Targeted metabolomics and transcript profiling of methyltransferases in three coffee species

Andrea Montis[a,b,*], Cédric Delporte[a,b], Yusaku Noda[c], Piet Stoffelen[d], Caroline Stévigny[a], Christian Hermans[e], Pierre Van Antwerpen[a,b], Florence Souard[f,g]

[a] RD3 Unit of Pharmacognosy, Bioanalysis and Drug Discovery, Faculty of Pharmacy, Université libre de Bruxelles, Campus Plaine, CP 205NA09, 1050 Brussels, Belgium

[b] APFP Analytical platform of the Faculty of Pharmacy, Faculty of Pharmacy, Université libre de Bruxelles, Campus Plaine, CP 205NA5, 1050 Brussels, Belgium

[c] Current address: National Institutes for Quantum Science and Technology (QST), Takasaki Institute of Advanced Quantum Science, Gunma, 370-1292, Japan

[d] Meise Botanic Garden, Domein van Bouchout, Nieuwe laan 38, 1860 Meise, Belgium

[e] Crop Production and Biostimulation Laboratory, Brussels Bioengineering School, Université libre de Bruxelles, Campus Plaine, CP 245, 1050 Brussels, Belgium

[f] Département de Pharmacochimie Moléculaire, UMR 5063 CNRS, Université Grenoble Alpes, 470 rue de la chimie, 38400 Saint-Martin d'Hères, France

[g] DPP Department – Unit of Pharmacology, Pharmacotherapy and Pharmaceutical care, Faculty of Pharmacy, Université libre de Bruxelles, Campus Plaine, CP 205NA5, 1050 Brussels, Belgium

* Corresponding author at: RD3 Unit of Pharmacognosy, Bioanalysis and Drug Discovery, Faculty of Pharmacy, Université libre de Bruxelles, Campus Plaine, CP 205NA09, 1050 Brussels, Belgium.

E-mail address: florence.souard@ulb.be

ABSTRACT

Xanthines are the most known alkaloids found in coffee plants. Three *Coffea* species grown in a controlled greenhouse environment were the focus of this research. *Coffea arabica* and *C. canephora* are two first principal commercial species and commonly known as arabica and robusta, respectively. *C. anthonyi* is a new small leaved species from Central Africa. The xanthine metabolites in flower, fruit and leaf extracts were compared using both targeted and untargeted metabolomics approaches. We evaluated how the xanthine derivatives and FQA isomers link to the expression of biosynthetic genes encoding *N*- and *O*-methyltransferases. Theobromine built up in leaves of *C. anthonyi* because caffeine biosynthesis was hindered in the absence of synthase gene expression. Green fruits nevertheless managed to express these genes and they produce caffeine. Given that *C. anthonyi* evolved successfully over time, these findings put into question the defensive role of caffeine in leaves. An overview of the histolocalisation of xanthines in the different flower parts of *Coffea arabica* was also provided. The gynoecium contained more theobromine than the flower buds or petals. This could be attributed to increased caffeine biosynthesis before fructification. The presence of theophylline and the absence of theobromine in the petals indicate that caffeine is catabolized more in the petals than in the gynoecium.

Keywords

Coffea

Xanthines

N-methyltranferase

LC-(HR)MS

Metabolomics

Transcriptomics

1. Introduction

Many plant species contain alkaloids, a large family of secondary metabolites with a nitrogen-containing heterocyclic ring (Dey et al., 2020). Caffeine (1,2,3-trimethylxanthine) is a purine alkaloid that plays a role in plant protection. It accounts for one or two percents dry weight of coffee (Ashihara et al., 2008). That molecule is synthesized from xanthosine in a four-step pathway involving three methylation reactions and one ribose hydrolysis (**Fig. 1**). The methylation steps are sequentially catalysed by the xanthosine 7-methyltransferase (XMT), 7-methylxanthine methyltransferase/theobromine synthase (MXMT) and 3,7-dimethylxanthine methyltransferase/caffeine synthase (DXMT). The coenzyme involved in these methyl group transfers is the S-Adenosyl Methionine (SAM). In coffee plants, different isoforms with dual functions are characterized : MXMT1 and MXMT2/CTS2 are essentially involved in theobromine biosynthesis, while DXMT1 and DXMT2/CCS1 in both theobromine and caffeine biosynthesis (Mizuno et al., 2003; Uefuji et al., 2003) (**Fig. 1**).

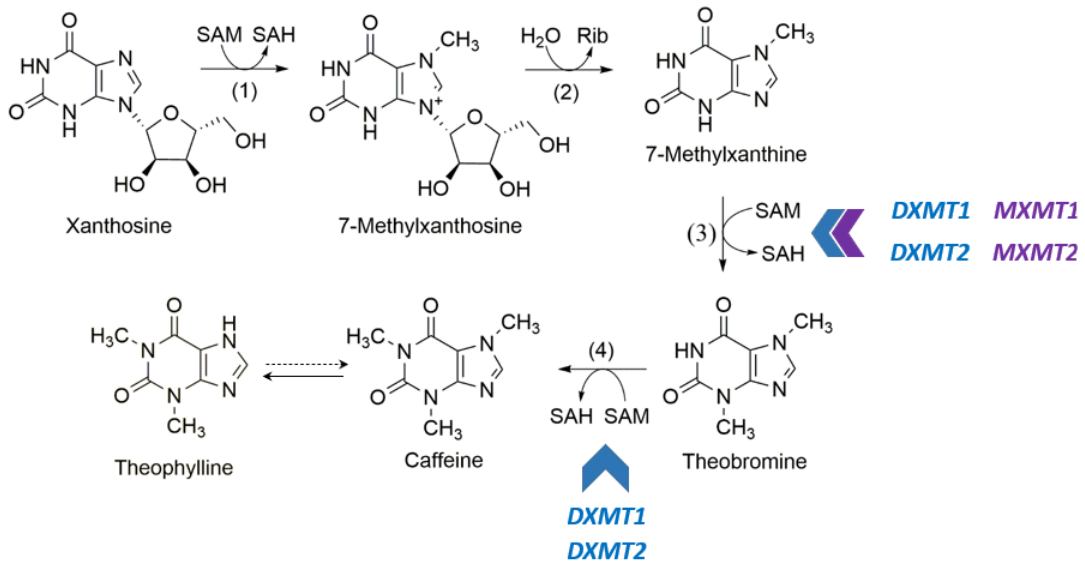

Fig 1. The major biosynthetic pathway of caffeine in coffee plants. The pathway shows metabolites and genes encoding enzymes investigated in this work. (1) 7-methylxanthosine synthase; (2) *N*-methylnucleosidase; (3) theobromine synthase (enzymatic activity exerted by DXMT1, DXMT2, MXMT1 and MXMT2); (4) caffeine synthase (enzymatic activity exerted by DXMT1 and DXMT2). S-adenosyl-L-methionine (SAM) is converted into S-adenosyl-L-homocysteine (SAH) during steps catalysed by *N*-methyltransferases. The ribose unit (Rib) is lost upon conversion of 7-Methylxanthosine into 7-Methylxanthine. Adapted from (Montis et al., 2021).

Caffeine is primarily synthesized in the young leaves and developing endosperms of fruits (Zheng and Ashihara, 2004). Every tissue expresses *DXMT2* and *MXMT2*, while flower buds lowly express *DXMT2* (Mizuno et al., 2003). In *Coffea arabica* (Arabica), transcripts of *DXMT1* are abundant during fruit development, while those of *MXMT1* low (Perrois et al., 2015). Mature leaves contain

lower *DXMT1* transcripts compared to younger ones. In *C. canephora* (Robusta), the expression of the same genes is showing a different pattern (Perrois et al., 2015). For instance, *MXMT1* is not expressed in young leaves while *DXMT1/2* is more in that species.

Caffeine is catabolised into theophylline in mature tissues, after being synthesized in young beans and leaves. The catabolic pathway involves a multistep reaction wherein theophylline is produced after one demethylation (Fig. 1). That catabolite loses the two remaining methyl groups and is subsequently converted into 3-methylxanthine and xanthine (Ashihara et al., 2008). An alternative catabolic pathway has been proposed in certain *Coffea* species, including *abeokutae*, *dewevrei* and *liberica*. The mature leaves of these species lose caffeine in favour of theacrine (4-methylaxanthine), which is then converted into liberine (Petermann and Baumann, 1983).

Most studies on the expression of genes involved in caffeine biosynthesis have been conducted in *C. arabica* and *C. canephora*. Indeed, these two major species are commercially cultivated for beverage production (Montis et al., 2021). Other species have received less attention. This is the case for *C. anthonyi*, an endemic to Cameroon, Congo and Gabon. It closely relates to *C. eugenioides*, the maternal progenitor of *C. arabica* (Stoffelen et al., 2009), while *C. canephora* is the paternal progenitor of the latter one (Bawin et al., 2021). Furthermore it grows in lower altitude and could be an alternative to the two other species in the context of the climate change and temperature increase.

In addition, the bitterness of coffee is attributed to the breakdown of chlorogenic acids (CGAs) into phenolic compounds during roasting. The polyphenolic fraction of coffee tissues is primarily composed of caffeoylquinic acids (CQAs), feruloylquinic acids (FQAs) and coumaroylquinic acids (3-CoQA) in monomeric or dimeric forms (Chen, 2019). As far as the quinic acid derivatives are concerned, the caffeoyl-coenzyme A 3-O-methyltransferase (CCoAOMT) is involved in FQAs production through feruloyl-CoA (Clifford et al., 2017). In this context, it is generally admitted that genes encoding *N*- and *O*-methyltransferases are related (Li et al., 2017).

We previously compared the metabolomes of *C. arabica*, *canephora* and *anthonyi* (Montis et al., 2022). The fruits of *C. anthonyi* contain caffeine, albeit in lower concentration than those of *C. arabica*. This is surprising because caffeine is absent from leaves and phloem sap in that genotype, unlike the two other ones. Our assumption is a possible disruption of the caffeine biosynthetic pathway in leaves. Besides, *C. anthonyi* contains elevated FQA levels, and turns to be a valuable coffee species for improving human health. The purpose of this study on these three *Coffea* species, is to quantify the xanthine derivatives and FQA isomers in fruits and leaves, and to correlate these observations with the expression of genes encoding *N*- and *O*-methyltransferases. In addition, we

want to clarify whether the accumulation of caffeine in the fruits of *C. arabica* was primarily due to phloem translocation from leaves or to efficient biosynthetic activity in different flower parts.

2. Materials and methods

2.1 Chemicals and reagents

Ultra-high purity water was obtained by filtration using a Milli-Q system from Millipore (Bedford, MA, USA). MS quality acetonitrile, formic acid (FA) and trifluoroacetic acid (TFA) were purchased from Fisher Scientific, while caffeine (> 99 %), theobromine (> 98 %) and theophylline (> 99 %) were purchased from Sigma-Aldrich (Steinheim, Germany).

2.2 Plant material

The coffee trees were grown in the tropical greenhouses of the Meise Botanic Garden (Meise, Belgium) with the same environmental conditions: natural daylight, temperature > 20 °C and relative humidity. The codes of coffee trees are archived, and vouchers deposited in the Herbarium of the Garden. All samples were harvested in the morning between 9 am and 12 am. Some 30 samples of leaves from *C. anthonyi*, *C. arabica* and *C. canephora* were collected. We distinguished between two developmental stages: L1 consisted of young leaves, and L2 of mature leaves. The first ones were recently expanded with light green shade (Salgado et al., 2008) and they measured respectively 2 cm wide and 3 cm long in *C. arabica* and *C. anthonyi*, and 6 cm wide and 10 cm long in *C. canephora*. The second ones were fully developed without marginal necrosis (Souard et al., 2018), and they measured respectively 6 cm wide and 12 cm long, and 10 cm wide and 18 cm long. Ten samples of fruits were picked from *C. anthonyi* and *C. arabica*. Flower parts were also collected from *C. arabica* before fructification: six buds from one plant, 18 gynoecium and 30 petals from five different plants. For metabolomic analysis, only the mature leaves (L2) and the ripened fruits (RF) were considered. For RNA extraction and expression study, the same organs, the young leaves (L1) and the green fruits (GF) were harvested.

2.3 Sample preparation and storage conditions

Samples for metabolomic analysis were dried after collection in sealed plastic bags filled with silica gel, which was replaced daily for seven days. Then, the leaves and flower petals were powdered with an ULTRA-TURRAX® Tube Drive control (Q-Lab, Vilvoorde, Belgium) during 10 min at 6 10^3 rotations min^{-1}. The gynoecia and flower buds were first grinded with a mortar and pestle, and submitted the same treatment during 20 min. The extraction consisted of suspending 15 mg of

homogenized sample powder in 1.5 mL of milliQ water, and placing for 5 min in 55 kHz ultrasonic bath (Souard et al., 2018). Three to eight extraction replicates were obtained. Samples were filtered through a cellulose acetate membrane (0.2 µm) and stored at −20 °C until proceeding to analysis. Fresh plant tissues for RNA extraction were flash-frozen liquid nitrogen and stored at −80 °C.

2.4 LC-HRMS analysis

Semi-polar metabolite fingerprints were monitored using a 1200 series rapid resolution LC (RRLC) system coupled to a 6520 series electrospray ionization (ESI)- quadrupole time-of-flight (QTOF) high-resolution mass spectrometer (HRMS) from Agilent Technologies (Waldbronn, Germany). A Poroshell 120 EC-C18 column (2.7 µm, 100 mm × 2.1 mm) from the same supplier was used to carry out the chromatographic separation. The column temperature was set to 55.0 °C. The mobile phases were composed of 0.025% of TFA and 0.075% of FA in water (solvent A), and in acetonitrile (solvent B) for positive ion mode analysis. The same LC-MS conditions were applied to the negative ionisation mode, by slightly changing the composition of the solvents. Solvent A was 20 mM ammonium formate pH 5.5 and solvent B acetonitrile. The flow rate was set to 0.5 mL min^{-1} and the applied gradient was: 0 min, 0 % B; 0–8 min, 0–10 % B; 8–9 min, 10–12.5 % B; 9–11 min, 12.5–15 % B; 11–17 min, 15–80 % B; 17–18 min, 80–100 % B; 18–19 min, 100 % B; 19–20 min, 100–0 % B; post-run 8 min at 0.5 mL min^{-1}. ESI-QTOF parameters are detailed in **supplemental data 1**.

2.5 Quantification of caffeine, theobromine and theophylline

Mass Hunter Quantitative Analysis software version B.10 (Agilent Technologies) was used to determine the concentration of caffeine, theobromine and theophylline in the extracts. Caffeine concentration was determined by interpolation of caffeine calibration curve ($y = -0.01141\ x^2 + 606.5\ x + 826739.8$; $R^2 = 0.9983$) drawn from 0.001 to 200 µg mL^{-1} using a caffeine standard. Theobromine and theophylline concentrations were determined by referring to calibration curves ($y = 1715.437707\ x + 0.708100$; $R^2 = 0.9933$; and $y = 2448.170265\ x + 0.731876$; $R^2 = 0.9945$) drawn from 0.001 to 20 µg mL^{-1}.

2.6 Data processing

ProteoWizard MSConvert tools (Version 3.03.9393, 64-bit) was used to convert raw data (.d Agilent ones) to .mzXML file. Peak Picking was chosen as a filter option. All data processing was performed on the Workflow4Metabolomics infrastructure (https://workflow4metabolomics.org) (accessed on 15/04/22). Detailed steps and parameters that were used for data processing are publicly available on the W4M workflow repository (https://usegalaxy.fr/u/amontis/h/coffee-published-2022) (accessed on 28/11/23) and as **supplemental data 1**. Each metabolite submitted to statistical analysis were called

according to their *m/Z* ratio (M) and retention time in sec (T); for example, M195T507 stands for caffeine.

Principle Component Analysis (PCA) and loadings plot were used to illustrate clustering and markers of this clustering. Biosigner algorithm (Rinaudo et al., 2016) was used in workflow4metabolomics/galaxy platform to confirm the markers of each cluster considering three binary classifiers: RF (random forest), SVM (Support vector machine) and PLS-DA (Partial least squares-discriminant analysis).

The annotation process involved first carefully reviewing the information about the different *m/z* values in the available literature, and then by computing the ion elemental formulas based on the *m/z* values with MassHunter Qualitative Analysis® (Version B.10) from Agilent Technologies. Standard samples were also injected with the same LC conditions to confirm the identity of the most significant metabolites. The workflow showing sample preparation from harvest to LC-MS analysis is shown in **supplemental data 2**.

2.7 RNA extraction, reverse transcription and gene expression

Total RNA was extracted from plant tissues with the Maxwell RSC Plant RNA Kit (Promega Benelux, the Netherlands) and reverse transcription was done with the GoScript Reverse cDNA Synthesis Kit (Promega). Quantitative PCR was performed using the GoTaq qPCR master mix (Promega) and the PicoReal Real-time PCR System (Thermo Fisher Scientific, Waltham, MA, USA). The thermal cycling conditions were as follows: 95 °C for 2 min, 40 cycles of 95 °C for 15 s, 60 °C for 1 min. The efficiency method, which compares the PCR efficiency of the reference genes with that of the target gene (Pfaffl, 2001), and the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001) were used to calculate the relative expression levels. The primers are listed in **supplemental data 3**. The workflow showing sample preparation for RT-qPCR experiments is shown in **supplemental data 4**.

3. Results

3.1 Quantification of caffeine, theobromine and theophylline in leaves and fruits

Caffeine, theobromine and theophylline were measured in the mature leaves (L2) and ripened fruits (RF) (**Tables 1 & 2**).

Table 1. Concentration of caffeine, theobromine and theophylline in mature leaves (L2) of *Coffea anthonyi*, *C. arabica* and *C. canephora*. Values are expressed as mg g^{-1} dry matter (mean ± SD, n = 3), N.D. = not detected.

	C. anthonyi	***C. arabica***	***C. canephora***
Caffeine	N.D.	7.3 ± 0.3	5.1 ± 0.5
Theobromine	0.18 ± 0.02	0.004 ± 0.001	0.02 ± 0.01
Theophylline	N.D.	N.D.	0.13 ± 0.02

Caffeine was detected in *C. arabica* and *C. canephora* but not in *C. anthonyi* leaves. Theobromine concentration in leaves ranged from lowest in *C. arabica* to greatest in *C. anthonyi*. Theophylline was only found in *C. anthonyi* leaves (**Table 1**).

Table 2. Concentration of caffeine, theobromine and theophylline in ripened fruits (RF) of *Coffea anthonyi* and *C. arabica*. Values are expressed as mg g^{-1} dry matter (mean ± SD, n = 3), N.D. = not detected.

	C. anthonyi	***C. arabica***
Caffeine	2.0 ± 0.2	2.7 ± 0.4
Theobromine	0.11 ± 0.01	0.04 ± 0.01
Theophylline	N.D.	N.D.

Caffeine concentration was greater in *C. arabica* than in *C. anthonyi* fruits (**Table 2**). Theobromine followed opposite pattern and theophylline was undetected in both species. Quantification was not done in *C. canephora* fruits because that auto-sterile species did not fructify in the greenhouse due to the absence of pollinators.

3.2 Expression of N-METHYLTRANSFERASES

As caffeine was detected in fruits but not in leaves of *C. anthonyi* (Montis et al., 2022), we looked at the relationship between levels of xanthines and transcripts of four genes (*MXMT1*, *MXMT2*, *DXMT1* and *DXMT2*) involved in caffeine biosynthesis. The survey was conducted in young (L1) and L2 leaves, as well as in green (GF) and red (RF) fruits. The *MXMT1* and *MXMT2* genes encoding the enzymes responsible for conversion of 7-methylxanthine into theobromine, were expressed abundantly in L1, lowly in L2 leaves and GF, and not at all in RF of *C. anthonyi* and *C. canephora*

(**Fig. 2 A, B**). In comparison with the other species, the transcript levels of *MXMT1* were undetected and the ones of *MXMT2* much more abundant in *C. canephora* leaves.

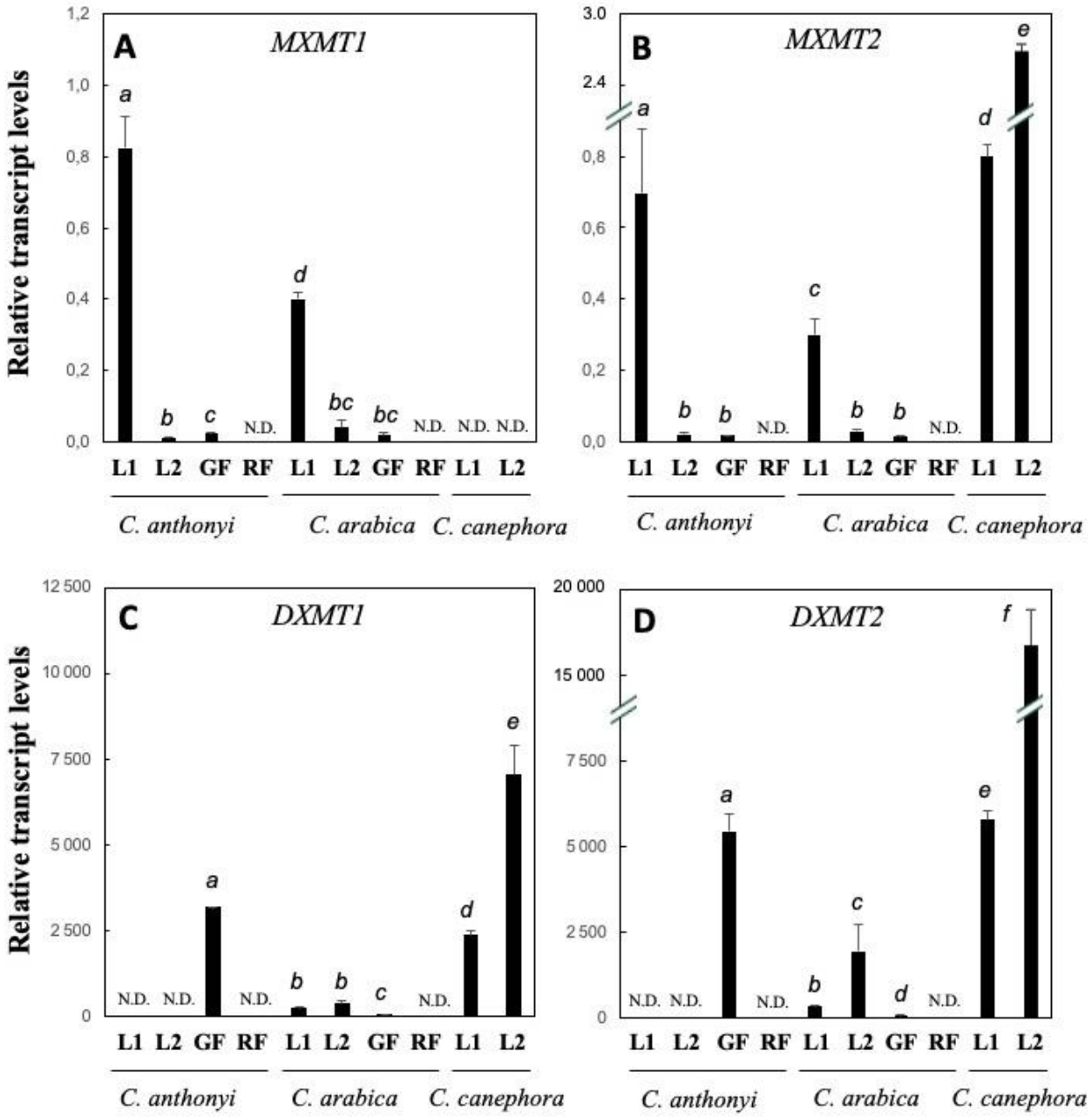

Fig. 2. Expression of genes involved in caffeine biosynthesis. Relative transcript levels of *MXMT1* (**A**), *MXMT2* (**B**), *DXMT1* (**C**) and *DXMT2* (**D**) in young (L1) and mature (L2) leaves of *Coffea anthonyi*, *arabica* and *canephora*, as well as in green (GF) and red fruits (RF) of *C. anthonyi* and C. *arabica*. N.D. = no data or values > 10 times lower than the lowest displayed value. *ACTIN*, *UBQ* and *GAPDH* were used as an internal reference genes. N = 3 technical replica ± std. Letters indicate statistical differences ($p < 0.05$).

The *DXMT1* and *DXMT2* genes encoding enzymes converting 7-methylxanthine into theobromine, and also theobromine into caffeine, were sequentially more exprrssed in *C. anthonyi* leaves, *C. anthonyi* green fruits and *C. arabica* leaves (**Fig 2 C,D**).

Interestingly, C. *anthonyi* green fruits have greater transcript levels of *DXMT1* and *DXMT2* compared to other organs. This suggests that these genes have an equal contribution to caffeine biosynthesis in fruits of that species. Next, in both mature and young leaves of *C. arabica* and *C. canephora, DXMT2* was more expressed than *DXMT1*. This indicates *DXMT2* could contribute more to caffeine biosynthesis in leaves.

3.3 Quantification of feruloylquinic acids in leaves

The FQAs were quantified in mature leaves (L2) with a targeted metabolomic approach in a negative ionization mode. Three ions with *m/z* = 367.1038, 367.1043 and 367.1044 were detected (**Fig. 3**). Based on the isotopic pattern and a literature survey, the formula $C_{17}H_{19}O_9$ was assigned to these three features. The calculated Δppm was about 2. As the standard compounds were not available for these three isomers, we performed a univariate Kruskal-Wallis test to detect the differences between the three *Coffea* species, based on the different metabolite intensities in the samples. The elution order for the three FQAs has been established according to the elution order of the three CQAs from which these compounds are originated from. Indeed, the 5-CQA eluted earlier than 4-CQA but later than 3-CQA (Montis et al., 2022) (**supplemental data 2**). Based on this assumption, the 5-FQA eluted earlier than 4-FQA but later than the 3-FQA. Given that 5-CQA is the most prevalent CQA in coffee leaves, that isomer should presumably be the one contributing most to the occurrence of FQAs.

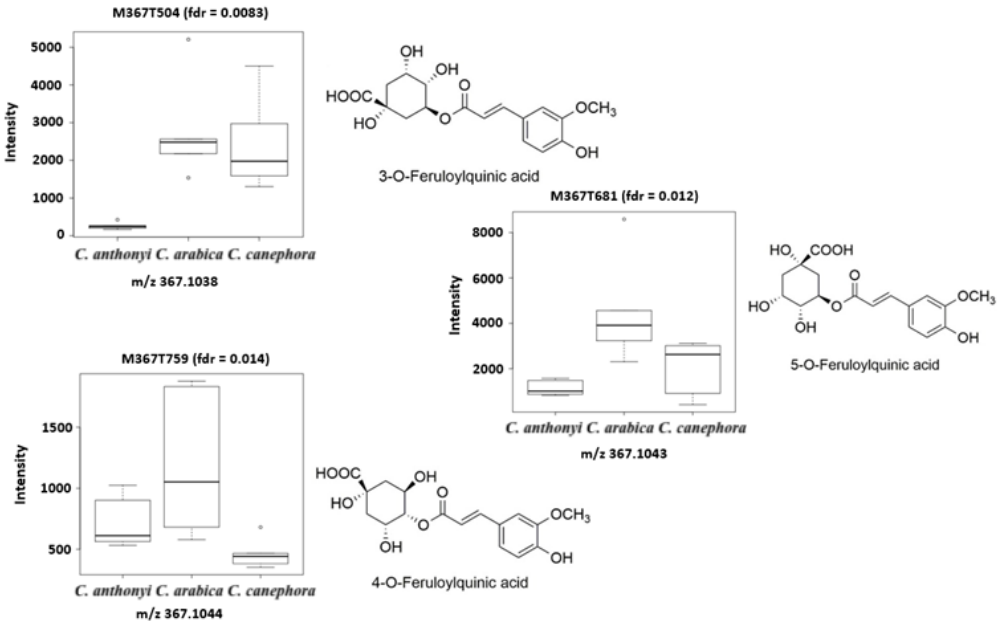

Fig. 3. Quantification of feruloylquinic acid isomers in leaves. Univariate analysis on the ion abundance (intensity) (Kruskal Walis test) showing the levels of the three isomers 3-, 4- and 5-O-Ferulolylquinic acids, in the leaves (L2) of *C. anthonyi*, *C. arabica* and *C. canephora*. N = 6 technical replicates, fdr = false discovery rate.

The greatest levels of 3- and 5-FQA were detected in *C. arabica* and *C. canephora* leaves, while the lowest ones in *C. anthonyi* (**Fig. 3**). Both isomers were present at nearly the same level in the two

first species. The level of 4-FQA was the greatest in *C. arabica* leaves. The fact that these three FQA isomers do not occur in the same ways must also be taken into account.

3.4 Expression of O-METHYLTRANSFERASE

Eventually, the expression of the gene encoding a selective *O*-methyltransferase involved in the conversion of the CQA into FQA was investigated. The transcript levels of *CCoAOMT* were measured in L1 and L2 of the three species (**Fig. 4**).

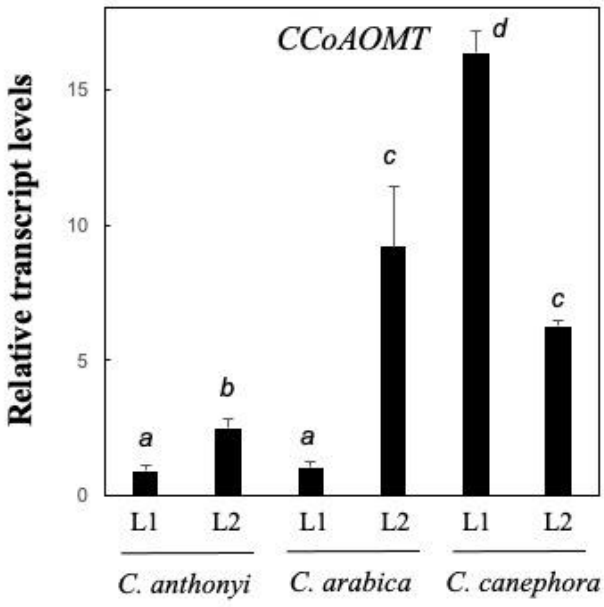

Fig. 4. Expression of genes involved in FQA biosynthesis. Relative transcript levels of *CCoAOMT* in young (L1) and mature (L2) leaves of *Coffea anthonyi*, *arabica* and *canephora*. *ACTIN*, *UBQ* and *GAPDH* were used as internal reference genes. N = 3 technical replica ± std. Letters indicate statistical differences ($p < 0.05$).

In L1 leaves, the *CCoAOMT* transcript levels of both *C. anthonyi* and *C. arabica* were the lowest ones, and those of *C. canephora* the greatest ones. In L2 leaves, the transcript levels of *C. anthonyi*, *C. canephora* and *C. arabica* were sequentially increasing.

3.4 Coffee flower metabolomics

An untargeted metabolomic approach was used to detect the main xanthine derivates present in flower buds, as well as in the gynoecium and petals of *C. arabica*. To have a clear overview of the histolocalisation of the main annotated xanthines, one PCA was first performed. The purpose was to highlight differences in the chemical composition between the three flower parts. A total of 1,345 variables/metabolites were detected. The PCA score plots showed a fair separation of the samples, with PC1 accounting for 59 % of the total variation and PC2 for 22 % (**Fig. 5A**).

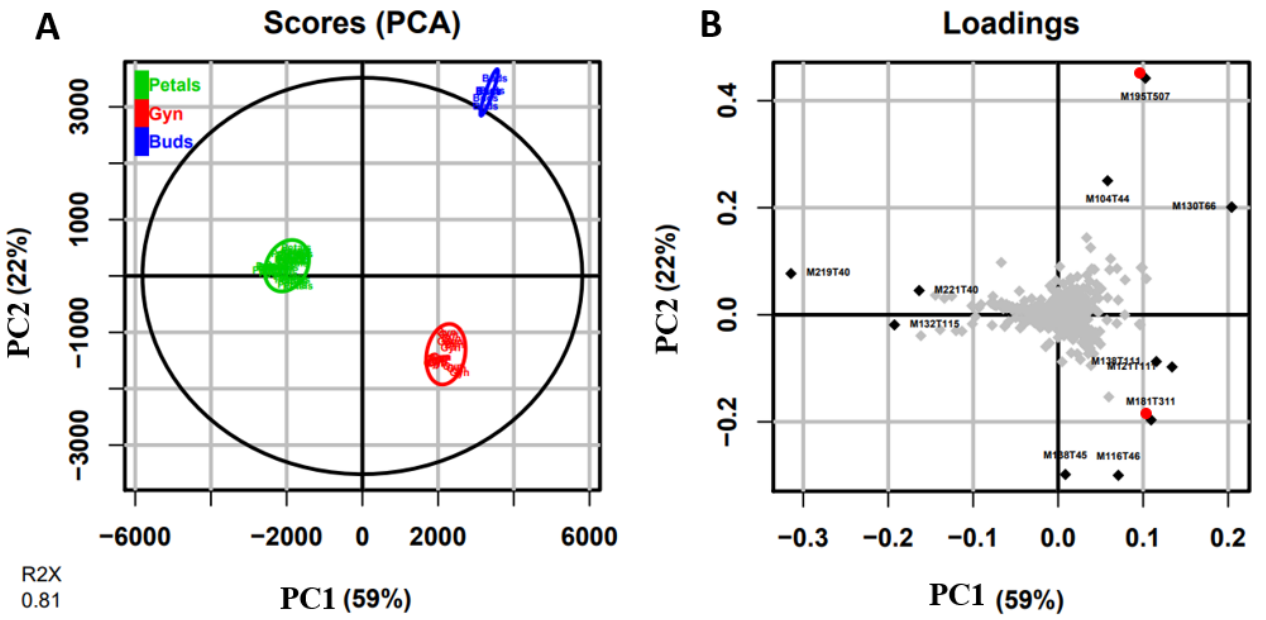

Fig 5. Principal component analysis of metabolome in flower parts of *Coffea arabica*. The PCA score plot (A) is showing the separation for the three different flower samples (flower bud, gynoecium and petals) analysed in ESI-MS (+) mode, and the loading plot (B) the main metabolites accountable for the separation. The first two principal components PC1 and PC2 show cumulative variance of 59 % and 22 %, respectively. Red points indicate M195T507 (annotated as caffeine) and M181T311 (theobromine).

The loading plot showed features responsible for the clustering (**Fig. 5B**). Most of these showed rather low retention time values (> 100 s). However, features M195T507 and M181T311 drew our attention, and were annotated as caffeine and theobromine, respectively. Since theobromine and theophylline are isomers of the same molecule, the identification of these compounds was performed by injecting both standard compounds and by checking their retention time. Interestingly, loadings nominates caffeine as a flower bud marker and theobromine as a gynoecium marker. However, theophylline was not pointed as a marker.

Caffeoylquinic acids were not detected in flowers, although previously found in fruits and leaves (Montis et al., 2022). The Biosigner results confirmed that caffeine and theobromine were the main discriminant metabolites across the samples. Surprisingly, theophylline appeared as a petal marker (**Fig. 6**). As a validation, these three xanthines were simultaneously quantified in the various flower parts (**Table 3**).

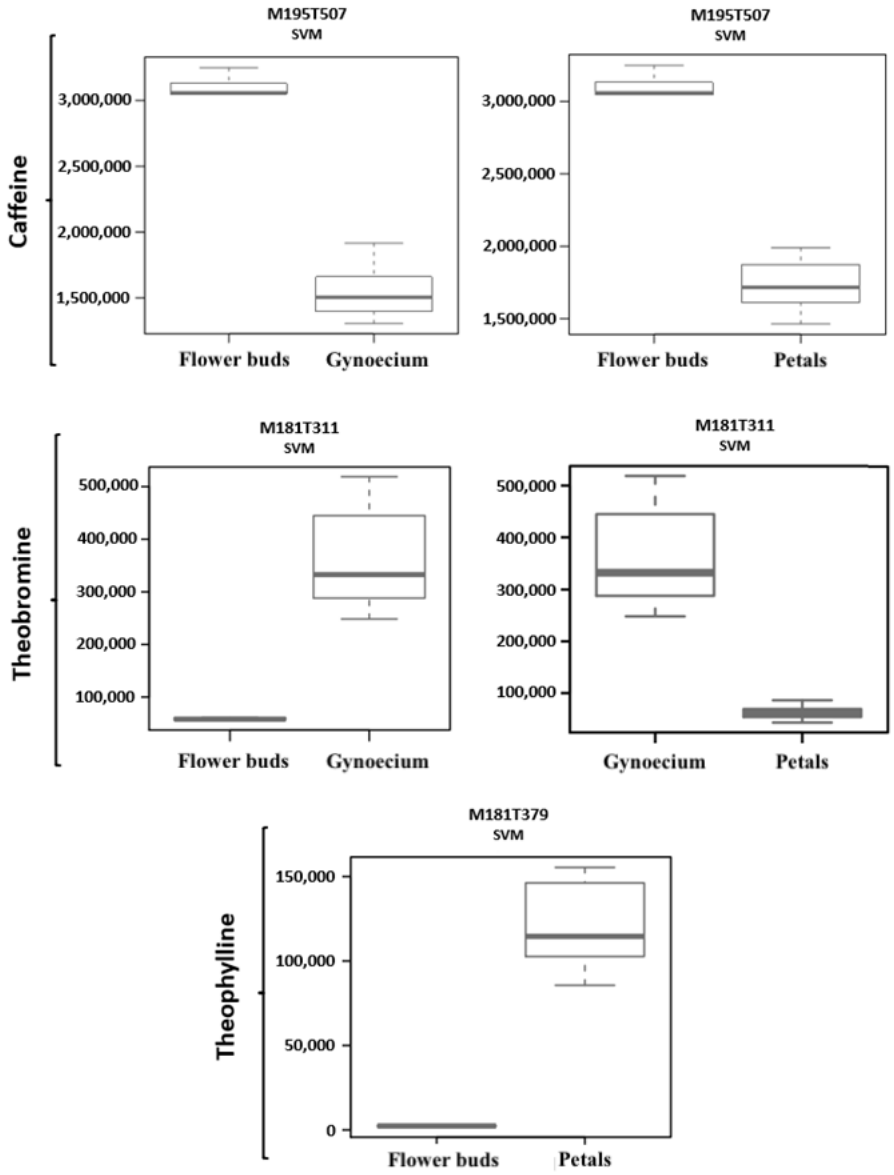

Fig. 6. Statistical comparison of the metabolite abundance illustrated by Biosigner. Boxplots show the normalised intensities of the main metabolites M195T507 (caffeine), M181T311 (theobromine) and M181T379 (theophylline) detected in the flower buds, gynoecium and petals. The boxplot illustrated here were obtained by Support Vector Machines (SVM).

Biosigner pointed that the main marker in the flower buds was caffeine (**Fig. 6**). The quantification showed caffeine in the flower buds was twice as more concentrated as in both the gynoecium and the petals (**Table 3**). Theobromine was returned as a gynoecium indicator (**Fig. 6**). Definitely, the gynoecium contained three times more theobromine than the other flower parts (**Table 3**). Finally, theophylline was returned as a petal marker (**Fig. 6**). That metabolite was largely detected in the petals, but undetectable in the flower bud and the gynoecium (**Table 3**).

Table 3. Concentration of caffeine, theobromine and theophylline in *C. arabica* flowers. Results are expressed as mg g^{-1} dry matter (mean ± SD, n = 5) N.D. = not detected

	Flower buds	*Gynoecium*	*Petals*
Caffeine	8.6 ± 0.6	4.01 ± 0.07	4.1 ± 0.1
Theobromine	0.07 ± 0.01	0.9 ± 0.2	0.07 ± 0.01
Theophylline	N.D.	N.D.	0.164 ± 0.004

4. Discussion

4.1 Quantification of xanthines associated with N-methyltransferases expression in leaves

The relationship between the metabolites and the transcript levels of genes involved in metabolites biosynthesis (summarized in **Fig. 7**) is first discussed.

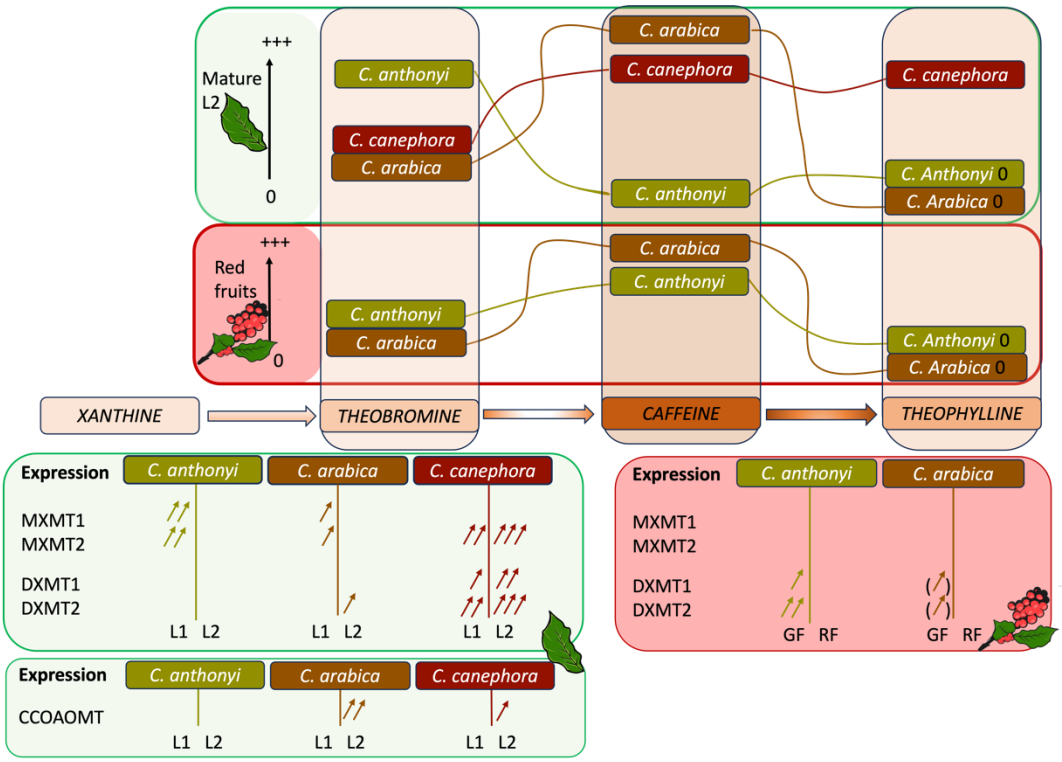

Figure 7. Overview of metabolites and gene transcripts measured in *C. anthonyi*, *C. arabica* and *C. canephora*. Samples were young (L1) and mature leaves (L2), green fruits (GF) and ripened fruits (RF). +++ high-level, 0 = not detected. Gene transcript levels are indicated by arrows.

Theobromine concentration was greater in *C. canephora* compared to *C. arabica* and *vice versa* for caffeine (**Table 1 & Fig. 7**). Theophylline was only detected in *C. canephora*. The species *C. anthonyi* stood out from the other two, by an elevated concentration of theobromine and an absence of caffeine in the old leaves. When it comes to gene expression, *C. canephora* was set apart by showing important transcript levels of *MXMT1*, *DXMT1* and *DXMT2* in the old leaves (**Fig. 2**). In contrast, these genes were lowly or not expressed in the other two genotypes (except for *DXMT2* in *C. arabica*). Conversely

in young leaves, *MXMT1* was expressed in *C. anthonyi* and *C. arabica* but not in *C. canephora*. The *MXMT2* transcript levels were found in young leaves of all three genotypes.

Previous studies (Monteiro et al., 2019; Montis et al., 2021) confirmed that theobromine concentrations in leaves were greater in *C. canephora* (in the range 0.1 – 0.8 mg g^{-1} dry matter) than in *C. arabica* (< 0.1 mg g^{-1} dry matter) (Fujimori and Ashihara, 1994). Nonetheless, values were ten times lower in the present study (**Table 1**), which may be due to theobromine being synthesized and quickly converted to caffeine. The concentrations of caffeine in *C. arabica* and *C. canephora* leaves were comparable to those in published reports (4-5 mg g^{-1} dry matter) (Monteiro et al., 2019). The concentrations of theophylline were previously reported in the range 0.1 mg g^{-1} dry matter in both *C. arabica* and *C. canephora* leaves (Spiller, 2019). Here, theophylline concentration in *C. canephora* was comparable, while it was not detected in *C. arabica* (**Table 1**). The absence of caffeine in *C. anthonyi* leaves may be the reason theophylline was not detected, underlying that theophylline is a catabolite of caffeine. However, theophylline was not detected in *C. arabica*, despite the higher amount of caffeine as also demonstrated by Fujimori and Ashihara (1994). The accumulation of caffeine during the leaf development is well known in *Coffea species* but theophylline is barely detected (Fujimori and Ashihara, 1994).

The *MXMT1* and *MXMT2* genes, encoding enzymes catalysing the conversion of 7-methylxanthine into theobromine, were preferentially expressed in young leaves of *C. arabica* (**Fig. 2 A,B**), matching previous observations (Uefuji et al., 2003). Perrois et al. (2015) reported that *DXMT1 and DXMT2*, which encode enzymes converting 7-methylxanthine into theobromine as well as theobromine into caffeine, were more expressed in *C. canephora* than *C. arabica*. Intriguingly, the caffeine was slightly greater in *C. arabica*, in contrast with what was measured by Perrois et al. (2015). In that study, *C. canephora* synthetised slightly more caffeine. Moreover, the conversion of theobromine and caffeine seemingly occurs in all the developmental stages of the *C. arabica* leaves (Ashihara, 2006). Conversely in *C. anthonyi* leaves, theobromine concentration was elevated and caffeine undetected (**Table 1**), while *DXMT* and *MXMT* genes were lowly or not expressed (**Fig. 2**). Indeed, *C. anthonyi* leaves contained low transcript levels of *MXMT1* and no *DXMT* transcript, while *C. canephora* import ant transcripts of *MXMT2* (**Fig. 2**).

The MXMT1 and MXMT2 enzymes differ by their Michaelis constants. Indeed, the *Km* values are respectively 148 and 251 (Uefuji et al., 2003), meaning that MXMT1 has a greater substrate affinity compared to MXMT2. This could also account for greater theobromine in C. *anthonyi* compared to *C. canephora* leaves. Notably, *MXMT1* was primarily expressed in the young leaves and theobromine was detected in greater concentrations in the mature leaves, indicating an accumulation of

theobromine during development. The absence of *DXMT* transcripts in *C. anthonyi* could therefore explain the absence of caffeine in leaves of that species.

4.2 Quantification of xanthines associated with N-methyltransferases expression in fruit

Fruits of *C. canephora* were not available due to culture conditions. Theobromine and caffeine profiles in ripened fruits (**Table 2**) were consistent with those found in old leaves (**Table 1**). Theobromine concentration was higher in mature fruits of *C. anthonyi* than *C. arabica*, and the opposite was true for caffeine. Theophylline was not detected in the fruits. The *DXMT1* and *DXMT2* transcript levels were important in green fruits of *C. anthonyi* (**Fig. 2C, D**).

In published reports, theobromine was detected within the same range (Mehari et al., 2016; Spiller, 2019), but caffeine was more important (in the range 5 - 23 mg g^{-1} dry matter) in *C. arabica* (Perrone et al., 2008; Alonso-Salces et al., 2009; Duarte et al., 2010; Gichimu et al., 2014; Mehari et al., 2016).

The *MXMT1* and *MXMT2* transcripts were very low in green and absent in red fruits of the two genotypes (**Fig. 2A, B**). These findings are in line with previous research. Early fruit developmental stages are shown to exhibit transient expression of *MXMT1* (Perrois et al., 2015). Also, *MXMT2* is early expressed but undetectable in yellow and red fruits of *C. arabica (Maluf et al., 2009)*, and untraceable in *C. canephora (Koshiro et al., 2006)*. The *DXMT1* and *DXMT2* genes were more expressed in green than red fruits of both species, and transcripts were much more abundant in *C. anthonyi* than *C. arabica* (**Fig. 2C, D**). A literature survey indicates *DXMT1* transcripts are detected during developmental stages before *C. arabica* and *C. canephora* fruits ripen, and *DXMT2* follows the same expression pattern (Koshiro et al., 2006; Ashihara et al., 2008). The *C. arabica* green fruits already reached their final size and were starting to turn green-yellow, which could account for the low *DXMT* transcript levels (Maluf et al., 2009).

We noted that *DXMT1* and *DXMT2* were overall more expressed than *MXMT1* and *MXMT2* across organs and species (**Fig. 2**). Hence, the two first genes probably have a greater contribution to theobromine accumulation in tissues, even though *MXMT1* and *MXMT2* are solely converting 7-methylxanthine into theobromine. Presumably, the disruption of caffeine biosynthesis in *C. anthonyi* leaves is caused by low expression of *DXMT1* and *DXMT2*, rather than of these selective genes involved in theobromine biosynthesis.

Finally, since the concentration of purine alkaloids in leaves and fruits somehow correlated positively with the expression of biosynthetic genes across the three species, we conclude that xanthine concentration is partially influenced by genetic factors. The absence of *DXMT* expression prevents caffeine synthesis, but the expression level is not well correlated with the quantity of caffeine found

in organs. Despite having more caffeine in every organ than the other two species, *C. arabica* had far fewer transcripts. Allelic variation in *DXMT* could account for that. Finally, phloem translocation could be negligeable. Indeed, phloem exudates contain no theobromine in all three species and very low caffeine only in *C. arabica* and *C. canephora* (Montis et al., 2022). This means that theobromine is most likely synthesized and stored *in situ*.

4.2 O-methyltransferase expression is related to N-methyltransferase expression

The 5-FQA concentration and *CCoAOMT* transcript levels in leaves of *C. anthonyi* were two or three times less important compared to the two other species (**Fig. 3, 4**). These results were surprising in view of the important CQAs levels previously reported in *C. anthonyi* leaves (Montis et al., 2022). A disruption in the FQAs biosynthetic pathway may occur when the *O*-methyltransferases are implicated (Clifford et al., 2017). We conclude that altered caffeine and FQAs biosynthesis in *C. anthonyi* leaves is likely due to low expression of genes encoding *N*- and *O*-methyltransferases.

4.3 Biosynthesis and catabolism of xanthines in flowers

Xanthine metabolites quantification was done in the flower buds, gynoecium and petals of *C. arabica* (**Fig. 6**, **Table 3**). The gynoecium, the flower tissue that will eventually become a fruit, had greater theobromine precursor levels compared to the flower buds or the petals. This could be due to a greater biosynthesis of caffeine prior to fructification. The presence of theophylline and the absence of theobromine in the petals, suggest that caffeine is more catabolized there than in the gynoecium. Thus, biosynthesis of caffeine happens in the gynoecium, while its catabolism occurs in the petals. As caffeine was found very low in the phloem (Montis et al., 2022), and both caffeine and theobromine elevated in the gynoecium (**Table 3**), we assume that caffeine is synthesized at an early developmental stage and stored in the fruits. The optimal defence theory may also account for a more efficient caffeine biosynthesis in the gynoecium. Accordingly, plant organs that are more susceptible to pest attacks produce greater quantities of specialized metabolites. These are often more concentrated in the fruits than in any other tissues, because they contain the seeds to be dispersed for ensuring species survival. For example, the angel's trumpet (*Brugmansia suaveolens*) fruits have twice as much scopolamine as the flowers and leaves (Meldau et al., 2012).

5. Conclusion

Three coffee species grown in a controlled greenhouse environment were the focus of this study. The purpose was to ascertain how metabolite concentration and biosynthetic gene expression relate to one another, especially in *C. anthonyi*, a new species indigenous from Central Africa. Theobromine accumulation in leaves of that genotype is likely the result of a disruption in caffeine biosynthesis caused by the lack of synthase gene expression. Still, green fruits regain expression and produce caffeine. The organ-specificity of caffeine biosynthesis emphasizes the significance of epigenetic and gene regulation. In fact, *C. anthonyi* conserved functional copies of synthase genes but does not express them, while *C. humblotiana* has mutated genes and caffeine is totally absent in all organs (Raharimalala et al., 2021). In contrast, the closest relative *C. eugenoides* synthesises low quantity of caffeine in leaves. Given that *C. anthonyi* successfully evolved over time, these findings call into question the defence role of caffeine in leaves. A survey of the *C. anthonyi* population divergence should validate this evolution in caffeine histolocalisation and the origin of such observation. Finally, after examining *C. arabica* flowers, we showed that caffeine synthesis already occurs in the gynoecium, at a very early fruit development stage.

5. References

Alonso-Salces RM, Serra F, Reniero F, Héberger K (2009) Botanical and geographical characterization of green coffee (*Coffea arabica* and *Coffea canephora*): chemometric evaluation of phenolic and methylxanthine contents. J Agric Food Chem **57**: 4224–4235

Ashihara H (2006) Metabolism of alkaloids in coffee plants. Braz J Plant Physiol **18**: 1–8

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

Chen X (2019) A review on coffee leaves: Phytochemicals, bioactivities and applications. Critical Reviews in Food Science and Nutrition **59**: 1008–1025

Clifford MN, Jaganath IB, Ludwig IA, Crozier A (2017) Chlorogenic acids and the acyl-quinic acids: discovery, biosynthesis, bioavailability and bioactivity. Nat Prod Rep **34**: 1391–1421

Dey P, Kundu A, Kumar A, Gupta M, Lee BM, Bhakta T, Dash S, Kim HS (2020) Analysis of alkaloids (indole alkaloids, isoquinoline alkaloids, tropane alkaloids). Recent Advances in Natural Products Analysis 505–567

Duarte GS, Pereira AA, Farah A (2010) Chlorogenic acids and other relevant compounds in Brazilian coffees processed by semi-dry and wet post-harvesting methods. Food Chemistry **118**: 851–855

Fujimori N, Ashihara H (1994) Biosynthesis of theobromine and caffeine in developing leaves of *Coffea arabica*. Phytochemistry **36**: 1359–1361

Gichimu BM, Gichuru EK, Mamati GE, Nyende AB (2014) Biochemical Composition Within *Coffea arabica* cv. Ruiru 11 and Its Relationship With Cup Quality. JFR **3**: 31

Koshiro Y, Zheng X-Q, Wang M-L, Nagai C, Ashihara H (2006) Changes in content and biosynthetic activity of caffeine and trigonelline during growth and ripening of *Coffea arabica* and *Coffea canephora* fruits. Plant Science **171**: 242–250

Li W, Zhang F, Wu R, Jia L, Li G, Guo Y, Liu C, Wang G (2017) A novel N-methyltransferase in Arabidopsis appears to feed a conserved pathway for nicotinate Detoxification among land plants and Is associated with lignin biosynthesis. Plant Physiology **174**: 1492–1504

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods **25**: 402–408

Maluf MP, da Silva CC, de Oliveira M de PA, Tavares AG, Silvarolla MB, Guerreiro O (2009) Altered expression of the caffeine synthase gene in a naturally caffeine-free mutant of Coffea arabica. Genet Mol Biol **32**: 802–810

Mehari B, Redi-Abshiro M, Chandravanshi BS, Atlabachew M, Combrinck S, McCrindle R (2016) Simultaneous Determination of Alkaloids in Green Coffee Beans from Ethiopia: Chemometric Evaluation of Geographical Origin. Food Anal Methods **9**: 1627–1637

Meldau S, Erb M, Baldwin IT (2012) Defence on demand: mechanisms behind optimal defence patterns. Ann Bot **110**: 1503–1514

Mizuno K, Okuda A, Kato M, Yoneyama N, Tanaka H, Ashihara H, Fujimura T (2003) Isolation of a new dual-functional caffeine synthase gene encoding an enzyme for the conversion of 7-methylxanthine to caffeine from coffee (Coffea arabica L.)1. FEBS Letters **534**: 75–81

Monteiro Â, Colomban S, Azinheira HG, Guerra-Guimarães L, Do Céu Silva M, Navarini L, Resmini M (2019) Dietary Antioxidants in Coffee Leaves: Impact of Botanical Origin and Maturity on Chlorogenic Acids and Xanthones. Antioxidants (Basel) **9**: 6

Montis A, Souard F, Delporte C, Stoffelen P, Stévigny C, Van Antwerpen P (2021) Coffee Leaves: An Upcoming Novel Food? Planta Med. doi: 10.1055/a-1533-0021

Montis A, Souard F, Delporte C, Stoffelen P, Stévigny C, Van Antwerpen P (2022) Targeted and Untargeted Mass Spectrometry-Based Metabolomics for Chemical Profiling of Three Coffee Species. Molecules **27**: 3152

Perrois C, Strickler SR, Mathieu G, Lepelley M, Bedon L, Michaux S, Husson J, Mueller L, Privat I (2015) Differential regulation of caffeine metabolism in Coffea arabica (Arabica) and Coffea canephora (Robusta). Planta **241**: 179–191

Perrone D, Donangelo CM, Farah A (2008) Fast simultaneous analysis of caffeine, trigonelline, nicotinic acid and sucrose in coffee by liquid chromatography-mass spectrometry. Food Chem **110**: 1030–1035

Petermann JB, Baumann TW (1983) Metabolic Relations between Methylxanthines and Methyluric Acids in Coffea L. Plant Physiol **73**: 961–964
Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res **29**: e45
Raharimalala N, Rombauts S, McCarthy A, Garavito A, Orozco-Arias S, Bellanger L, Morales-Correa AY, Froger S, Michaux S, Berry V, et al (2021) The absence of the caffeine synthase gene is involved in the naturally decaffeinated status of Coffea humblotiana, a wild species from Comoro archipelago. Sci Rep **11**: 8119
Rinaudo P, Boudah S, Junot C, Thévenot EA (2016) biosigner: A New Method for the Discovery of Significant Molecular Signatures from Omics Data. Front Mol Biosci **3**: 26
Salgado PR, Favarin JL, Leandro RA, Lima Filho OF de (2008) Total phenol concentrations in coffee tree leaves during fruit development. Sci agric (Piracicaba, Braz) **65**: 354–359
Souard F, Delporte C, Stoffelen P, Thévenot EA, Noret N, Dauvergne B, Kauffmann J-M, Van Antwerpen P, Stévigny C (2018) Metabolomics fingerprint of coffee species determined by untargeted-profiling study using LC-HRMS. Food Chem **245**: 603–612
Spiller GA (2019) Caffeine. doi: 10.1201/9780429126789
Stoffelen P, Noirot M, Couturon E, Bontems S, Block PD, Anthony F (2009) Coffea anthonyi, a new self-compatible Central African coffee species, closely related to an ancestor of Coffea arabica. TAXON **58**: 133–140
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
Zheng X, Ashihara H (2004) Distribution, biosynthesis and function of purine and pyridine alkaloids in Coffea arabica seedlings. Plant Science **166**: 807–813