Philippine Journal of Science
153 (3): 717-743, June 2024
ISSN 0031 - 7683
Date Received: 05 Oct 2023

Characterization of *N*-methyltransferase Genes in *Coffea liberica* var. *liberica*

Miguel V. Mondragon[1,2]*, Terrence Ferdinand S. Nagaño[1,2], and Daisy May C. Santos[1,2]

[1]Natural Sciences Research Institute, College of Science,
University of the Philippines Diliman, Diliman, Quezon City 1101 the Philippines
[2]Institute of Biology, College of Science,
University of the Philippines Diliman, Diliman, Quezon City 1101 the Philippines

Caffeine biosynthesis in *Coffea* evolved independently through tandem gene duplication and subsequent neofunctionalization of ancestral *N*-methyltransferases (NMTs). The bulk of our understanding of this process in *Coffea* is centered on two species: *C. arabica* and *C. canephora*. To gain a wider perspective of caffeine biosynthesis in *Coffea*, this study focused on the less popular but commercially available *C. liberica* var. *liberica*, locally known as "kapeng barako." This study aimed to determine the genomic and coding sequences of the three NMTs – mainly, xanthosine methyltransferase (XMT), 7-methylxanthine-*N*-methyltransferase (MXMT), and dimethylxanthine methyltransferase (DXMT) – through Sanger sequencing. Relative expression levels of these NMTs in the leaves, flower bud, green fruit, and red fruit stages were also determined through polymerase chain reaction amplification of synthesized cDNA and then gel electrophoresis. Results reveal that the genetic sequences of NMTs of *C. liberica* were similar to the *C. eugenioides*-derived NMTs of *C. arabica*. Results also indicate that there is a possibility that more than one NMT gene for each NMT cluster (XMT, MXMT, DXMT) exists. This study has also identified genomic sequences and transcripts, whose sequences were in between the *XMT* and *MXMT* gene clusters. The function of these novel sequences is yet to be determined. NMT expression levels in *C. liberica* were similar to those observed in *C. arabica* and *C. canephora*, wherein expression peaks at the green fruit stage and then diminishes, as the fruit ripens fully. Although the NMTs of *C. liberica* have a lot of similarities with *C. arabica* and *C. canephora*, discovering the function and role in caffeine biosynthesis of these novel genomic sequences and transcripts will help us further understand the evolution of caffeine biosynthesis in *Coffea*.

Keywords: caffeine, *Coffea*, Liberica, methyltransferase

INTRODUCTION

Caffeine is an ever-present psychostimulant in the modern world commonly consumed in popular beverages such as tea, carbonated soda, and coffee (Reyes and Cornelis 2018). The biosynthesis of caffeine (1,3,7-trimethylxanthine) begins with a series of reactions involving adenine nucleotide which produces xanthosine. The final steps involve the successive methylation of xanthosine at the 7-*N*, 3-*N*, and 1-*N* position by a variety of *N*-methyltransferases (NMTs), resulting in the following molecules: 7-methylxanthosine, 3,7-dimethylxanthine (theobromine), and finally, 1,3,7-trimethylxanthine (caffeine) (Ashihara *et al.* 1996). The three successive NMTs involved have been identified as xanthosine methyltransferase (XMT), 7-methylxanthine methyltransferase (MXMT), and 3,7-dimethylxanthine

*Corresponding author: mvmondragon@up.edu.ph

methyltransferase (DXMT) (Moisyadi *et al*. 1998; Ogawa *et al*. 2001; Uefuji *et al.* 2003).

Caffeine biosynthesis evolved independently in tea (*Camellia sinensis*) (Xia *et al.* 2017), cacao (*Theobroma cacao*) (Argout *et al.* 2011), and robusta coffee (*Coffea canephora*) (Denoeud *et al.* 2014). Tea caffeine synthases have a low degree of sequence similarity (< 40%) with coffee caffeine synthases; however, they have similar substrate specificities (Zhou *et al.* 2020). Contrastingly, NMTs such as XMT, MXMT, and DXMT within *Coffea* share high sequence similarity (> 80%) with each other; however, they have different substrate specificities. Caffeine biosynthesis in coffee was driven by tandem duplications and subsequent functional specialization of NMTs (Xu *et al.* 2020). Caffeine content also varies between different coffee species – from as low as 0% dry matter basis (dmb) in *Coffea pseudozanguebariae* to as high as 2.5% dmb in *C. canephora* (Anthony *et al.* 1993). However, the bulk of the research on NMTs of *Coffea* has mainly been focused on the Arabica (*C. arabica*) and robusta (*C. canephora*) varieties. For example, the crystal structures of two NMTs – namely, XMT and DXMT – from *C. canephora* have been elucidated by McCarthy and McCarthy (2007). These crystal structures would be essential in predicting the secondary structures of newly discovered homologous proteins. The sequences and expression analysis of these genes have also been studied in both *C. arabica* and *C. canephora* (Perrios *et al.* 2015). The highest NMT expression levels were observed in young leaves for both species. Expression levels were reduced by up to 10 and 70 times less during coffee bean development from the small green grain stage up to the red grain stage in both Robusta and Arabica, respectively. In the final stages of bean maturation, specifically the red grain stage, expression levels were almost undetectable (Perrois *et al.* 2015). Caffeine content also reflects the NMT expression levels observed. The highest caffeine concentration was found in young leaves by as much as 3% (percentage of dry weight) in robusta and 1.6% (dry weight) in Arabica (Perrois *et al*. 2015). In the fruit, caffeine was observed as early as the small green grain stage, and accumulated rapidly until the large green grain stage. A decrease in caffeine content was observed from the large green grain stage to the yellow grain stage. Caffeine content then stabilized during the yellow grain stage until the red grain stage (Perrois *et al*. 2015). This trend in caffeine accumulation aligns with caffeine's role as the plant's defense against pest infestation and herbivory (Kim *et al.* 2010). High caffeine content found in the young leaves and developing fruit protects them from pests and discourages herbivores from eating them until they mature.

In order to have a better understanding of NMTs and caffeine biosynthesis in coffee, there is a need to diversify research to include the wild and less popular species of *Coffea*. Obtaining additional NMT sequences from other *Coffea* species will help us determine which sites are conserved within *Coffea* and which are crucial for substrate recognition.

This study focused on the Liberica variety (*C. liberica* var. *liberica*) – a commercially cultivated coffee species in the Philippines, Malaysia, and Indonesia (PCBI 2020). Although not as popular as Arabica and Robusta, this type of coffee – locally known as "kapeng barako" – is slowly gaining followers in the local coffee scene. Currently, there are only five NMT sequences of *C. liberica* available in the GenBank database. Four of these are unpublished putative NMT mRNA sequences from *C. liberica* clones, whereas the last one is a genomic putative NMT sequence from *C. liberica* var. *dewevrei*. The identities of these putative NMT sequences have not been further analyzed in any published journal. Preliminary analysis of these coding sequences indicated that three were putative DXMT sequences, whereas the other one was a putative XMT sequence. Representation of all the NMTs – specifically, XMT, MXMT, and DXMT – would be ideal in trying to understand caffeine synthesis specifically in *C. liberica*. Although *C. liberica* var. *dewevrei* is genetically similar to *C. liberica* var. *liberica*, intraspecies genetic differences may be present in these NMT genes.

Currently, there is insufficient data available regarding the genomic and coding sequences of the NMT genes of *C. liberica* var. *liberica*. Studies on the expression of these NMT genes in *C. liberica* var. *liberica* also have not been studied yet. Therefore, this study aimed to sequence the genomic and coding sequences of the three NMTs – namely, XMT, MXMT, and DMXT – in *C. liberica* var. *liberica*. Expression levels of these NMTs were also measured during the different developmental stages to determine if the expression profile of Liberica is different or similar to Arabica and Robusta.

MATERIALS AND METHODS

Sample Collection, DNA and RNA Extraction, and cDNA Synthesis

Coffee samples were obtained from three *C. liberica* var. *liberica* trees located at the Institute of Biology, University of the Philippines (UP) Diliman. These Liberica trees were obtained from the Philippine Coffee Board. In Arabica and Robusta, NMT expression levels varied depending on the development stage of the sample (Perrois *et al.* 2015). For this study, the extended BBCH (Biologische Bundesanstalt, Bundessortenamt, und Chemische Industrie) scale was used as a guide when identifying the different developmental stages (Arcila-Pulgarin *et al.* 2002). Young leaf and tissue samples from various stages of fruit development – namely,

flower bud (BBCH 71), small green fruit (BBCH 73-75), and ripe red fruit (BBCH 85-88) – were collected. Three samples of each tissue type were obtained. All samples were sliced to at most 0.5 cm in at least one dimension and submerged in five volumes of RNAlater solution (Invitrogen, CAT No.: AM7024). Samples were stored at –20 °C for at most nine months until all samples were collected and were ready for extraction.

Tissue samples suspended in RNAlater solution were taken out and pulverized using mortar and pestle. Genomic DNA was extracted from the young leaf samples using the DNEasy Plant Mini Kit (Qiagen, CAT No.: 69104). Meanwhile, total RNA was extracted from all the sample tissues using the ISOLATE II RNA Mini Kit (Bioline, CAT No.: BIO-52072). The concentration and quality of the extracted DNA and RNA were measured using an Epoch Microplate Spectrophotometer (BioTek - Winooski, Vermont USA). The first-strand cDNA synthesis was performed following the protocol of Viva cDNA Synthesis Kit (Vivantis, CAT No.: cDSK01-050) using oligo d(T) primers. The cDNA samples were then stored at –20 °C until further processing.

gDNA and cDNA Amplification and Sequencing of Caffeine-related N-methyltransferases (NMTs)

Based on initial polymerase chain reaction (PCR) experiments, the combination of CAF2-ATG and CAF2-STOP primers (Perrois *et al.* 2015) produced the brightest bands when used with Liberica genomic DNA samples. These were subsequently used for the non-specific amplification of NMT genes. The amplified PCR products were then excised and purified from an agarose gel. These were then cloned using the pGEM®-T Easy Vector system (Promega, Cat No.: A1360), following the manufacturer's protocol. Colonies positive in blue and white screening were then separately cultivated in lysogeny broth at 37 °C with constant shaking (150 rpm) for 12 h. Plasmids were extracted using the ISOLATE II Plasmid Mini Kit (Bioline, Cat No.: BIO-52056), following the manufacturer's protocol.

Using the plasmid DNA as a template, the target gene was amplified again with their respective primers. PCR products were run through agarose gel to check if the transformation was successful. Afterward, the products were excised, purified, and then sent to Macrogen Inc. (Seoul, South Korea) for Sanger sequencing. The optimal length for Sanger sequencing is only around 800 base pairs (bp), whereas the NMT genes are around 2000 bp in length. Therefore, a primer located in the middle (CAF1-MIDR) was designed based on the *Coffea* NMT sequences from Perrois *et al.* (2015) to overlap a portion of the sequence. Three primers – namely, CAF2-ATG, CAF2-STOP, and CAF1-MIDR – were utilized in sequencing each of the samples.

A total of 17 clones were successfully transformed carrying a copy of one genomic NMT gene. BLAST results of their sequence against the NCBI database revealed that three clones (C3, C12, and C15) (Figure 1) had the *C. arabica XMT2* gene (JX978515) as their top result when sorted by E value and max score. On the other hand, the top result for the other 14 was a putative NMT gene from *C. liberica* var. *dewevrei* (AY362825). Thirteen (13) of these clones had the *C. arabica MXMT2* gene (JX978512) as the 2nd best result, whereas one clone (C2) had the *C. arabica XMT2* gene. No clones likely carried a putative DXMT gene. Therefore, a reverse primer (CcDXMT_R2) specific for the *DXMT* gene was designed based on the *Coffea* DXMT sequences of Perrois *et al.* (2015) and used together with the CAF2-ATG primer. Amplified products also underwent the same transformation process. Out of the five colonies chosen for cultivation, three clones (R1, R3, and R4) were found to successfully carry a putative *DXMT* gene.

For the amplification of the coding sequence, the CAF2-ATG primer and the reverse primer specific for each NMT gene – namely, CcXMT_R, CcMXMT_R, and CcDXMT_R2 for *XMT*, *MXMT*, and *DXMT*, respectively – were used. The list of all primers used is found in Table 1. After

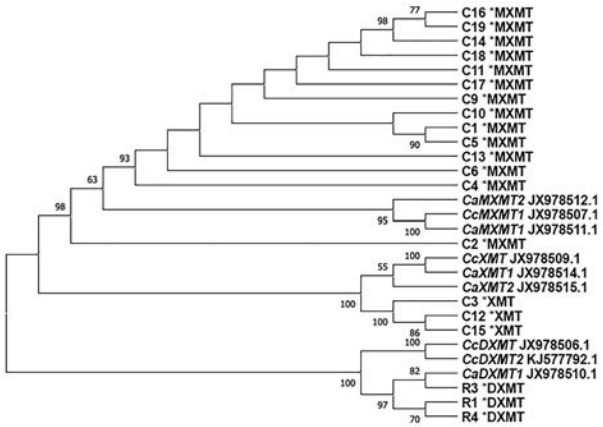


Figure 1. Cladogram of genomic *N*-methyltransferase (NMT) genes from *Coffea* constructed using the neighbor-joining (NJ) method based on 2090 nucleotides of *C. liberica* var. *liberica* sequences (numbered clones labeled C and R) obtained in this study and reference sequences of *C. arabica* and *C. canephora* from Perrois *et al.* (2015) using the maximum composite likelihood method of DNA substitution. The tree is rooted in the dimethylxanthine *N*-methyltransferase (*DXMT*) genes. Values on nodes represent bootstrap percentages out of 1000 bootstrap samples. Numbered clones labeled with C were amplified using nonspecific NMT primers – namely, CAF2-ATG and CAF2-STOP – whereas clones labeled with R were amplified using the CAF2-ATG and CcDXMT_R2 primers. Legend: [Ca] *C. arabica*; [Cc] *C. canephora*; [XMT] xanthosine methyltransferase; [MXMT] 7-methylxanthine-*N*-methyltransferase; [DXMT] dimethylxanthine methyltransferase.

Table 1. Primer pairs used in the study to amplify specific and non-specific *N*-methyltransferases (NMT). Primer CAF1-MIDR was utilized during Sanger sequencing to provide sequence overlaps in the amplification of the genomic NMT gene.

Amplification target	Forward primer used (name and sequence)	Reverse primer used (name and sequence)	Amplicon inserted into
Non-specific NMT	CAF2-ATG [a] (ATGGAGCTCCAAGAAGTCCTGCG) CAF1-MIDR [b] (CCCAATTCAATCACCAAACCGC)	CAF2-STOP [a] (ATTCACGTCTGACTTCTCTGGCT)	Clones labeled C (genomic)
XMT-specific	CAF2-ATG [a] (ATGGAGCTCCAAGAAGTCCTGCG)	CcXMT1_R [a] (ATGGGTTCGTAAACTGCTCTAACG)	Clones labeled E (coding)
MXMT-specific	CAF2-ATG [a] (ATGGAGCTCCAAGAAGTCCTGCG)	CcMXMT1_R [a] (AGGATGGGTTCGTAAACTGATCTAA)	Clones labeled F (coding)
DXMT-specific	CAF2-ATG [a] (ATGGAGCTCCAAGAAGTCCTGCG)	CcDXMT_R2 (TCATCGCAGGATACTGGGGA)	Clones labeled R (genomic), and G (coding)

[a]Primers from Perrois *et al.* (2015)
[b]Additional primer utilized during Sanger sequencing

amplification, the same methods as above were used for transforming and sequencing the specific NMT cDNAs. After transformation, only five colonies from each were randomly chosen for cultivation. For sequencing, only the CAF2-ATG and the corresponding specific NMT reverse primer were used. The CAF1-MIDR was not used anymore because the expected product length was only around 1000 bp.

PCR was performed in 50-µL reactions as follows: 32.5 µL of nuclease-free water, 10 µL of 5x MyTaq Reaction Buffer (Bioline, Cat No.: BIO-21106), 1 µL of MyTaq DNA polymerase (Bioline, Cat No.: BIO-21106), 2 µL each of the forward and reverse primers (10 µM), and 2.5 µL of the DNA template. For non-specific primer amplification, the PCR running condition used was the following: initial denaturation at 95 °C for 4 min, 40 cycles of 95 °C for 15 s, 58 °C for 30 s, 72 °C for 1 min and 30 s, and a final extension at 72 °C for 5 min. For the primer amplification of specific NMTs, the PCR running conditions used were the following: initial denaturation at 95 °C for 5 min, 35 cycles of 94 °C for 15 s, 60 °C for 30 s, 72 °C for 1 min, and a final extension at 72 °C for 5 min.

Sequence Data Analysis

All sequencing results were aligned and analyzed using the Mega 11 software (Stecher *et al.* 2020). For the analysis of genomic sequences, consensus sequences were aligned against the NMT genomic sequences from Perrois *et al.* (2015). For the initial tree constructions, cladograms were constructed using the neighbor-joining (NJ) method with the maximum composite likelihood model of DNA substitution. Pairwise distances between each sequence were also measured using the maximum composite likelihood model of DNA substitution and assuming a heterogenous pattern of lineage and gamma distribution. For the maximum likelihood (ML) trees, the general time reversible (GTR) model of substitution was used (Guindon *et al.* 2010). All sites were considered in the analysis, and all the consensus trees were derived from 1,000 bootstrap replicates.

For the analysis of coding sequences, the obtained *C. liberica* cDNA sequences were aligned and compared with the mRNA sequences from Perrois *et al.* (2015). For the analysis of protein sequence and structure, the cDNA sequences were translated into their amino acid sequence and aligned with the translated mRNA sequences from Perrois *et al.* (2015). Jalview v2 (Waterhouse *et al.* 2009) was used to align and analyze the protein sequences. For the prediction of the secondary structure of the protein, the jnetpred, JNETHMM, and JNETPSSM programs were used (Drozdetskiy *et al.* 2015). For the analysis of the effects of the mutations, the SAAMBE-3D web server (Pahari *et al.* 2020) was used in reference to the protein structure of *C. canephora* DXMT(2EFJ). Although this program used protein-protein interactions as its training basis for prediction, this supplementary data can give us an idea if these mutations are disruptive or not to the NMT protein.

In order to ascertain the identity of the obtained NMTs, the cDNA sequences were aligned with all the *Coffea* NMT protein sequences available in NCBI. The protein sequence for salicylic acid methyltransferase from *Gardenia jasminoides* (Rubiaceae) was also included to serve as an outgroup. For the ML trees, the Jones-Taylor-Thornton plus gamma (JTT+G) model of substitution was used (Guindon *et al.* 2010). All sites were considered in the analysis and all the consensus trees were derived from 1,000 bootstrap replicates.

Analysis of N-methyltransferase (NMT) Expression during Fruit Development

To standardize the expression between leaves, flower bud, green fruit, and red fruit, the total RNA concentration of each sample was measured using an Epoch Microplate Spectrophotometer (BioTek, Winooski, Vermont, USA) and diluted to 150 ng and used as the template for cDNA synthesis. *Coffea* expression reference genes from a previous study (Cruz *et al.* 2009) – specifically, *GAPDH* (glyceraldehyde-3-phosphate dehydrogenase) and *RPL39* (large ribosomal subunit 39) – were used to check if cDNA synthesis was successful.

Expression of NMTs was determined by amplifying with both non-specific primers (CAF2-ATG and CAF2-STOP) and specific primers (CAF2-ATG with either CcXMT1_R, CcMXMT1_R, or CcDXMT_R2) (Table 1). Amplicons were run through 1% agarose gel for visualization.

RESULTS AND DISCUSSION

Amplification and Alignment of the Genomic Sequences of the N-methyltransferase (NMT) Genes

Alignment of all the clones with the genomic *Coffea* NMT genes from Perrois *et al*. (2015) reveals the probable identities of the clones (Figure 1). Out of the 14 putative *MXMT* clones, 13 clones were genetically similar except for random one-base pair mutations throughout their sequences (Appendix Figure I). None of these random mutations were present in at least two clones, which suggests that these were most likely brought upon by PCR error. An amplicon length of around 2000 bps and the use of the inherently non-proofreading *Taq* polymerase, which has an error rate of 1.8 x 10^{-4} errors/base/doubling (Potapov and Ong 2017), might have caused these base substitution errors. Interestingly, the clone (C2) was different from the rest. Its first and second introns were similar to the *XMT* genes (*CaXMT2*, C3, C12, and C15); however, its 3rd intron was closer to the *MXMT* gene (Appendix Figure I). The third intron is notable between the *XMT* and *MXMT* genes because it contains a major 303-bp indel, which is present in the *XMT* gene.

Sequence Comparison of the Three N-methyltransferases (NMTs) in *Coffea liberica*

Aside from the informative region in the fourth exon, which is the basis of the specific NMT reverse primers, another region at Site 1097–1123 (Appendix Figure I) is also beneficial in differentiating between the three NMTs. The third intron region, which exhibits multiple indels, is also very variable between the three (Appendix Figure I).

Sequence Comparison of Genomic N-methyltransferases (NMTs) between *Coffea* Species

The *XMT*, *MXMT*, and *DXMT* gDNA sequences were aligned with their respective *Coffea* reference sequences from Perrois *et al*. (2015). For the *XMT* gene, the reference sequence closest to the putative *C. liberica* XMT sequences – in terms of pairwise distance – was the *CaXMT2* gene from *C. arabica* (JX978515) (Appendix Table I). An insertion (1266_1269insTTAG) at the 3rd intron region was present in the *CaXMT2* reference sequence and *C. liberica* sequences (Appendix Figure I), but absent in the *CaXMT1* (JX978514.1) and *C. canephora* reference sequence (JX978509). An indel in Site 1836–1874 in the 4th exon region was also present in *CaXMT2* and *C. liberica* sequences but absent in the rest. Differences between *C. liberica* sequences and the rest – *C. arabica* and *C. canephora* – are summarized in Table 2.

Table 2. List of informative sites that can differentiate putative *C. liberica N*-methyltransferases from their respective *C. arabica* and *C. canephora N*-methyltransferases from Perrois *et al.* (2015). Multiple sequence alignments can be found in Appendix Figure I.

	1st intron	2nd exon	2nd intron	3rd exon	3rd intron	4th exon
XMT	122_127insCA 174T>C 188C>T 210A>C	281C>T 282G>T 296A>G	761A>T 766T>G 801C>A 830T>G		1263A>G 1271T>C 1294T>C 1384A>G 1464T>C 1561G>C 1599G>A	1731A>T 1738G>A
MXMT			759A>C 761A>T 801C>A 819T>C 830T>G	936T>C 1156A>G	1201T>C Multiple indels	1727T>A 1763A>G 1852A>C 2012A>G
DXMT	173G>C	437A>G	696del 835A>G	1014G>T 1144A>C	1204A>T 1224A>T 1244_1247ins 1390_1529del	1717G>C 1738A>G 1773T>C

For the *MXMT* gene, the reference sequence closest to the putative *C. liberica* MXMT sequences – in terms of pairwise distance – was the *CaMXMT2* gene from *C. arabica* (JX978512) (Appendix Table I). Aside from various substitutions in the 2nd intron, 3rd exon, and 4th exon, the largest amount of variation between the *Coffea* species can be found in the 3rd intron region (Table 2; Appendix Figure I).

For the *DXMT* gene, the reference sequence closest to the putative *C. liberica* DXMT sequences – in terms of pairwise distance – is the *CaDXMT1* gene from *C. arabica* (JX978510.1) (Appendix Table I). Unfortunately, the gDNA of *DXMT* was only sequenced up to Site 1846 as a limitation of using a *DXMT*-specific primer located at Site 1861–1880. A number of substitutions can be found from the 1st intron up to the 4th exon. Similar to the *MXMT* gene, the greatest variation can be seen in the 3rd intron, which includes a 3-bp insertion and a major deletion at Site 1390–1529 (Table 2).

The *CaXMT2* (JX978515), *CaMXMT2* (JX978512), and *CaDXMT1* (JX978510) genes all come from the *C. eugenioides* subgenome of *C. arabica* (Perrois *et al.* 2015). As stated earlier, these reference genes were also the closest to their respective putative *C. liberica* NMT sequences in terms of pairwise distance (Appendix Table I). This result is in contrast with previous literature where in most phylogenetic trees *C. liberica* clusters closer to *C. canephora* than to *C. eugenioides* (Hamon *et al.* 2017; Charr *et al.* 2020; Huang *et al.* 2023).

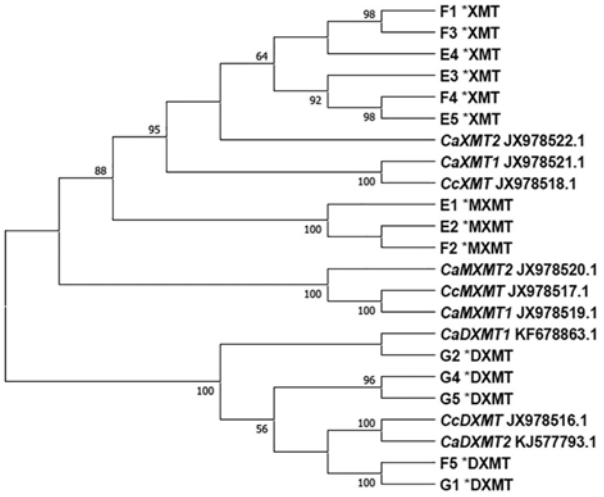


Figure 2. Cladogram of the coding sequences of *N*-methyltransferase genes from *Coffea* constructed using the neighbor-joining (NJ) method based on 962 nucleotides of *C. liberica* var. *liberica* sequences (numbered clones labeled E, F, and G) obtained from this study and reference sequences of *C. arabica* and *C. canephora* from Perrois *et al.* (2015) using the maximum composite likelihood method of DNA substitution. The tree is rooted in the dimethylxanthine *N*-methyltransferase (*DXMT*) genes. Values on nodes represent bootstrap percentages out of 1000 bootstrap samples. Numbered clones labeled with E, F, and G were amplified using the CAF2-ATG and their respective reverse primers – namely, CcXMT1_R, CcMXMT1_R, and CcDXMT_R2. Legend: [Ca] *C. arabica*; [Cc] *C. canephora*; [XMT] xanthosine methyltransferase; [MXMT] 7-methylxanthine-*N*-methyltransferase; [DXMT] dimethylxanthine methyltransferase.

Amplification and Alignment of the Coding Sequences of the N-methyltransferase (NMT) Genes

Alignment of the sequences with the NMT mRNAs from Perrois *et al.* (2015) reveals some unexpected results, as seen in the cladogram (Figure 2). Clones carrying an amplified product of the CAF2-ATG and CcDXMT_R2 primer (G1, G2, G4, and G5) (Table 1) had sequences close to the *DXMT* genes, as expected. Clones transformed by amplicons from the CAF2-ATG and CcXMT_R primers were labeled E1–E5, whereas clones transformed by amplicons from the CAF2-ATG and CcMXMT_R primers were labeled F1–F5 (Table 1). However, these clones did not cluster as expected. Six clones had sequences closer to the *XMT* gene (E3, E4, E5, F1, F3, and F4), whereas three clones were in between the *XMT* and *MXMT* genes (E1, E2, and F2). Lastly, one clone was close to the *DXMT* gene (F5).

Among the *XMT* putative clones, Clones F1–F3 were identical to each other, whereas Clone F4 was identical to E5. There is also a high sequence similarity between these two pairs with only nine base substitutions interspersed within the gene. In some sites, they also share the same base substitution compared to the *CaXMT2* reference gene. It is more probable that these pairs are heterozygous alleles of the same *XMT* gene, rather than coming from different *XMT* genes. Clone E4 was missing the whole 3rd exon which would suggest that it is a splice variant.

Protein Sequence Comparison of N-methyltransferases (NMTs) between *Coffea* Species

XMT, MXMT, and DXMT are part of the motif B′ methyltransferase family (Zhou *et al.* 2020). There are four conserved motifs shared by this family – namely, Motifs A, B′, C, and YFFF – promptly marked in Figure 3 (Zhou *et al.* 2020). As stated earlier, one clone (E4) was missing the whole 3rd exon. This region contains a number of secondary structures, as well as the motif YFFF (Figure 3). Such drastic loss might likely lead to a non-functional protein product or would result in a protein with a different function altogether.

Amino acid positions that support the clustering of the three NMTs, as noted by Perrois *et al.* (2015), were marked

with a red asterisk in Figure 3. In these informative sites, most of the sample sequences were in congruence with their putative NMT groups based on the initial cladogram. In the instances of nonsynonymous mutations, amino acid substitutions were within their functional group clustering such as positively or negatively charged, polar uncharged, or hydrophobic. An exception to this was the putative *MXMT* clones (E1, E2, and F2). In some sites, these clones were similar to the *MXMT* group – whereas in other sites, they were similar to the *XMT* group.

The work of McCarthy and McCarthy (2007) on the crystallographic structure of XMT and DXMT from *C. canephora* also uncovered particular sites of interest. Ser-316 was identified to be crucial for XR substrate specificity in XMT. Ser-316 corresponds to Ser-329 in the alignment (Figure 3) due to the 13 amino acid (aa) indel at Site 305–317 present in the *CaXMT2* and *C. liberica* sequences. Tyr-321 and Tyr-356, which are Tyr-334 and Tyr-369 (Figure 3), have also been hypothesized to help identify the substrate for XMT. Although tyrosine is also conserved in DXMT, McCarthy and McCarthy (2007) stated that the Tyr-334 in DXMT is too far to form a hydrogen bond with either 7-methylxanthine or theobromine substrates. It instead hydrogen bonds with Ser-238, thereby changing its orientation for optimal interaction with the substrate. Meanwhile, Tyr-369 in DXMT was hypothesized to disrupt substrate binding due to conformational changes in the β5-α6 loop. Interestingly, *XMTs* in *C. liberica* also contained a 13-aa insertion similar to the *DXMT* in *C. arabica*, which might lead to similar conformational changes as those above. However, there were also differences between the inserted sequence, which is DYPLRSHVQVYSD in *C. liberica* XMT and DYQGRSHSPVSCD in *C. arabica* DXMT. Whether these amino acid changes have an effect on the aforementioned issues above is yet to be determined in future studies. His-160, or His-161 in Figure 3, was found to be conserved in MXMT and DXMT, wherein it functions to stabilize the substrate (McCarthy and McCarthy 2007). A 161H>Q substitution has been observed in two *DXMT* putative clones (F5 and G1). Based on the SAAMBE-3D web server (Pahari *et al.* 2020), a slight destabilizing change of 0.48 in binding free energy was predicted. However, it was also predicted that this change would not disrupt the protein-protein interaction. Three sites have also been identified to be important for substrate specificity of DXMT and MXMT. These were Phe-27, Ser-237, and Ile-266 for DXMT and Ala-27, Pro-237, and Phe-266 for MXMT – which correspond to Sites 28, 238, and 267 in Figure 3 (McCarthy and McCarthy 2007). The putative *MXMTs* (E1, E2, and F2) had Val-28, Val-238, and Leu-267, which are closer to *XMT* than to *MXMT*. While putative *DXMTs* possessed the expected amino acids in these positions, Clones F5 and G1 had Thr-238 and Thr-267. While both of these amino acid substitutions were found non-disruptive to protein-protein interactions according to SAAMBE-3D, 238S>T was predicted to have a destabilizing change of 0.20 in binding free energy, whereas 267I>T had a stabilizing change of –0.08.

The ML tree constructed from these amino acid sequences can be found in Figure 4. Designations of the obtained cDNA samples were similar to the initial cladogram in terms of clustering (Figure 2). The putative *DXMT* samples (F5, G1, G2, G4, and G5) still clustered together with all the available *DXMT* sequences. Similarly, the putative *XMT* samples (E3, E4, E5, F1, F3, and F4) also clustered with the known *XMT* sequences. The putative *MXMT* samples (E1, E2, and F2) still clustered together in between the available *XMT* and *MXMT* sequences.

Comparing the putative *MXMT* cDNA samples with the putative *MXMT* gDNA samples, their sequences were considerably different and could be assumed to be different genes (Figure 5). The function of these putative *MXMT* mRNAs (E1, E2, and F2) can only be speculated for now. Future binding experiments, as well as protein crystallography, might be able to help us determine its specific substrate whether it is xanthosine like XMT or 7-methylxanthine like MXMT (McCarthy and McCarthy 2007).

For *XMT*, the spliced putative *XMT* gDNAs (C3, C12, and C15) clustered with the putative *XMT* cDNAs (F1, F3, F4, E3, E4, and E5) (Figure 5). Looking into their sequences, the mutations present were most likely due to heterozygous allele differences. It is highly likely that cDNAs F1, F3, E3, E4, and E5 were derived from the sequenced gDNAs (C3, C12, and C15). Although the E4 cDNA clustered together with the others, it still possessed several base mutations. Some of these base substitutions led to missense mutations such as *Ala23Ser, Leu27Met*, and *Val28Phe*, to name a few. This raises the possibility that this transcript might not have been derived from the gDNAs sequenced.

For *DXMT*, only one putative *DXMT* cDNA clustered with the spliced putative gDNA (R1, R3, and R4) (Figure 5). These samples also clustered close to the *C. eugenioides*-derived *DMXT* of *C. arabica*. The other putative *DXMT* cDNAs (F5, G1, G4, and G5) clustered closer to the *C. canephora DMXT* and the *C. canephora*-derived *DMXT* of *C. arabica*. These transcripts were most likely derived from a genomic DNA different from the one sequenced in this study.

In summary, there is one *XMT* gene in *C. liberica* genetically similar to the *C. eugenioides*-derived *XMT* gene of *C. arabica*. There is also a possibility that another *XMT* gene exists from which the E4 cDNA was derived. However, this cannot be ascertained until a more rigorous

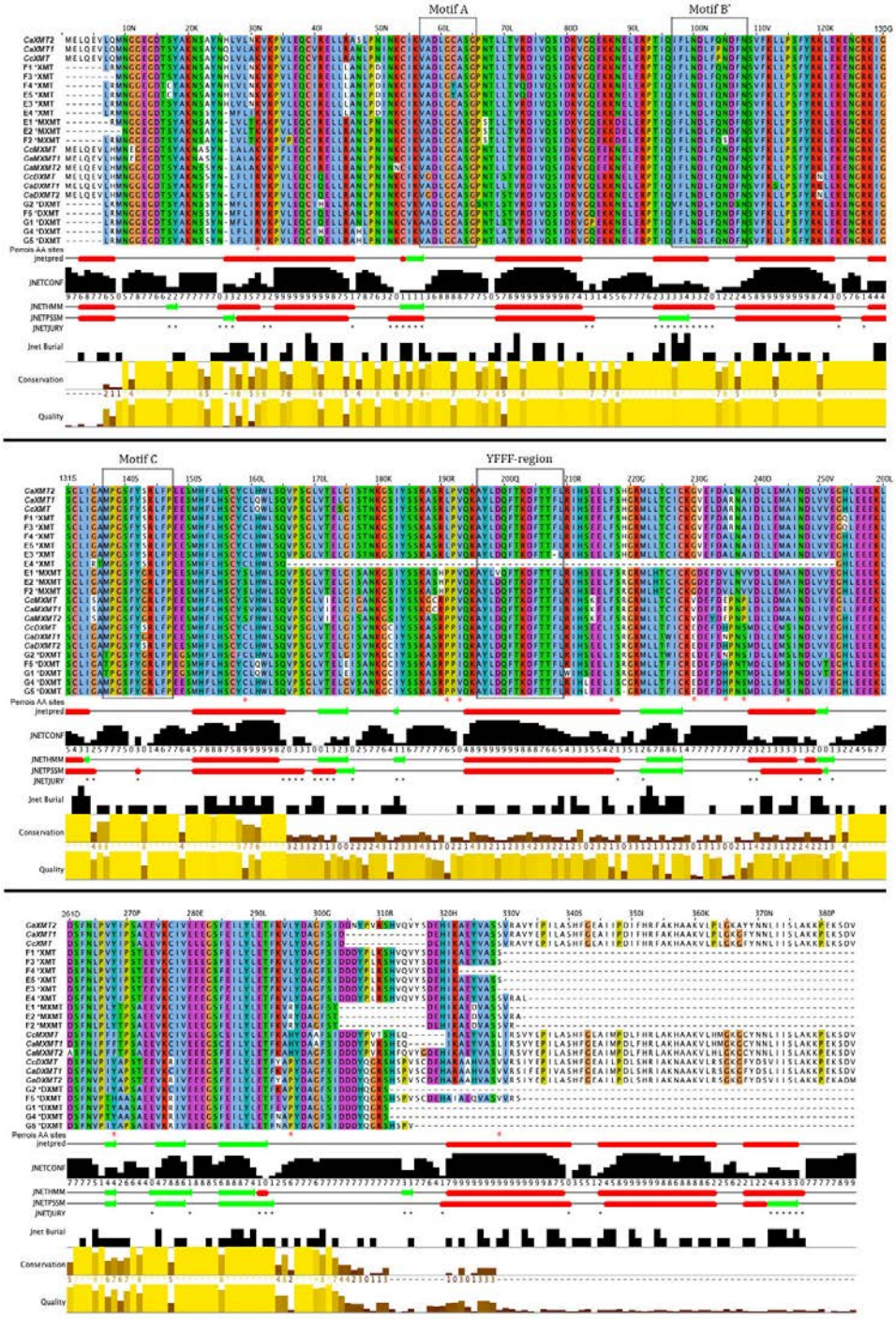


Figure 3. Multiple sequence alignment of coding sequences of *N*-methyltransferase genes from *Coffea liberica* var. *liberica* together with reference sequences of *C. arabica* and *C. canephora* from Perrois *et al.* (2015) (Waterhouse *et al.* 2009). Amino acid positions of interest indicated in the paper of Perrois *et al.* (2015) are indicated with a red asterisk. Results of jnetpred, JNETHMM, and JNETPSSM – which predict the secondary structure of the protein – are also included (Drozdetskiy *et al.* 2015). SAM-binding motifs (A, B', and C) and the conserved YFFF region are highlighted by the gray boxes (McCarthy and McCarthy 2007). Numbered clones labeled with E, F, and G were amplified using the CAF2-ATG and their respective reverse primers – namely, CcXMT1_R, CcMXMT1_R, and CcDXMT_R2. GenBank accession numbers are as follows: *CcXMT1* (JX978518); *CaXMT1* (JX978521); *CaXMT2* (JX978522); *CcMXMT1* (JX978517); *CaMXMT1* (JX978519); *CaMXMT2* (JX978520); *CcDXMT* (JX978516); *CaDXMT* (KF678863); *CaDXMT2* (KJ577793).

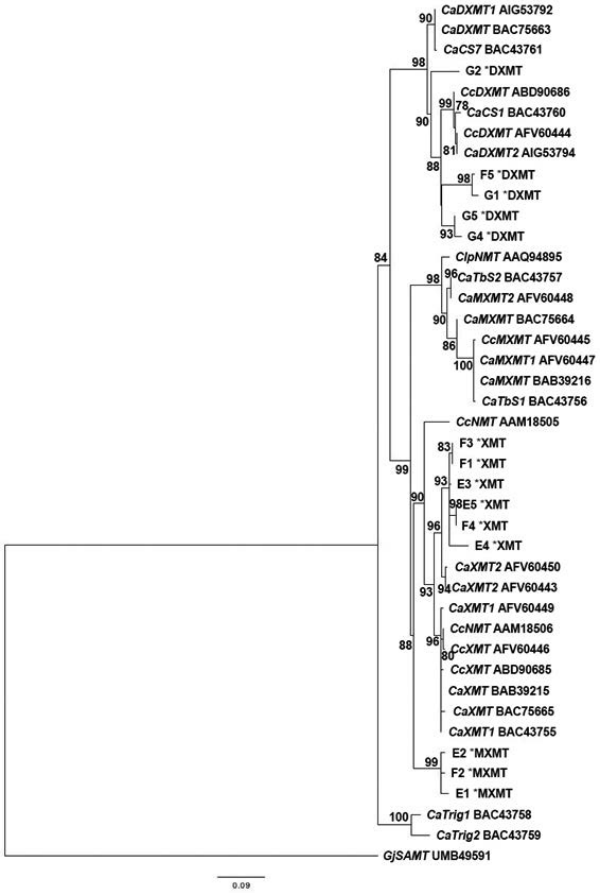


Figure 4. Maximum likelihood tree of the protein sequences of *N*-methyltransferase genes from *Coffea* based on 329 amino acids of *C. liberica* var. *liberica* sequences (numbered clones labeled E, F, and G) obtained from this study and all the available protein sequences of *Coffea* from the NCBI database using the JTT+G model of substitution. The tree is rooted in the salicylic acid methyltransferase (*SAMT*) of *Gardenia jasminoides* (Rubiaceae). Values on nodes represent bootstrap percentages out of 1000 bootstrap samples. Scale bars represent nine amino acid substitutions per 100 amino acids. Numbered clones labeled with E, F, and G were amplified using the CAF2-ATG and their respective reverse primers – namely, CcXMT1_R, CcMXMT1_R, and CcDXMT_R2. Legend: [Ca] *C. arabica*; [Cc] *C. canephora*; [Cl] *C. liberica*; [Gj] *G. jasminoides*; [DXMT] dimethylxanthine methyltransferase; [CS] caffeine synthase; [NMT] *N*-methyltransferase; [TbS] theobromine synthase; [MXMT] 7-methylxanthine-*N*-methyltransferase; [XMT] xanthosine methyltransferase; [TrigS] trigonelline synthase.

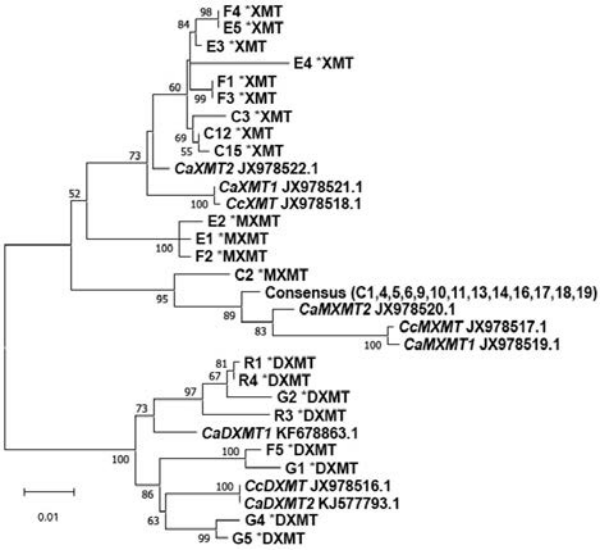


Figure 5. Maximum likelihood tree of the spliced genomic sequences (C and R) and coding sequences (E, F, and G) of *N*-methyltransferase genes from *Coffea* based on 924 nucleotide positions of *C. liberica* var. *liberica* sequences obtained from this study and reference sequences of *C. arabica* and *C. canephora* from Perrois *et al.* (2015) using the general time reversible (GTR) model of DNA substitution. The tree is rooted in the dimethylxanthine *N*-methyltransferase (*DXMT*) genes. Values on nodes represent bootstrap percentages out of 1000 bootstrap samples. Scale bar represents one amino acid substitution per 100 amino acids. Numbered clones labeled with C were amplified using nonspecific NMT primers – namely, CAF2-ATG and CAF2-STOP – whereas clones labeled with R were amplified using the CAF2-ATG and CcDXMT_R2 primers. Numbered clones labeled with E, F, and G were amplified using the CAF2-ATG and their respective reverse primers – namely, CcXMT1_R, CcMXMT1_R, and CcDXMT_R2. Legend: [Ca] *C. arabica*; [Cc] *C. canephora*; [XMT] xanthosine methyltransferase; [MXMT] 7-methylxanthine-*N*-methyltransferase; [DXMT] dimethylxanthine methyltransferase.

sequencing attempt like whole genome sequencing of *C. liberica* is done. There is also one *MXMT* gene in *C. liberica* genetically similar to the *C. eugenioides*-derived *MXMT* gene of *C. arabica*. However, there seem to be two other genes whose sequences are in between both *XMT* and *MXMT* genes – the one in C2 and the one where the E1, E2, and F2 cDNAs were derived from. Their function in the caffeine biosynthesis is yet to be determined. Lastly, there is also one *DXMT* gene in *C. liberica* genetically similar to the *C. eugenioides*-derived *DXMT* gene of *C. arabica*. However, there seems to be another *DMXT* from which the F5, G1, G4, and G5 cDNAs were derived. Presence of duplicated NMT genes is not unlikely, as this is the hypothesized origin of caffeine biosynthesis in *Coffea* (Denoeud *et al.* 2014). A recent study by Raharimalala *et al.* (2021) discovered numerous tandem duplications of *NMT* genes, some of which were designated as pseudogenes, in the genomes of *C. canephora* and *C. humblotiana*. A high percentage of transposable elements, as high as 63.4% in one region, have been found in these *NMT* regions (Raharimalala *et al*. 2021). Whether or not

these other NMT genes provide the same function as those established is still to be determined.

Expression of N-methyltransferases (NMTs) during Fruit Development in *C. liberica*

Synthesis of cDNA from the various tissue samples was successful as evidenced by the amplification of the *GAPDH* and *RPL39* reference genes (Appendix Figures IIA and B). For the *DXMT*-specific primer, faint bands were observed in the leaf and flower bud tissues, whereas bright bands were seen in the green fruit tissues (Figure 6C). The *XMT* and *MXMT*-specific primers had similar expression profiles. Bright bands were observed in the leaf, flower bud, and green fruit tissues (Figures 6A and B). High gene expression in the leaves is in congruence with the caffeine content data of Arabica and Robusta, as the highest caffeine content was found in young leaves (Perrois *et al*. 2015). Caffeine was also observed to significantly accumulate during bean development, peaking at the large green grain stage (Perrois *et al*. 2015). High expression of *XMT*, *MXMT*, and *DXMT* – as suggested by the bright bands in the green fruit tissues – mirrors this increase in caffeine content. Bands in the red fruit tissue samples were mostly absent except for faint bands in one and two samples for *XMT* (Figure 6A) and *MXMT* (Figure 6B), respectively. This suggests that XMT, MXMT, and DXMT expression was minimal in the red fruit stage, hence leading to minimal caffeine synthesis. A similar trend was also observed in the study of Perrois *et al*. (2015), as caffeine content was found to decrease as the fruit ripens from the large green grain stage to the red grain stage.

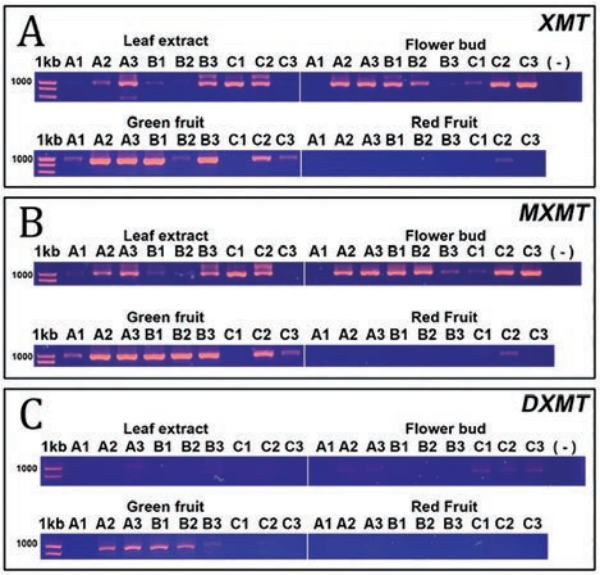


Figure 6. Agarose gel showing PCR products after amplifying the synthesized cDNA from leaf, flower bud, green fruit, and red fruit extracts using primers specific for xanthosine methyltransferase (*XMT*) (A), 7-methylxanthine-*N*-methyltransferase (*MXMT*) (B), and dimethylxanthine methyltransferase (*DXMT*) (C). Samples were obtained from individual *Coffea liberica* var. *liberica* trees (A, B, and C) in triplicate. A 1000-bp (1 kb) ladder was used as standard, and a negative control (–) was also included.

CONCLUSION

Caffeine biosynthesis in *Coffea* evolved independently from other lineages, thus highlighting the importance of expanding research from just *C. arabica* and *C. canephora*. This study focused on the NMTs of *C. liberica*. The genomic sequences of the *XMT*, *MXMT*, and *DXMT* genes have been sequenced in this study. The NMTs of *C. liberica* were genetically distinguishable from *C. arabica* and *C. canephora*, especially in the 3rd intron and 4th exon of the genes.

This study has also partially sequenced the coding sequences of the same genes. Key amino acid positions of the NMTs were mostly conserved in *C. liberica*. While there were a few differences in the amino acid sequences, these substitutions do not seem to disrupt substrate interaction. Interestingly, three clones (E1, E2, and F2) of NMT transcripts were found to cluster in between the *XMT* and *MXMT* clusters based on the ML tree. The function and binding specificity of the product of these transcripts are yet to be determined. Another clone (E4) was missing the entire 3rd exon, which suggests that it might be a splice variant. Several secondary structures are present in this region, which might make the product of this transcript non-functional.

The NMT expression levels of *C. liberica* were also observed to be similar to *C. arabica* and *C. canephora*. Expression of NMTs was high at the green fruit developmental stage and diminishes to very low imperceptible levels once the fruit has ripened.

The results of this study show how similar and different the NMTs of *C. liberica* are from *C. arabica* and *C. canephora*. Understanding the origin and function of these novel homologous genes and transcripts will help us more deeply understand the evolution of biosynthesis in *Coffea*.

RECOMMENDATIONS

The use of newer sequencing methods might provide a more thorough representation of all the NMTs in *C. liberica*. Discovery of duplicated NMT genes in the genome, as well as all the splice variants of these NMT genes, is possible with newer technologies. Due to the high sequence similarity of the different NMTs, the use

of technology with high specificity and sensitivity such as the Taqman probe used by Perrois *et al.* (2015) might be able to provide more accurate expression results. Expanding the scope to other *Coffea* species will also further strengthen the assumed conserved and critical regions in the different NMT genes.

ACKNOWLEDGMENTS

The authors would like to acknowledge the support of the Natural Sciences and Research Institute, College of Science, UP Diliman (Project Code: BIO-22-1-02), as well as the Institute of Biology, College of Science, UP Diliman.

STATEMENT ON CONFLICT OF INTEREST

The authors declare no conflict of interest.

NOTES ON APPENDICES

The complete appendices section of the study is accessible at https://philjournsci.dost.gov.ph

REFERENCES

ANTHONY F, CLIFFORD M, NOIROT M. 1993. Biochemical diversity in the genus *Coffea* L.: chlorogenic acids, caffeine, and mozambioside contents. Genetic Resources and Crop Evolution 40: 61–70.

ARCILA-PULGARIN J, BUHR, L, BLEIHOLDER H, HACK, H, UMEIR, U, WICKE H. 2002. Application of the extended BBCH scale for the description of the stages of coffee (*Coffea spp*.). Annals of Applied Biology 141(1): 19–27.

ARGOUT X, SALSE J, AURY JM, GUILTINAN MJ, DROC G, GOUZY J, ALLEGRE M, CHAPARRO C, LEGAVRE T, MAXIMOVA SN, ABROUK M, MURAT F, FOUET O, POULAIN J, RUIZ M, ROGUET Y, RODIER-GOUD M, BARBOSA-NETO JF, SABOT F, KUDRNA D, AMMIRAJU JS, SCHUSTER SC, CARLSON JE, SALLET E, SCHIEX T, DIEVART A, KRAMER M, GELLEY L, SHI Z, BÉRARD A, VIOT C, BOCCARA M, RISTERUCCI AM, GUIGNON V, SABAU X, AXTELL MJ, MA Z, ZHANG Y, BROWN S, BOURGE M, GOLSER W, SONG X, CLEMENT D, RIVALLAN R, TAHI M, AKAZA JM, PITOLLAT B, GRAMACHO K, D'HONT A, BRUNEL D, INFANTE D, KEBE I, COSTET P, WING R, MCCOMBIE WR, GUIDERDONI E, QUETIER F, PANAUD O, WINCKER P, BOCS S, LANAUD C. 2011. The genome of *Theobroma cacao*. Nature Genetics 43(2): 101–108.

ASHIHARA H, MONTEIRO A, GILLIES F, CROZIER A. 1996. Biosynthesis of caffeine in leaves of coffee. Plant Physiology 111: 747–753.

CHARR JC, GARAVITO A, GUYEUX C, CROUZILLAT D, DESCOMBES P, FOURNIER C, LY SN, RAHARIMALALA EN, RAKOTOMALALA JJ, STOFFELEN P, JANSSENS S, HAMON P, GUYOT R. 2020. Complex evolutionary history of coffees revealed by full plastid genomes and 28,800 nuclear SNP analyses, with particular emphasis on *Coffea canephora* (Robusta coffee). Molecular Phylogenetics and Evolution 151: 106906.

CRUZ F, KALAOUN S, NOBILE P, COLOMBO C, ALMEIDA J, BARROS LMG, ROMANO E, GROSSI-DE-SA MF, VASLIN M, ALVES-FERREIRA M. 2009. Evaluation of coffee reference genes for relative expression studies by quantitative real-time RT-PCR. Molecular Breeding 23: 607–616.

DENOEUD F, CARRETERO-PAULET L, DEREEPER A, DROC G, GUYOT R, PIETRELLA M, ZHENG C, ALBERTI A, ANTHONY F, APREA G, AURY J, BENTO P, BERNARD M, BOCS S, CAMPA C, CENCI A, COMBES M, CROUZILLAT D, DA SILVA C, DADDIEGO L, DE BELLIS F, DUSSERT S, GARSMEUR O, GAYRAUD T, GUIGNON V, JAHN K, JAMILLOUX V, JOËT T, LABADIE K, LAN T, LECLERCQ J, LEPELLEY M, LEROY T, LI L, LIBRADO P, LOPEZ L, MUÑOZ A, NOEL B, PALLAVICINI A, PERROTTA G, PONCET V, POT D, PRIYONO, RIGOREAU M, ROUARD M, ROZAS J, TRANCHANT-DUBREUIL C, VANBUREN R, ZHANG Q, ANDRADE A, ARGOUT X, BERTRAND B, DE KOCHKO A, GRAZIOSI G, HENRY R, JAYARAMA, MING R, NAGAI C, ROUNSLEY S, SANKOFF D, GIULIANO G, ALBERT V, WINCKER P, LASHERMES P. 2014. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. Science 345: 1181–1184.

DROZDETSKIY A, COLE C, PROCTER J, BARTON G. 2015. JPred4: a protein secondary structure prediction server. Nucleic Acids Research 43(W).

GUINDON S, DUFAYARD JF, LEFORT V, ANISIMOVA M, HORDIJK W, GASCUEL O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology 59(3): 307–321.

HAMON P, GROVER CE, DAVIS AP, RAKOTOMALALA JJ, RAHARIMALALA NE, ALBERT VA, SREENATH HL, STOFFELEN P, MITCHELL SE, COUTURON E, HAMON S, DE KOCHKO A, CROUZILLAT D, RIGOREAU M, SUMIRAT U, AKAFFOU S, GUYOT R. 2017. Genotyping-by-sequencing provides the first well-resolved phylogeny for coffee (*Coffea*) and insights into the evolution of caffeine content in its species: GBS coffee phylogeny and the evolution of caffeine content. Molecular Phylogenetics and Evolution 109: 351–361.

HUANG X, BAI X, XIE Z, FAHAD S, GBOKIE T, LU Y, GUO T, LI J, ZHANG Z, WU W, YI K. 2023. *De novo* transcriptome assembly of *Coffea liberica* reveals phylogeny and expression atlas of *phenylalanine ammonia-lyase* genes in *Coffea* species. Industrial Crops and Products 192: 116029.

KIM YS, CHOI YE, SANO H. 2010. Plant vaccination: Stimulation of defense system by caffeine production in planta. Plant Signaling & Behavior 5(5): 489–493.

MCCARTHY AA, MCCARTHY JG. 2007. The structure of two *N*-methyltransferases from the caffeine biosynthetic pathway. Plant Physiology 144: 879–889.

MOISYADI S, NEUPANE KR, STILES JI. 1998. Cloning and characterization of a cDNA encoding xanthosine-N^7-methyltransferase from coffee (*Coffea arabica*). Acta Horticulturae 461: 367–377.

OGAWA M, HERAI Y, KOIZUMI N, KUSANO T, SANO H. 2001. 7-Methylxanthine methyltransferase of coffee plants. Journal of Biological Chemistry 276: 8213–8218.

PAHARI S, LI G, MURTHY AK, LIANG S, FRAGOZA R, YU H, ALEXOV E. 2020. SAAMBE-3D: predicting effect of mutations on protein-protein interactions. International Journal of Molecular Sciences 21(7): 2563.

[PCBI] Philippine Coffee Board Inc. 2020. Everything You Need to Know about Liberica, the Unique Philippine Coffee. Retrieved on 26 Sep 2023 from https://philcoffeeboard.com/everything-you-need-to-know-about-liberica-the-unique-philippine-coffee/

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(1): 179−191.

POTAPOV V, ONG J. 2017. Examining sources of error in PCR by single-molecule sequencing. PLoS ONE 12(1): e0169774.

RAHARIMALALA N, ROMBAUTS S, MCCARTHY A, GARAVITO A, OROZCO-ARIAS S, BELLANGER L, MORALES-CORREA AY, FROGER S, MICHAUX S, BERRY V, S METAIRON S, FOURNIER C, LEPELLEY M, MUELLER L, COUTURON E, HAMON P, RAKOTOMALALA J-J, DESCOMBES P, GUYOT R, CROUZILLAT D. 2021. The absence of the caffeine synthase gene is involved in the naturally decaffeinated status of *Coffea humblotiana*, a wild species from Comoro archipelago. Scientific Reports 11: 8119.

REYES CM, CORNELIS MC. 2018. Caffeine in the diet: country-level consumption and guidelines. Nutrients 10(11): 1772.

STECHER G, TAMURA K, KUMAR S. 2020. Molecular Evolutionary Genetics Analysis (MEGA) for macOS. Molecular Biology and Evolution.

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 Physiology 132: 372–380.

WATERHOUSE AM, PROCTER JB, MARTIN DMA, CLAMP M, BARTON GJ. 2009. Jalview version 2: A multiple sequence alignment and analysis workbench. Bioinformatics 25(9): 1189–1191.

XIA E, ZHANG H, SHENG J, LI K, ZHANG Q, KIM C, ZHANG Y, LIU Y, ZHU T, LI W, HUANG H, TONG Y, NAN H, SHI C, SHI C, JIANG J, MAO S, JIAO J, ZHANG D, ZHAO Y, ZHAO Y, ZHANG L, LIU Y, LIU B, YU Y, SHAO S, NI D, EICHLER E, GAO L. 2017. The tea tree genome provides insights into tea flavor and independent evolution of caffeine biosynthesis. Molecular Plant 10(6): 866−877.

XU Z, PU X, GAO R, DEMURTAS OC, FLECK SJ, RICHTER M, HE C, JI A, SUN W, KONG J, HU K, REN F, SONG J, WANG Z, GAO T, XIONG C, YU H, XIN T, ALBERT VA, GIULIANO G, CHEN S, SONG J. 2020. Tandem gene duplications drive divergent evolution of caffeine and crocin biosynthetic pathways in plants. BMC Biology 18(1): 63.

ZHOU MZ, YAN CY, ZENG Z, LUO L, ZENG W, HUANG YH. 2020. *N*-methyltransferases of caffeine biosynthetic pathway in plants. Journal of Agricultural and Food Chemistry 68(52): 15359–15372.

APPENDICES

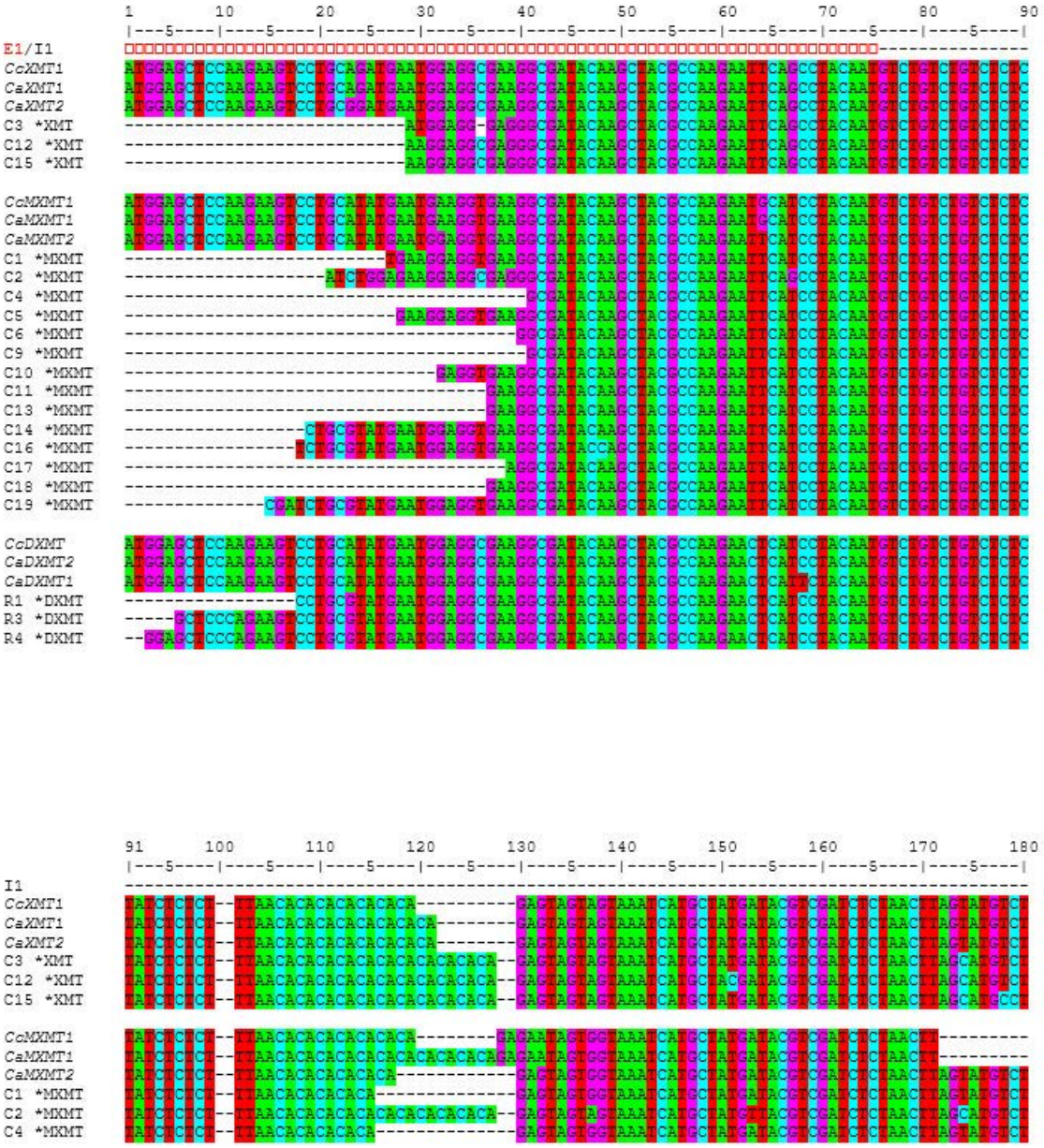
Table I. Pairwise distance values between sequences from genomic *N*-methyltransferases of *C. liberica* var. *liberica* and the reference sequences from Perrois *et al.* (2015) (highlighted blue). Sequences labeled with C were amplified using non-specific NMT primers – namely, CAF2-ATG and CAF2-STOP. Sequences labeled with R were amplified using the CAF2-ATG and the DXMT-specific reverse primer, *CcDXMT*_R2. Putative NMT designation is indicated alongside the label with an asterisk (*) based on the reference sequence with the lowest pairwise distance value (highlighted yellow). GenBank accession numbers are as follows: *CcXMT1* (JX978518); *CaXMT1* (JX978521); *CaXMT2* (JX978522); *CcMXMT1* (JX978517); *CaMXMT1* (JX978519); *CaMXMT2* (JX978520); *CcDXMT* (JX978516); *CaDXMT* (KF678863); *CaDXMT2* (KJ577793).

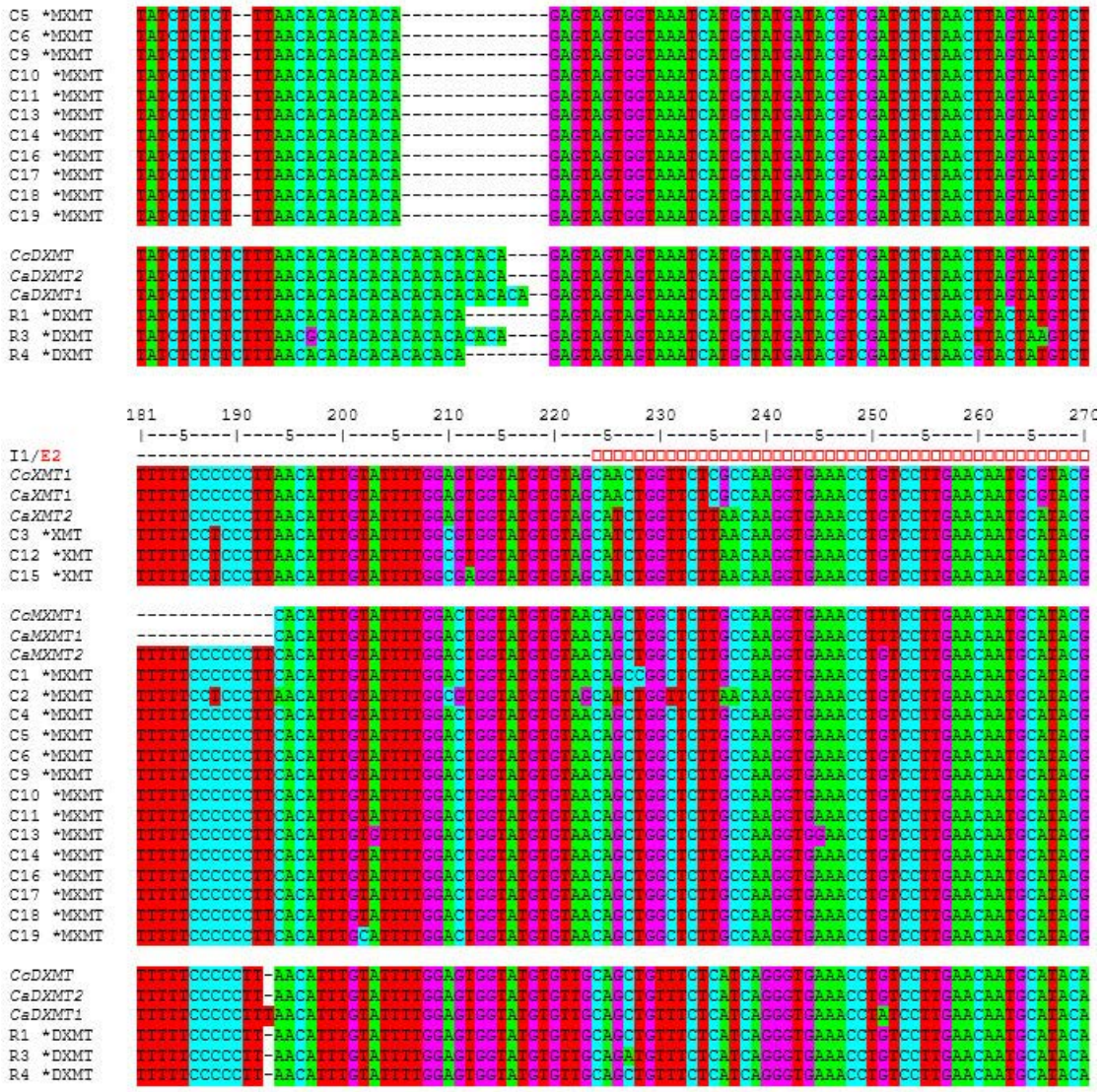
	CcXMT1	*CaXMT1*	*CaXMT2*	C3*XMT	C12*XMT	C15*XMT	*CcMXMT1*	*CaMXMT1*	*CaMXMT2*	C1*MXMT	C2*MXMT	C4*MXMT	C5*MXMT	C6*MXMT
CcXMT1														
CaXMT1	0.00151													
CaXMT2	0.01803	0.01756												
C3*XMT	0.03389	0.03344	0.02481											
C12*XMT	0.02653	0.02606	0.01779	0.00929										
C15*XMT	0.03206	0.03161	0.02310	0.01502	0.00717									
CcMXMT1	0.05996	0.05889	0.06155	0.07644	0.07151	0.07748								
CaMXMT1	0.06700	0.06591	0.06783	0.08136	0.07709	0.08239	0.00578							
CaMXMT2	0.05004	0.04969	0.04839	0.06132	0.05649	0.06231	0.02717	0.03244						
C1*MXMT	0.06254	0.06258	0.05955	0.06624	0.06195	0.06744	0.04342	0.04009	0.02596					
C2*MXMT	0.05852	0.05848	0.05204	0.04472	0.03937	0.04459	0.05823	0.05608	0.04323	0.02752				
C4*MXMT	0.06026	0.06030	0.05809	0.06394	0.05968	0.06665	0.04530	0.04277	0.02820	0.01168	0.02911			
C5*MXMT	0.06041	0.06045	0.05748	0.06484	0.06056	0.06603	0.04065	0.03791	0.02467	0.00727	0.02559	0.01044		
C6*MXMT	0.05748	0.05752	0.05462	0.06042	0.05820	0.06391	0.03899	0.03394	0.02231	0.00612	0.02329	0.00798	0.00489	
C9*MXMT	0.05804	0.05808	0.05517	0.06174	0.05755	0.06303	0.03888	0.03603	0.02290	0.00671	0.02259	0.00982	0.00549	0.00428
C10*MXMT	0.05706	0.05709	0.05422	0.06220	0.05867	0.06414	0.03796	0.03508	0.02151	0.00546	0.02377	0.00859	0.00424	0.00305
C11*MXMT	0.05792	0.05796	0.05575	0.06443	0.06022	0.06572	0.03880	0.03668	0.02351	0.00732	0.02511	0.01045	0.00609	0.00489
C13*MXMT	0.05816	0.05819	0.05527	0.06178	0.05956	0.06527	0.03966	0.03684	0.02295	0.00673	0.02456	0.00861	0.00550	0.00428
C14*MXMT	0.05864	0.05867	0.05510	0.06269	0.05991	0.06538	0.04038	0.03764	0.02323	0.00850	0.03002	0.00982	0.00667	0.00427
C16*MXMT	0.06083	0.06087	0.05722	0.06417	0.06138	0.06691	0.04112	0.03842	0.02584	0.01036	0.03201	0.01170	0.00851	0.00612
C17*MXMT	0.05587	0.05591	0.05306	0.06166	0.05747	0.06299	0.03747	0.03455	0.02094	0.00487	0.02257	0.00797	0.00365	0.00244
C18*MXMT	0.05579	0.05583	0.05299	0.06162	0.05743	0.06291	0.03742	0.03450	0.02092	0.00487	0.02254	0.00797	0.00364	0.00244
C19*MXMT	0.06067	0.06071	0.05707	0.06269	0.05991	0.06538	0.04170	0.03904	0.02448	0.00789	0.02937	0.00920	0.00606	0.00366
CcDXMT	0.10159	0.10171	0.10310	0.11974	0.11189	0.11986	0.10287	0.12418	0.10726	0.10155	0.09995	0.09006	0.09778	0.09501
CaDXMT2	0.10361	0.10374	0.10381	0.11975	0.11192	0.11988	0.10503	0.12647	0.10929	0.10155	0.09996	0.09007	0.09779	0.09502
CaDXMT1	0.09890	0.09861	0.09684	0.11226	0.10549	0.11242	0.08818	0.11047	0.10071	0.09582	0.09564	0.08532	0.09360	0.09013
R1*DXMT	0.06974	0.06930	0.06429	0.07745	0.07357	0.07875	0.08648	0.08642	0.07194	0.08386	0.08161	0.07875	0.08219	0.07869
R3*DXMT	0.07446	0.07404	0.07051	0.08444	0.08038	0.08565	0.09179	0.09003	0.07756	0.09295	0.09200	0.08782	0.09125	0.08775
R4*DXMT	0.06843	0.06799	0.06303	0.07760	0.07217	0.07734	0.08403	0.08381	0.07059	0.08315	0.08090	0.07803	0.08147	0.07797

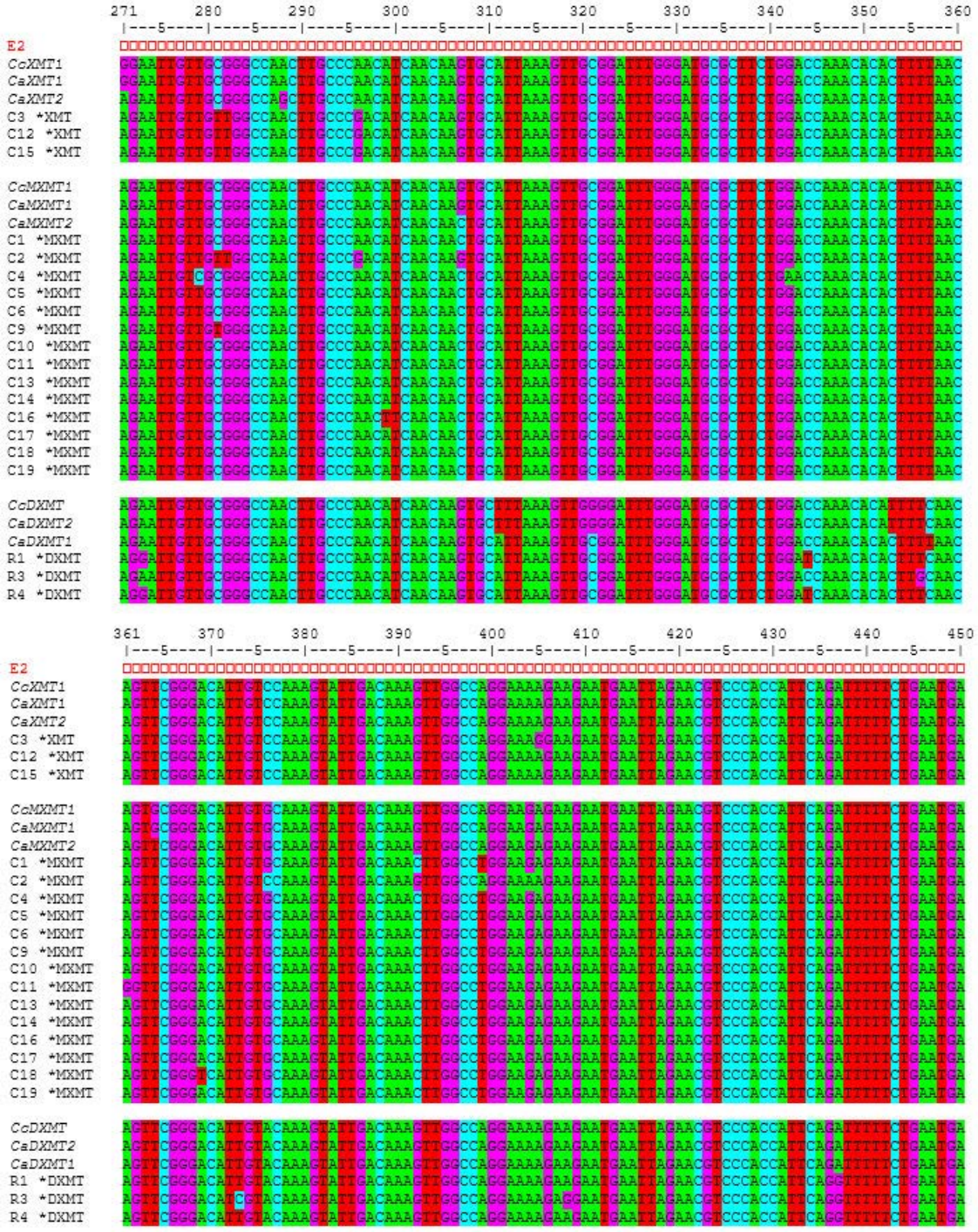
Table I. Cont.

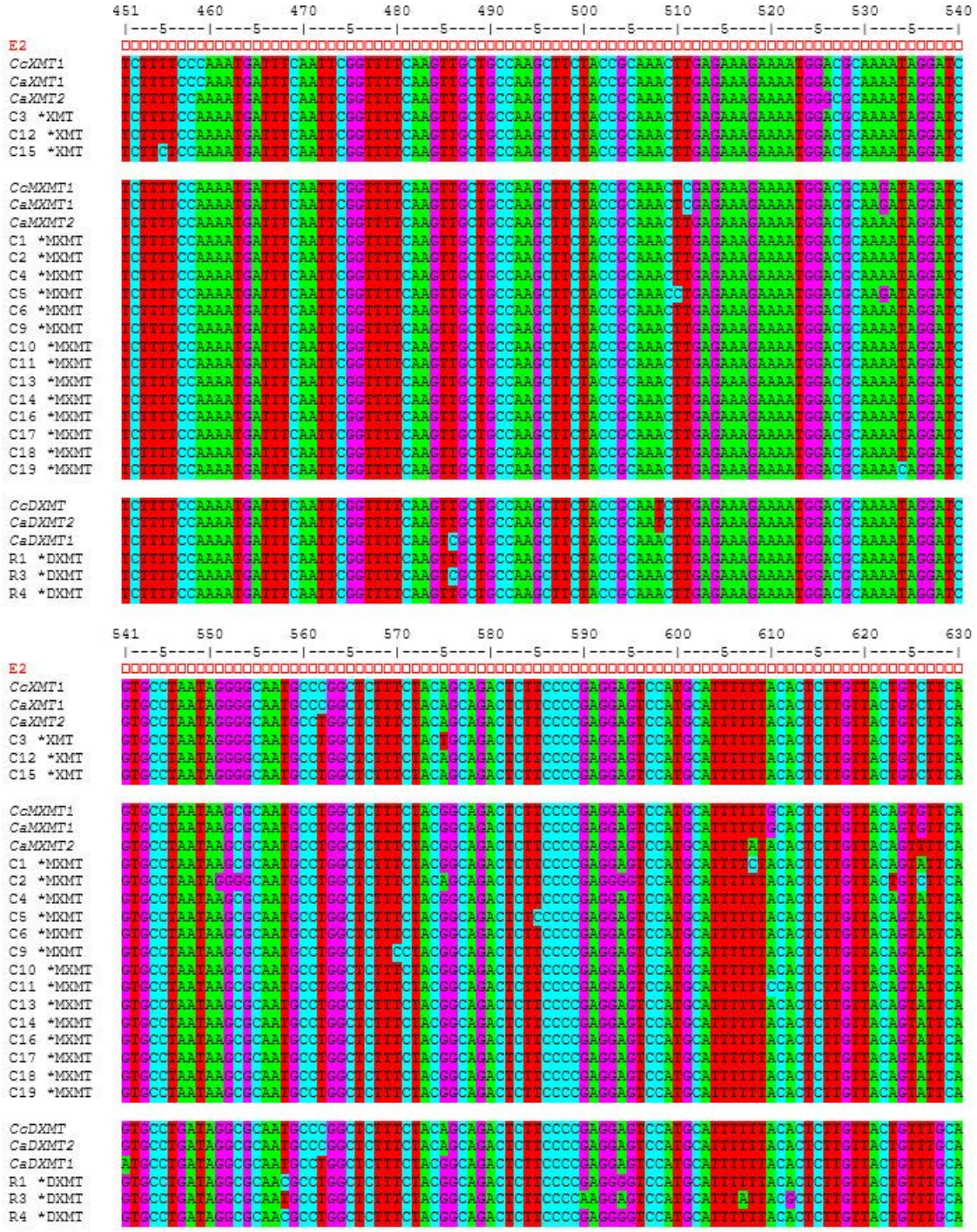
	C9*MXMT	C10*MXMT	C11*MXMT	C13*MXMT	C14*MXMT	C16*MXMT	C17*MXMT	C18*MXMT	C19*MXMT	*CcDXMT*	*CaDXMT2*	*CaDXMT1*	R1*DXMT	R3*DXMT	R4*DXMT
CcXMT1															
CaXMT1															
CaXMT2															
C3*XMT															
C12*XMT															
C15*XMT															
CcMXMT1															
CaMXMT1															
CaMXMT2															
C1*MXMT															
C2*MXMT															
C4*MXMT															
C5*MXMT															
C6*MXMT															
C9*MXMT															
C10*MXMT	0.00366														
C11*MXMT	0.00549	0.00426													
C13*MXMT	0.00490	0.00366	0.00550												
C14*MXMT	0.00487	0.00364	0.00548	0.00488											
C16*MXMT	0.00673	0.00547	0.00733	0.00674	0.00664										
C17*MXMT	0.00304	0.00182	0.00365	0.00305	0.00304	0.00488									
C18*MXMT	0.00304	0.00182	0.00365	0.00305	0.00304	0.00487	0.00121								
C19*MXMT	0.00426	0.00303	0.00487	0.00427	0.00421	0.00603	0.00243	0.00243							
CcDXMT	0.09708	0.09585	0.09763	0.09570	0.09562	0.10024	0.09474	0.09461	0.09847						
CaDXMT2	0.09709	0.09586	0.09764	0.09570	0.09563	0.10025	0.09475	0.09462	0.09848	0.00150					
CaDXMT1	0.09138	0.09019	0.09193	0.09081	0.09075	0.09457	0.08908	0.08896	0.09284	0.05879	0.06055				
R1*DXMT	0.07959	0.07743	0.07940	0.07934	0.07825	0.08162	0.07612	0.07690	0.07902	0.05523	0.05523	0.02432			
R3*DXMT	0.08868	0.08641	0.08846	0.08839	0.08716	0.09061	0.08512	0.08590	0.08955	0.06129	0.06130	0.02807	0.01925		
R4*DXMT	0.07887	0.07671	0.07868	0.07862	0.07754	0.08091	0.07540	0.07618	0.07989	0.05407	0.05408	0.02350	0.00487	0.01786	

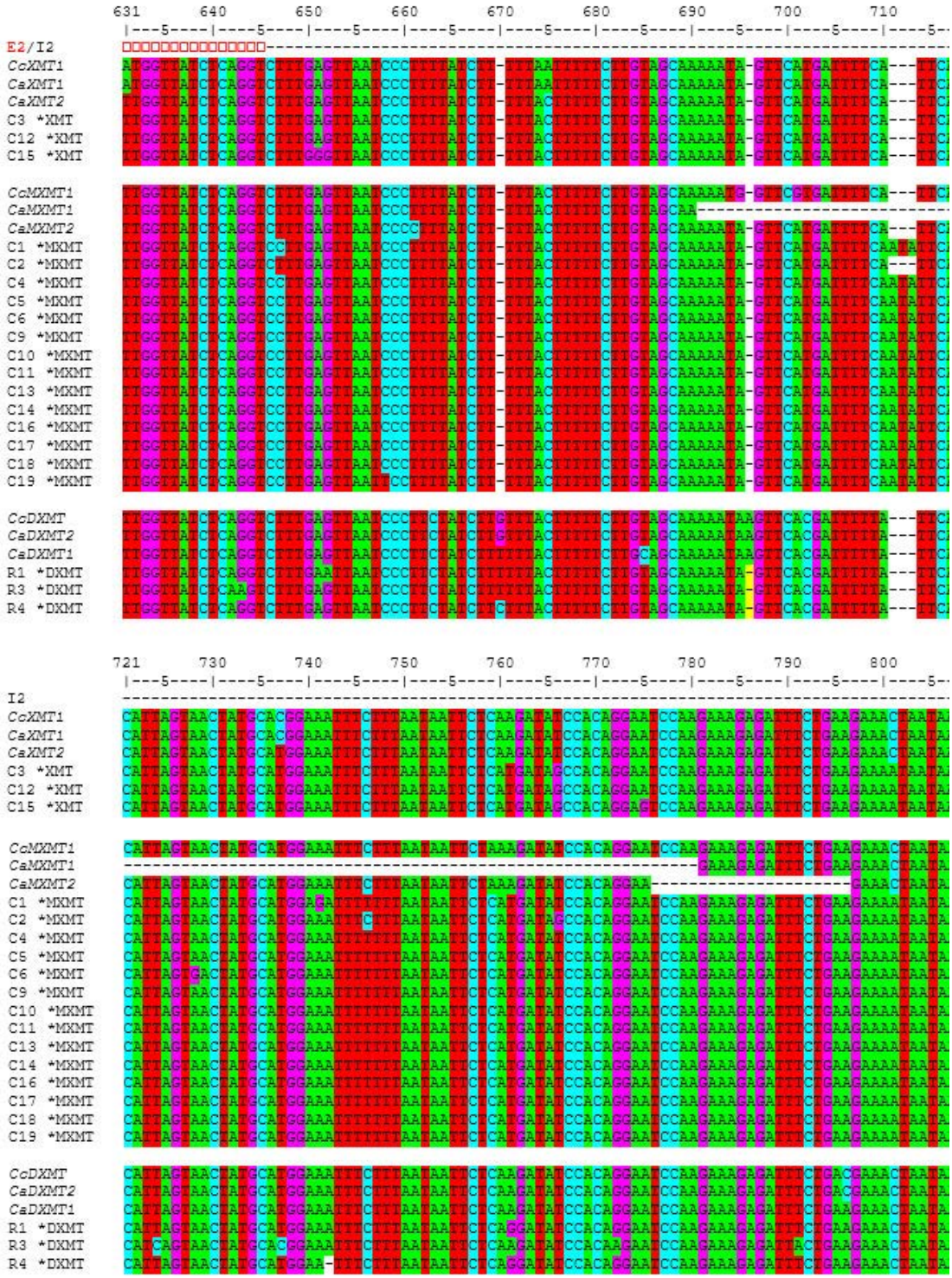
Figure I. Sequence alignment of genomic *N*-methyltransferases from *C. liberica* var. *liberica* with closest reference sequences from Perrois *et al.* (2015). Sequences labeled with C were amplified using non-specific NMT primers – namely, CAF2-ATG and CAF2-STOP. Sequences labeled with R were amplified using the CAF2-ATG and the DXMT-specific reverse primer, *CcDXMT*_R2. Putative NMT designation is indicated alongside the label with an asterisk (*). Exon regions are numbered and indicated as red boxes, whereas intron regions are indicated as black dashes. GenBank accession numbers are as follows: *CcXMT1* (JX978518); *CaXMT1* (JX978521); *CaXMT2* (JX978522); *CcMXMT1* (JX978517); *CaMXMT1* (JX978519); *CaMXMT2* (JX978520); *CcDXMT* (JX978516); *CaDXMT* (KF678863); *CaDXMT2* (KJ577793).

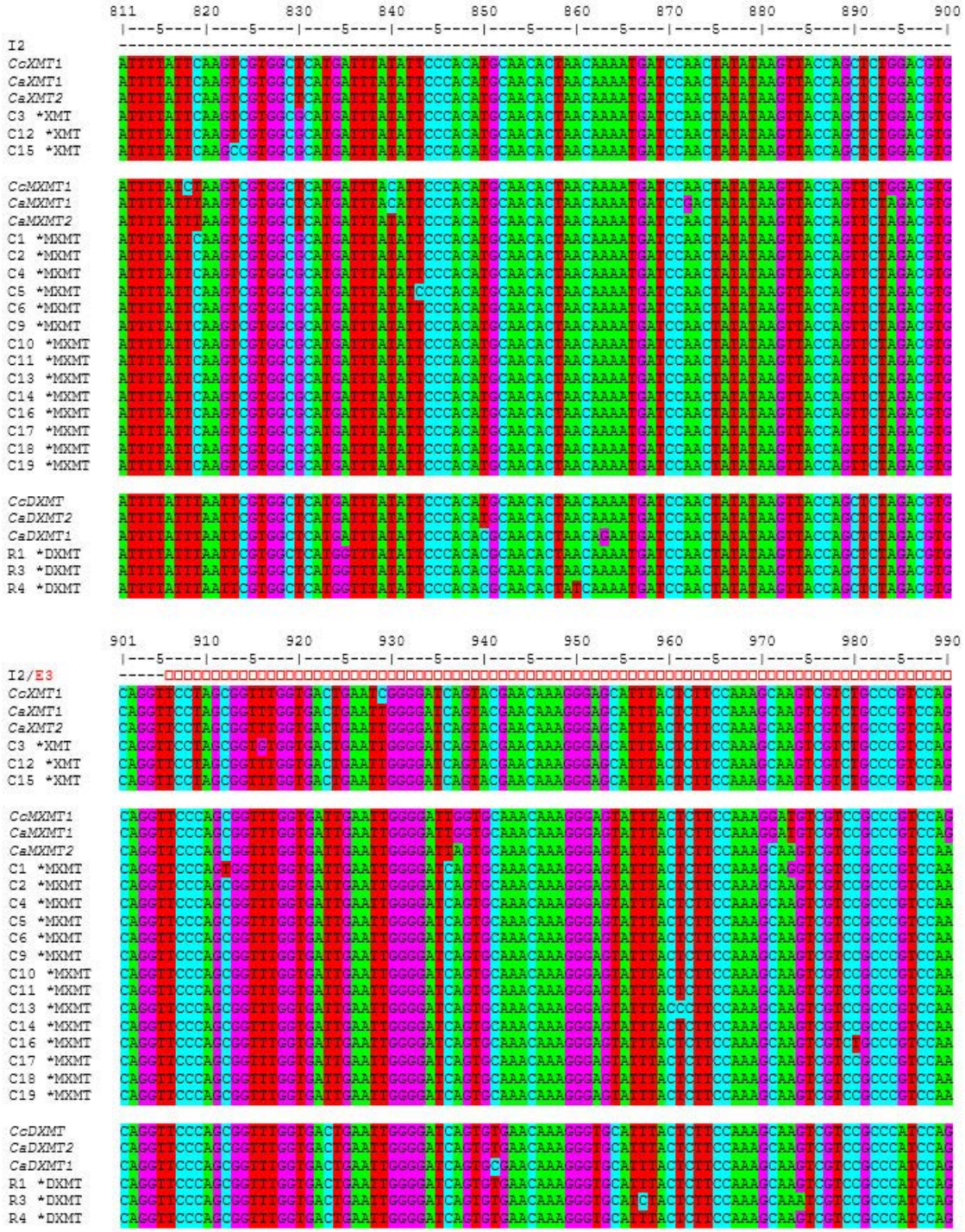


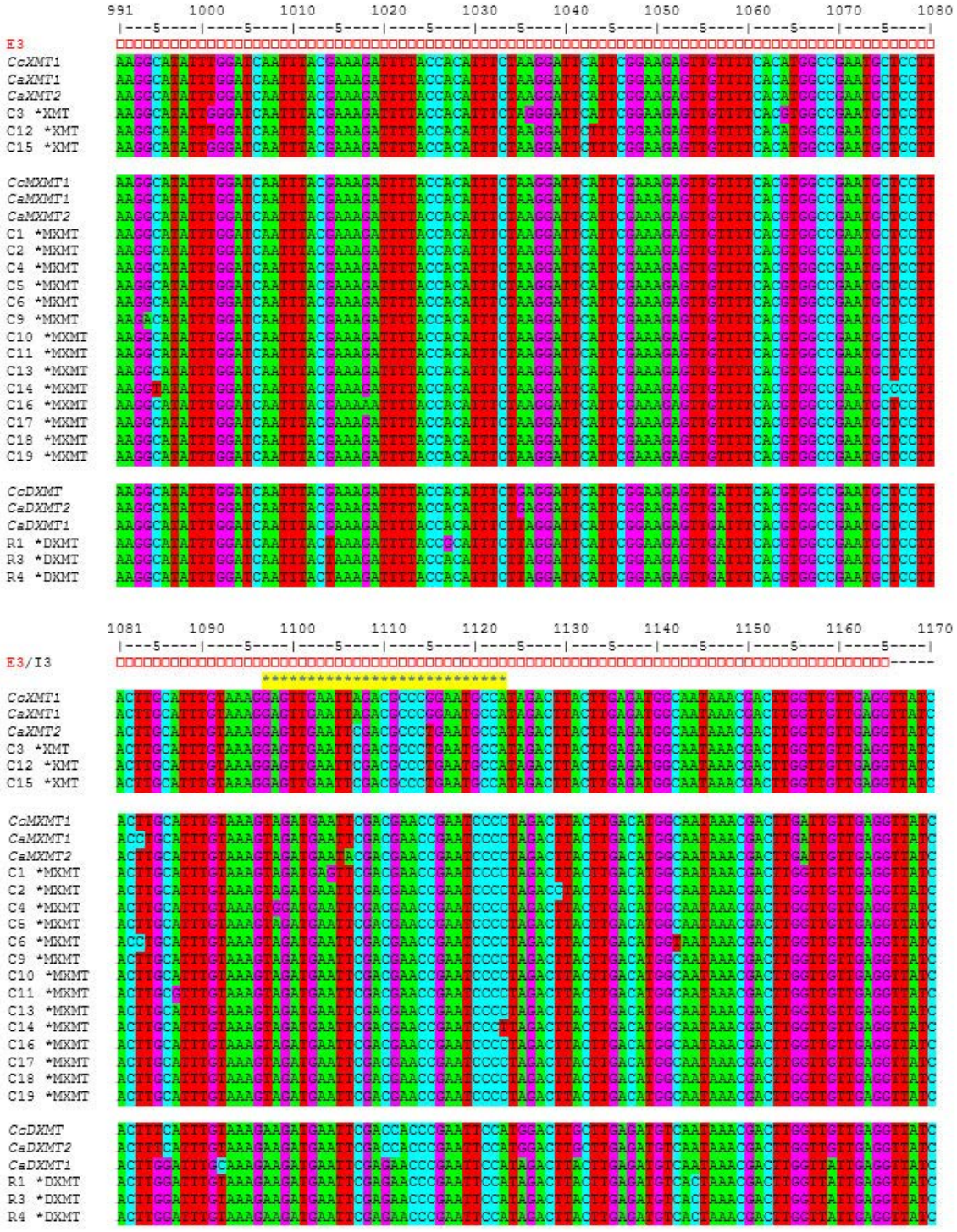

C5 *MXMT
C6 *MXMT
C9 *MXMT
C10 *MXMT
C11 *MXMT
C13 *MXMT
C14 *MXMT
C16 *MXMT
C17 *MXMT
C18 *MXMT
C19 *MXMT
CoDXMT
CaDXMT2
CaDXMT1
R1 *DXMT
R3 *DXMT
R4 *DXMT
181
190
200
210
220
230
240
250
260
I1/E2
CoXMT1
CaXMT1
CaXMT2
C3 *XMT
C12 *XMT
C15 *XMT
CoMXMT1
CaMXMT1
CaMXMT2
C1 *MXMT
C2 *MXMT
C4 *MXMT
C5 *MXMT
C6 *MXMT
C9 *MXMT
C10 *MXMT
C11 *MXMT
C13 *MXMT
C14 *MXMT
C16 *MXMT
C17 *MXMT
C18 *MXMT
C19 *MXMT
CoDXMT
CaDXMT2
CaDXMT1
R1 *DXMT
R3 *DXMT
R4 *DXMT

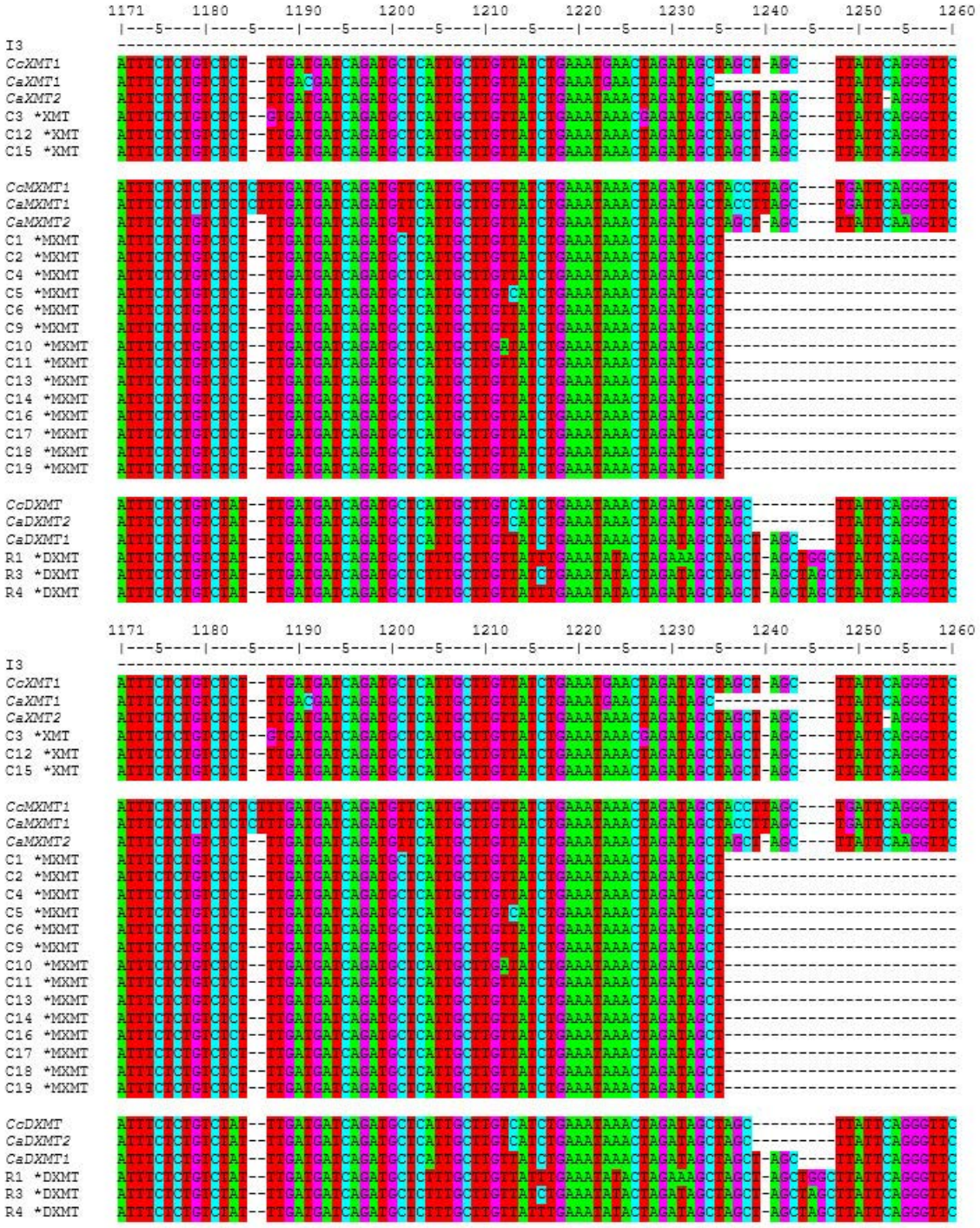

271 280 290 300 310 320 330 340 350 360
E2
CcXMT1
CaXMT1
CaXMT2
C3 *XMT
C12 *XMT
C15 *XMT
CcMXMT1
CaMXMT1
CaMXMT2
C1 *MXMT
C2 *MXMT
C4 *MXMT
C5 *MXMT
C6 *MXMT
C9 *MXMT
C10 *MXMT
C11 *MXMT
C13 *MXMT
C14 *MXMT
C16 *MXMT
C17 *MXMT
C18 *MXMT
C19 *MXMT
CcDXMT
CaDXMT2
CaDXMT1
R1 *DXMT
R3 *DXMT
R4 *DXMT
361 370 380 390 400 410 420 430 440 450
E2
CcXMT1
CaXMT1
CaXMT2
C3 *XMT
C12 *XMT
C15 *XMT
CcMXMT1
CaMXMT1
CaMXMT2
C1 *MXMT
C2 *MXMT
C4 *MXMT
C5 *MXMT
C6 *MXMT
C9 *MXMT
C10 *MXMT
C11 *MXMT
C13 *MXMT
C14 *MXMT
C16 *MXMT
C17 *MXMT
C18 *MXMT
C19 *MXMT
CcDXMT
CaDXMT2
CaDXMT1
R1 *DXMT
R3 *DXMT
R4 *DXMT

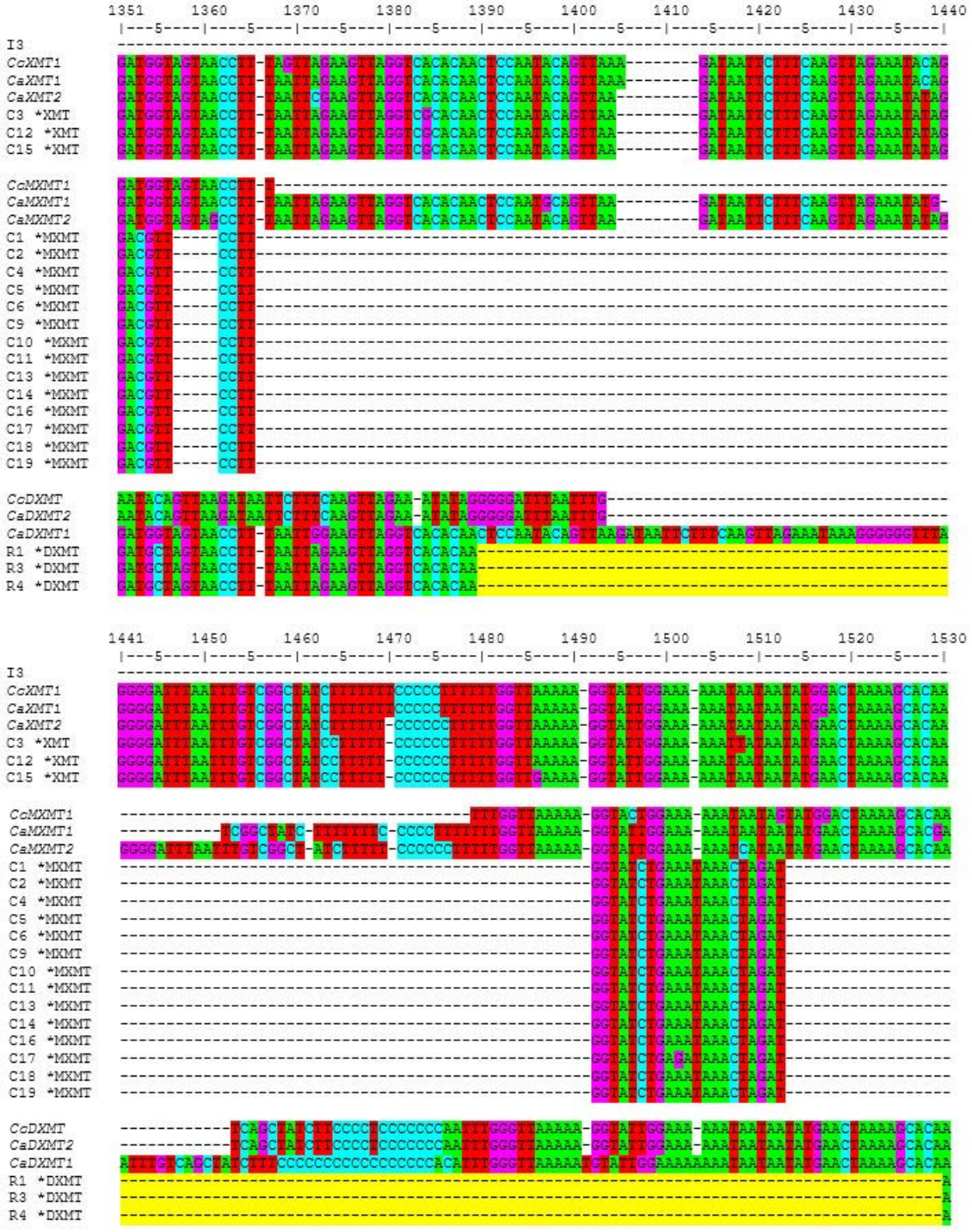

451 460 470 480 490 500 510 520 530 540
E2
CcXMT1
CaXMT1
CaXMT2
C3 *XMT
C12 *XMT
C15 *XMT
CcMXMT1
CaMXMT1
CaMXMT2
C1 *MXMT
C2 *MXMT
C4 *MXMT
C5 *MXMT
C6 *MXMT
C9 *MXMT
C10 *MXMT
C11 *MXMT
C13 *MXMT
C14 *MXMT
C16 *MXMT
C17 *MXMT
C18 *MXMT
C19 *MXMT
CcDXMT
CaDXMT2
CaDXMT1
R1 *DXMT
R3 *DXMT
R4 *DXMT
541 550 560 570 580 590 600 610 620 630
E2
CcXMT1
CaXMT1
CaXMT2
C3 *XMT
C12 *XMT
C15 *XMT
CcMXMT1
CaMXMT1
CaMXMT2
C1 *MXMT
C2 *MXMT
C4 *MXMT
C5 *MXMT
C6 *MXMT
C9 *MXMT
C10 *MXMT
C11 *MXMT
C13 *MXMT
C14 *MXMT
C16 *MXMT
C17 *MXMT
C18 *MXMT
C19 *MXMT
CcDXMT
CaDXMT2
CaDXMT1
R1 *DXMT
R3 *DXMT
R4 *DXMT

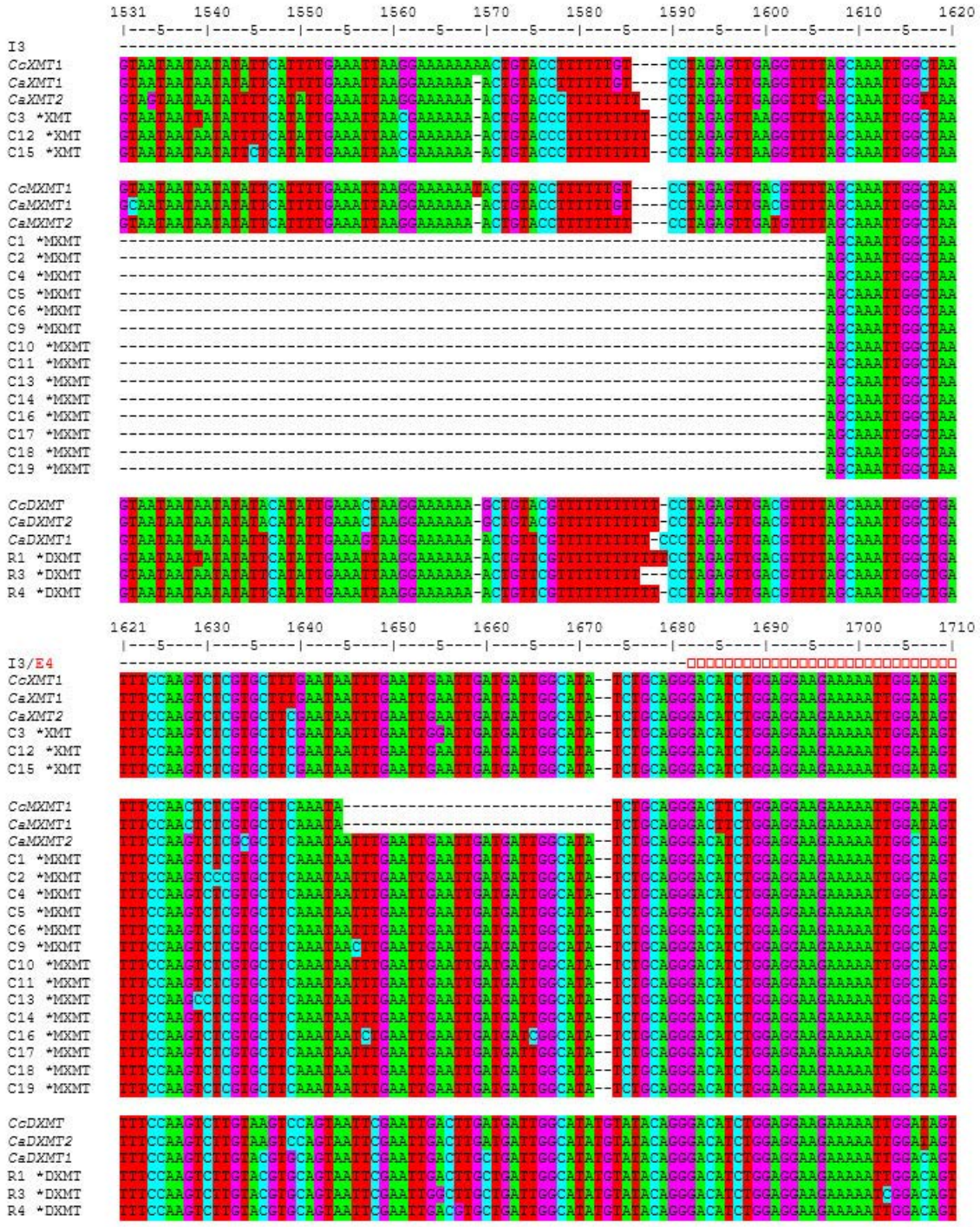

631 640 650 660 670 680 690 700 710
E2/I2
CcXMT1
CaXMT1
CaXMT2
C3 *XMT
C12 *XMT
C15 *XMT
CcMXMT1
CaMXMT1
CaMXMT2
C1 *MXMT
C2 *MXMT
C4 *MXMT
C5 *MXMT
C6 *MXMT
C9 *MXMT
C10 *MXMT
C11 *MXMT
C13 *MXMT
C14 *MXMT
C16 *MXMT
C17 *MXMT
C18 *MXMT
C19 *MXMT
CcDXMT
CaDXMT2
CaDXMT1
R1 *DXMT
R3 *DXMT
R4 *DXMT
721 730 740 750 760 770 780 790 800
I2
CcXMT1
CaXMT1
CaXMT2
C3 *XMT
C12 *XMT
C15 *XMT
CcMXMT1
CaMXMT1
CaMXMT2
C1 *MXMT
C2 *MXMT
C4 *MXMT
C5 *MXMT
C6 *MXMT
C9 *MXMT
C10 *MXMT
C11 *MXMT
C13 *MXMT
C14 *MXMT
C16 *MXMT
C17 *MXMT
C18 *MXMT
C19 *MXMT
CcDXMT
CaDXMT2
CaDXMT1
R1 *DXMT
R3 *DXMT
R4 *DXMT


811 820 830 840 850 860 870 880 890 900
I2
CcXMT1
CaXMT1
CaXMT2
C3 *XMT
C12 *XMT
C15 *XMT
CcMXMT1
CaMXMT1
CaMXMT2
C1 *MXMT
C2 *MXMT
C4 *MXMT
C5 *MXMT
C6 *MXMT
C9 *MXMT
C10 *MXMT
C11 *MXMT
C13 *MXMT
C14 *MXMT
C16 *MXMT
C17 *MXMT
C18 *MXMT
C19 *MXMT
CcDXMT
CaDXMT2
CaDXMT1
R1 *DXMT
R3 *DXMT
R4 *DXMT
901 910 920 930 940 950 960 970 980 990
I2/E3


991 1000 1010 1020 1030 1040 1050 1060 1070 1080
E3
CcXMT1
CaXMT1
CaXMT2
C3 *XMT
C12 *XMT
C15 *XMT
CcMXMT1
CaMXMT1
CaMXMT2
C1 *MXMT
C2 *MXMT
C4 *MXMT
C5 *MXMT
C6 *MXMT
C9 *MXMT
C10 *MXMT
C11 *MXMT
C13 *MXMT
C14 *MXMT
C16 *MXMT
C17 *MXMT
C18 *MXMT
C19 *MXMT
CcDXMT
CaDXMT2
CaDXMT1
R1 *DXMT
R3 *DXMT
R4 *DXMT
1081 1090 1100 1110 1120 1130 1140 1150 1160 1170
E3/I3
CcXMT1
CaXMT1
CaXMT2
C3 *XMT
C12 *XMT
C15 *XMT
CcMXMT1
CaMXMT1
CaMXMT2
C1 *MXMT
C2 *MXMT
C4 *MXMT
C5 *MXMT
C6 *MXMT
C9 *MXMT
C10 *MXMT
C11 *MXMT
C13 *MXMT
C14 *MXMT
C16 *MXMT
C17 *MXMT
C18 *MXMT
C19 *MXMT
CcDXMT
CaDXMT2
CaDXMT1
R1 *DXMT
R3 *DXMT
R4 *DXMT


1171 1180 1190 1200 1210 1220 1230 1240 1250 1260
I3
CcXMT1
CaXMT1
CaXMT2
C3 *XMT
C12 *XMT
C15 *XMT
CcMXMT1
CaMXMT1
CaMXMT2
C1 *MXMT
C2 *MXMT
C4 *MXMT
C5 *MXMT
C6 *MXMT
C9 *MXMT
C10 *MXMT
C11 *MXMT
C13 *MXMT
C14 *MXMT
C16 *MXMT
C17 *MXMT
C18 *MXMT
C19 *MXMT
CcDXMT
CaDXMT2
CaDXMT1
R1 *DXMT
R3 *DXMT
R4 *DXMT
1171 1180 1190 1200 1210 1220 1230 1240 1250 1260
I3
CcXMT1
CaXMT1
CaXMT2
C3 *XMT
C12 *XMT
C15 *XMT
CcMXMT1
CaMXMT1
CaMXMT2
C1 *MXMT
C2 *MXMT
C4 *MXMT
C5 *MXMT
C6 *MXMT
C9 *MXMT
C10 *MXMT
C11 *MXMT
C13 *MXMT
C14 *MXMT
C16 *MXMT
C17 *MXMT
C18 *MXMT
C19 *MXMT
CcDXMT
CaDXMT2
CaDXMT1
R1 *DXMT
R3 *DXMT
R4 *DXMT


1351 1360 1370 1380 1390 1400 1410 1420 1430 1440
I3
CcXMT1
CaXMT1
CaXMT2
C3 *XMT
C12 *XMT
C15 *XMT
CcMXMT1
CaMXMT1
CaMXMT2
C1 *MXMT
C2 *MXMT
C4 *MXMT
C5 *MXMT
C6 *MXMT
C9 *MXMT
C10 *MXMT
C11 *MXMT
C13 *MXMT
C14 *MXMT
C16 *MXMT
C17 *MXMT
C18 *MXMT
C19 *MXMT
CcDXMT
CaDXMT2
CaDXMT1
R1 *DXMT
R3 *DXMT
R4 *DXMT
1441 1450 1460 1470 1480 1490 1500 1510 1520 1530
I3
CcXMT1
CaXMT1
CaXMT2
C3 *XMT
C12 *XMT
C15 *XMT
CcMXMT1
CaMXMT1
CaMXMT2
C1 *MXMT
C2 *MXMT
C4 *MXMT
C5 *MXMT
C6 *MXMT
C9 *MXMT
C10 *MXMT
C11 *MXMT
C13 *MXMT
C14 *MXMT
C16 *MXMT
C17 *MXMT
C18 *MXMT
C19 *MXMT
CcDXMT
CaDXMT2
CaDXMT1
R1 *DXMT
R3 *DXMT
R4 *DXMT


1531 1540 1550 1560 1570 1580 1590 1600 1610 1620
I3
CcXMT1
CaXMT1
CaXMT2
C3 *XMT
C12 *XMT
C15 *XMT
CcMXMT1
CaMXMT1
CaMXMT2
C1 *MXMT
C2 *MXMT
C4 *MXMT
C5 *MXMT
C6 *MXMT
C9 *MXMT
C10 *MXMT
C11 *MXMT
C13 *MXMT
C14 *MXMT
C16 *MXMT
C17 *MXMT
C18 *MXMT
C19 *MXMT
CcDXMT
CaDXMT2
CaDXMT1
R1 *DXMT
R3 *DXMT
R4 *DXMT
1621 1630 1640 1650 1660 1670 1680 1690 1700 1710
I3/E4
CcXMT1
CaXMT1
CaXMT2
C3 *XMT
C12 *XMT
C15 *XMT
CcMXMT1
CaMXMT1
CaMXMT2
C1 *MXMT
C2 *MXMT
C4 *MXMT
C5 *MXMT
C6 *MXMT
C9 *MXMT
C10 *MXMT
C11 *MXMT
C13 *MXMT
C14 *MXMT
C16 *MXMT
C17 *MXMT
C18 *MXMT
C19 *MXMT
CcDXMT
CaDXMT2
CaDXMT1
R1 *DXMT
R3 *DXMT
R4 *DXMT


1711 1720 1730 1740 1750 1760 1770 1780 1790 1800
E4
CcXMT1
CaXMT1
CaXMT2
C3 *XMT
C12 *XMT
C15 *XMT
CcMXMT1
CaMXMT1
CaMXMT2
C1 *MXMT
C2 *MXMT
C4 *MXMT
C5 *MXMT
C6 *MXMT
C9 *MXMT
C10 *MXMT
C11 *MXMT
C13 *MXMT
C14 *MXMT
C16 *MXMT
C17 *MXMT
C18 *MXMT
C19 *MXMT
CcDXMT
CaDXMT2
CaDXMT1
R1 *DXMT
R3 *DXMT
R4 *DXMT
1801 1810 1820 1830 1840 1850 1860 1870 1880 1890
E4
CcXMT1
CaXMT1
CaXMT2
C3 *XMT
C12 *XMT
C15 *XMT
CcMXMT1
CaMXMT1
CaMXMT2
C1 *MXMT
C2 *MXMT
C4 *MXMT
C5 *MXMT
C6 *MXMT
C9 *MXMT
C10 *MXMT
C11 *MXMT
C13 *MXMT
C14 *MXMT
C16 *MXMT
C17 *MXMT
C18 *MXMT
C19 *MXMT
CcDXMT
CaDXMT2
CaDXMT1
R1 *DXMT
R3 *DXMT
R4 *DXMT


1891 1900 1910 1920 1930 1940 1950 1960 1970 1980
E4
CcXMT1
CaXMT1
CaXMT2
C3 *XMT
C12 *XMT
C15 *XMT
CcMXMT1
CaMXMT1
CaMXMT2
C1 *MXMT
C2 *MXMT
C4 *MXMT
C5 *MXMT
C6 *MXMT
C9 *MXMT
C10 *MXMT
C11 *MXMT
C13 *MXMT
C14 *MXMT
C16 *MXMT
C17 *MXMT
C18 *MXMT
C19 *MXMT
CcDXMT
CaDXMT2
CaDXMT1
R1 *DXMT
R3 *DXMT
R4 *DXMT
1981 1990 2000 2010 2020 2030 2040 2050 2060 2070
2071 2080

```
C2  *MXMT    ------------
C4  *MXMT    ------------
C5  *MXMT    ------------
C6  *MXMT    ------------
C9  *MXMT    ------------
C10 *MXMT    ------------
C11 *MXMT    ------------
C13 *MXMT    ------------
C14 *MXMT    ------------
C16 *MXMT    ------------
C17 *MXMT    ------------
C18 *MXMT    ------------
C19 *MXMT    ------------

CcDXMT       TCAGACGTGTAA
CaDXMT2      GCAGACATGTAA
CaDXMT1      TCAGACGTGTAA
R1  *DXMT    ------------
R3  *DXMT    ------------
R4  *DXMT    ------------
```

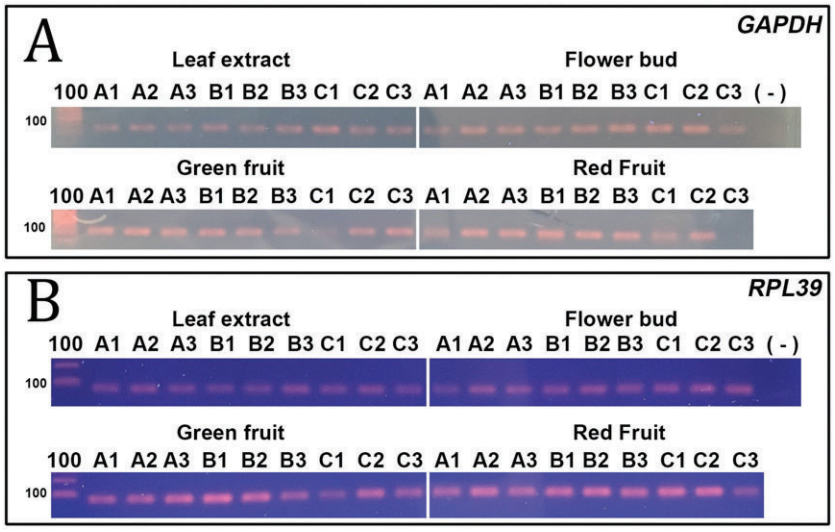


Figure II. Agarose gel showing PCR products after amplifying the synthesized cDNA from leaf, flower bud, green fruit, and red fruit extracts using primers specific for the reference genes, glyceraldehyde 3-phosphate dehydrogenase (*GAPDH)* (A) and ribosomal protein L39 (*RPL39)* (B). Samples were obtained from individual *Coffea liberica* var. *liberica* trees (A, B, and C) in triplicate. A 100-bp (100) ladder was used as standard, and a negative control (–) was also included.