. 9, 1348–1359.

Mazzafera, P., Carvalho, A., 1991. Breeding for low seed caffeine content of coffee (*Coffea* L.) by interspecific hybridization. Euphytica 59, 55–60.

Mazzafera, P., Crozier, A., Magalhães, A.C., 1991. Caffeine metabolism in *Coffea arabica* and other species of coffee. Phytochemistry 30, 3913–3916.

Mazzafera, P., Crozier, A., Sandberg, G., 1994a. Studies on the metabolic control of caffeine turnover in developing endosperms and leaves of *Coffea arabica* and *Coffea dewevrei*. J. Agric. Food Chem. 42, 1423–1427.

Mazzafera, P., Wingsle, G., Olsson, O., Sandberg, G., 1994b. S adenosyl-L-methionine:theobromine 1-*N*-methyltransferase, an enzyme catalyzing the synthesis of caffeine in coffee. Phytochemistry 37, 1577–1584.

McCarthy, A.A., McCarthy, J.G., 2007. The structure of two *N*-methyltransferases from the caffeine biosynthetic pathway. Plant Physiol. 144, 879–889.

Missouri Botanical Garden, 2012. Tropicos—<http://www.tropicos.org/>, (accessed in Feb., 2012).

Mithöfer, A., Boland, W., 2012. Plant defense against herbivores: chemical aspects. Annu. Rev. Plant Biol. 63, 431–450.

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

Mizuno, K., Okuda, A., Kato, M., Yoneyama, N., Tanaka, H., Ashihara, H., Fujimura, T., 2003b. 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.). FEBS Lett. 534, 75–81.

Nagata, T., Sakai, S., 1985. Purine base pattern of *Camellia irrawadiensis*. Phytochemistry 24, 2271–2272.

Oliveira, E.R.N., 2010. Características morfofisiológicas e bioquímicas de clones de guaraná *Paullinia cupana* Kunt. var. *sorbilis* (Mart.) Paullinia cupana Ducke cultivados sob plantio comercial. Botânica. Instituto Nacional de Pesquisas da Amazônia – INPA, Manau.

Pires, J.M., 1949. Guaraná e cupana. Sociedade de Agrônomos e Veterinários do Pará 1, 9–20.

Raju, K.I., Gopal, N.H., 1979. Distribution of caffeine in arabica and robusta coffee plants. J. Coffee Res. 9, 83–90.

Rezaian, M., Krake, L., 1987. Nucleic acid extraction and virus detection in grapevine. J. Virol. Methods 17.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Sartor, R.M., Mazzafera, P., 2000. Caffeine formation by suspension cultures of *Coffea dewevrei*. Braz. Arch. Biol. Technol. 43, 61–69.

Schimpl, F.C., 2013. Teores de metilxantinas e metabolismo de cafeína em frutos de guaraná (*Paullinia cupana* var. *sorbilis* Kunth.). Instituto de Biologia. Universidade Estadual de Campinas, Campinas, p. 72.

Schimpl, F.C., da Silva, J.F., Gonçalves, J.F.d.C., Mazzafera, P., 2013. Guarana: revisiting a highly caffeinated plant from the Amazon. J. Ethnopharmacol. 150, 14–31.

Schmidt, F., 1941. O guaraná, sua cultura e indústria. Serviço de Informação Agrícola, Ministério da Agricultura, Rio de Janeiro.

Senanayake, U., Wijesekera, R., 1971. Theobromine and caffeine content of the cocoa bean during its growth. J. Sci. Food Agric. 22, 262–263.

Silvarolla, M.B., Mazzafera, P., Fazuoli, L.C., 2004. Plant biochemistry: a naturally decaffeinated arabica coffee. Nature 429, 826.

Spoladore, D.S.B., Marco Antonio Milan, Sáes, Luis Alberto, 1987. Teor de cafeína em sementes matrizes do guaranazeiro. Bragantia 46, 425–429.

Suzuki, T., 1972. Participation of S-adenozylmethionine in biosynthesis of caffeine in tea plant. FEBS Lett. 24, 18–20.

Suzuki, T., Waller, G.R., 1984a. Biodegradation of caffeine – Formation of theophylline and theobromine from caffeine in mature *Coffea arabica* fruits. J. Sci. Food Agric. 35, 66–70.

Suzuki, T., Waller, G.R., 1984b. Biosynthesis and biodegradation of caffeine, theobromine, and theophylline in *Coffea arabica* L. fruits. J. Agric. Food Chem. 32, 845–848.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

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, 372–380.

Weckerle, C.S., Stutz, M.A., Baumann, T.W., 2003. Purine alkaloids in *Paullinia*. Phytochemistry 64, 735–742.

Ye, C., Lin, Y., Zhou, H., Cheng, F., Li, X., 1997. Isolation and analysis of purine alkaloids from *Camellia ptilophylla* Chang. Acta Sci. Nat. Univ. Sunyatseni 36, 30–33.

Yoneyama, N., Morimoto, H., Ye, C.X., 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, 125–135.

Please cite this article in press as: Schimpl, F.C., et al. Molecular and biochemical characterization of caffeine synthase and purine alkaloid concentration in guarana fruit. Phytochemistry (2014), http://dx.doi.org/10.1016/j.phytochem.2014.04.018